



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>

École Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITÉ DE LORRAINE

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

Par Fanny RISSER

**Études d'un mécanisme enzymatique et d'interactions inter-protéiques
au sein de voies complexes de biosynthèse de polycétides**

Le 16 décembre 2019

Membres du jury :

Rapporteurs :	Mme Mirjam CZJZEK	Directrice de recherche, CNRS, Roscoff
	M Jean-Luc PERNODET	Directeur de recherche, CNRS, Paris
Examineurs :	Mme Kira WEISSMAN	Professeure, Université de Lorraine, Nancy Directrice de thèse
	M Benjamin CHAGOT	Maître de conférence, Université de Lorraine, Nancy Co-directeur de thèse
	Mme Claudine MAYER	Professeure, Université Paris Diderot, Paris
	M Thierry OSTER	Professeur, Université de Lorraine, Nancy
Membre invitée :	Mme Pascale TSAN	Maître de conférence, Université de Lorraine, Nancy

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

ABRÉVIATIONS

A : domaine d'adénylation
ACP : *acyl carrier protein*
ADN : acide désoxyribonucléique
ARN : acide ribonucléique
AT : acyl-transférase
Bcc : *Burkholderia cepacia complex*
BDL : benzenediol lactone
C : domaine de condensation
C_{ter} : extrémité C-terminale
CD : dichroïsme circulaire
°DD : domaine de *docking* situé à l'extrémité C-terminale de sa sous-unité
CFP : *cyan fluorescent protein*
CLF : *chain length factor*
CoA : coenzyme A
Cryo-EM : cryo-microscopie électronique
Da : Dalton
DATD : *N,N'*-diallyl-L-tartardiamide
DD : domaine de *docking*
DH : déshydratase
DHD : déshydratase *docking*
DEB : 6-*deoxyerythronolide B*
DEBS : 6-*deoxyerythronolide B synthase*
EMAA : éthyl 2-méthylacétoacétate
ER : énoyl réductase
FAD : flavine adénine dinucléotide
FAS : synthase d'acide gras
HC : hétérocyclase
IDP : protéine intrinsèquement désordonnée
IDR : région intrinsèquement désordonnée
ITC : titration calorimétrique isotherme
kb : kilobase
 k_{cat} : constante catalytique
 K_{d} : constante de dissociation
 K_{M} : constante de Michaelis

Abréviations

KR : cétoréductase

KS : cétoynthase

LINKS : *Laterally-Interacting Ketosynthase Sequences*

MAO : monoamine oxydase

MAT : malonyl/acétyl-transférase

MCAT : malonyl-CoA : ACP acyltransférase

mFAS : synthase d'acide gras de mammifère

MRSA : *Staphylococcus aureus* résistant à la méthicilline

MS : spectrométrie de masse

GC/MS : spectrométrie de masse couplée à la chromatographie gazeuse

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

MT : méthyltransférase

^NDD : domaine de *docking* situé à l'extrémité N-terminale de sa sous-unité

NRP : peptide non ribosomal

NRPS : synthétases de peptide non ribosomique

N_{ter} : extrémité N-terminale

ORF : cadre ouvert de lecture

PCP : *peptidyl carrier protein*

pb : paire de base

PKS : polycétide synthase

hrPKS : PKS itérative hautement réductrice

nrPKS : PKS itérative non réductrice

prPKS : PKS itérative partiellement réductrice

Ppant : phosphopantéthéine

PPT : phosphopantéthéine transférase

PUFA : acide gras polyinsaturé

RiPPs : peptide ribosomique modifié de manière post-traductionnelle

RMN : résonance magnétique nucléaire

SAM : S-adénosyl-méthionine

SAXS : diffusion des rayons X aux petits angles

TE : thioestérase

VRSA : *Staphylococcus aureus* résistant à la vancomycine

wt : *wild type*

YFP : *yellow fluorescent protein*

6-dEB : 6-désoxyerythronolide B

6-MSA : acide 6-méthylsalicylic

SOMMAIRE

INTRODUCTION	1
I. Les métabolites secondaires	1
I.A. Diversité chimique et fonctionnelle	1
I.B. Essor et « âge d'or » des métabolites secondaires	3
I.B.1. Antibiotiques	3
I.B.2. Anticancéreux	4
I.B.3. Autres cibles thérapeutiques	5
I.C. Apparition des résistances antimicrobiennes et besoin de nouvelles molécules thérapeutiques	6
I.D. Différentes stratégies de découverte et de développement de nouvelles molécules	7
I.D.1. Hémi-synthèse et synthèse chimique	7
I.D.2. <i>Genome mining</i>	10
I.D.3. Biologie synthétique des PKS	12
II. Les polycétides synthases	13
II.A. Découverte	13
II.B. Organisation en <i>cluster</i>	14
II.C. Les trois types de PKS	17
II.C.1. Les PKS de type III	17
II.C.2. Les PKS de type II	18
II.C.3. Les PKS de type I	20
II.C.3.a. Les PKS de type I itératives	20
II.C.3.b. Les PKS de type I modulaires	22
II.C.3.b.i. Les PKS <i>cis</i> -AT	23
II.C.3.b.ii. Les PKS <i>trans</i> -AT	24
II.C.3.b.iii. Différence évolutive entre PKS <i>cis</i> - et <i>trans</i> -AT	27
III. Différentes stratégies d'ingénierie des PKS de type I modulaire	28
III.A. Ingénierie des domaines	29
III.A.1. Quelques exemples de mutations spécifiques	29
III.A.1.a. Inactivation des domaines de modification optionnels	29
III.A.1.b. Changement de la spécificité de substrat des domaines AT	30
III.A.1.c. Changement de la stéréochimie des domaines KR	31
III.A.1.d. Mutasynthèse	32
III.A.2. Échange et insertion de domaines	32

III.B. Ingénierie des modules	36
III.C. Le cas des PKS <i>trans</i>-AT	39
III.D. Questions en suspens	40
III.E. Les enzymes post-PKS, une cible de choix pour l'ingénierie ?	41
IV. Structures et dynamique des modules et des sous-unités	44
IV.A. La mFAS comme modèle structural d'un module de PKS modulaire	45
IV.B. Structure de modules de PKS à moyenne résolution	49
IV.B.1. Structure d'un module entier par cryo-EM	49
IV.B.2. Structures d'une construction modulaire et bimodulaire par SAXS	53
IV.C. Les domaines de <i>docking</i>	56
IV.C.1. Mise en évidence	56
IV.C.2. Différentes classes de domaines de <i>docking</i>	58
IV.C.2.a. Domaines de <i>docking</i> de classe 1	58
IV.C.2.a.i. Classe 1a	58
IV.C.2.a.ii. Classe 1b	60
IV.C.2.b. Domaines de <i>docking</i> de classe 2	61
IV.C.2.c. Domaines de <i>docking</i> particuliers chez les PKS <i>trans</i> -AT	63
IV.C.2.d. Domaines de <i>docking</i> de classe 3	64
IV.C.3. Un code de spécificité hiérarchique	66
IV.D. Organisation macromoléculaire de PKS <i>trans</i>-AT	67
V. Présentation de nos systèmes d'études	72
V.A. La virginiamycine synthase, une PKS <i>trans</i>-AT	72
V.B. L'enacyloxine synthase, une PKS hybride <i>cis</i>-/<i>trans</i>-AT	74
V.C. LkcE, une enzyme post-PKS	77
HYPOTHÈSES ET OBJECTIFS DE TRAVAIL	83

RÉSULTATS	87
Caractérisation d'une nouvelle classe de <i>docking</i> chez une PKS de type <i>trans</i> -AT	87
Caractérisation des interactions entre sous-unités au sein d'une PKS hybride, l'enacyloxine synthase	91
Caractérisation d'une enzyme post-PKS, LkcE	95
DISCUSSION ET PERSPECTIVES	101
BIBLIOGRAPHIE	113

INTRODUCTION

I. Les métabolites secondaires

I.A.Diversité chimique et fonctionnelle

Les métabolites secondaires sont des biomolécules synthétisées par différents organismes : les bactéries, les champignons, les plantes mais aussi les animaux (Demain, 2006). Ils sont considérés comme non-essentiels à la vie de l'organisme producteur, au contraire des métabolites primaires parmi lesquels se trouvent les acides aminés, les carbohydrates et les acides nucléiques. Les métabolites secondaires jouent néanmoins un rôle non négligeable dans la survie des organismes les synthétisant et dans l'adaptabilité de ces derniers à leur environnement. En effet ces molécules peuvent jouer le rôle d'agent chélatant, d'hormone sexuelle ou d'effecteur de différenciation, entre autres. Par exemple chez les bactéries, ces molécules sont utilisées comme agent de symbiose avec d'autres bactéries, des plantes, des nématodes, des insectes, ou encore des animaux supérieurs, mais leur rôle le plus connu est celui d'arme chimique, c'est à dire d'antibiotique, permettant d'éliminer les espèces occupant le même biotope (Demain et Fang, 2000) .

Bien que les métabolites secondaires soient synthétisés principalement à partir de petites molécules simples issues du métabolisme primaire, comme l'acétyl-coenzyme A (acétyl-CoA) ou les acides aminés, ces molécules présentent une grande diversité chimique et fonctionnelle. Six grandes classes (Kabear et al., 2014) sont distinguées selon leur structure et donc selon leur voie de biosynthèse (Figure 1 page 2) :

- Les phénylpropanoïdes, parmi lesquels se trouvent la lignine, les tannins ou les coumarines, qui sont issus du catabolisme de la phénylalanine.
- Les alcaloïdes sont des composés azotés, le plus souvent hétérocycliques et dérivés d'acides aminés. La caféine, la nicotine ou encore la morphine font partie de cette famille.
- Les terpènes et terpénoïdes comme le caoutchouc et la vitamine A sont synthétisés à partir d'unités de base à cinq carbones dérivant de l'isoprène, lui-même provenant de l'acétyl-CoA.
- Les hétérosides ou glycosides, principalement synthétisés par les végétaux, sont des dérivés d'ose modifiés par une molécule non glucidique de nature chimique très variée comme un alcool, un thiol, ou un stéroïde. La salicyline est produite par l'écorce de saule blanc à partir d'une molécule de glucose et d'une molécule d'alcool salicylique ; cette molécule une fois métabolisée forme l'acide salicylique, un anti-inflammatoire.

- Les peptides et acides aminés non protéinogènes, à différencier des vingt-deux acides aminés dits à différencier des vingt-deux acides aminés incorporés durant la synthèse polypeptidique lors de la traduction de l'ARNm par les ribosomes. Les peptides issus de synthétases de peptide non ribosomique (NRPS) sont classés dans cette famille, ainsi que les peptides ribosomiques modifiés de manière post-traductionnelle (RiPPs). De nombreuses molécules issues des NRPS sont utilisées en tant qu'antibiotiques comme l'actinomycine, la vancomycine ou encore la bacitracine.
- Et enfin les polycétides, des macromolécules synthétisées par condensations itératives de blocs acyl-CoA par les polycétides synthases (PKS). Ces enzymes sont présentes chez les bactéries, les plantes et les champignons. Le resvératrol, un antioxydant trouvé dans la peau du raisin et prévenant les maladies coronariennes, fait partie de cette famille de molécules.

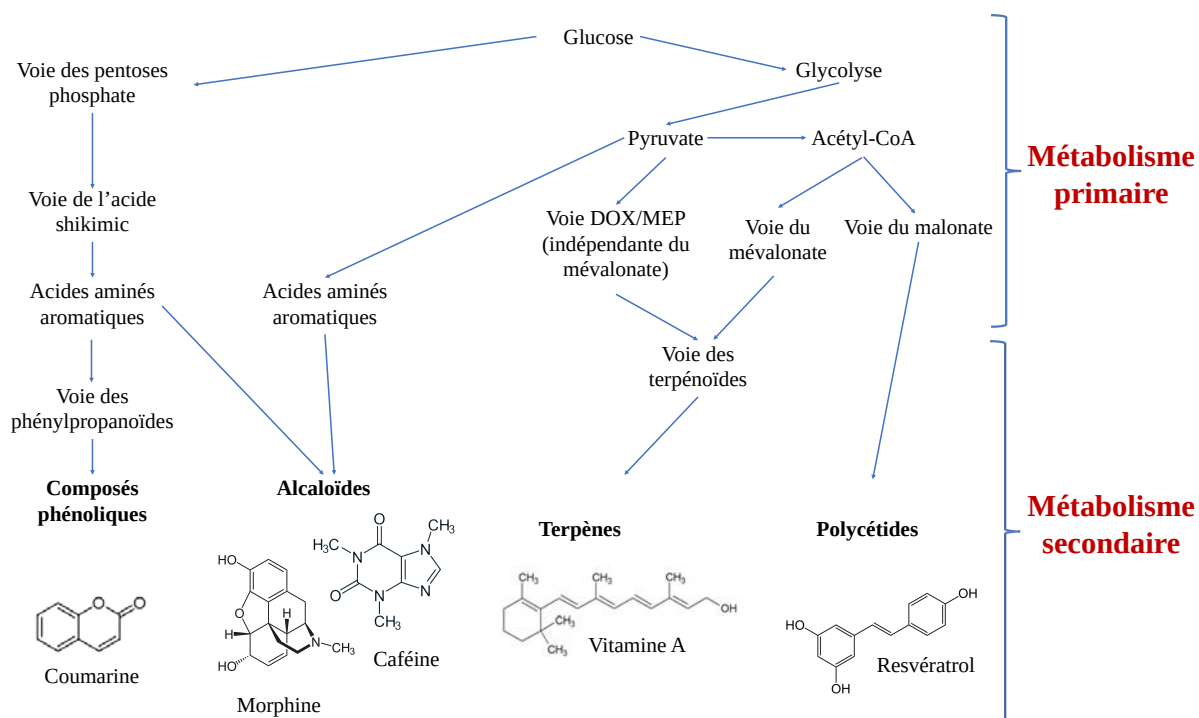


Figure 1 : Voies de biosynthèse des métabolites secondaires découlant du métabolisme primaire chez la plante (adaptée de Ghasemzadeh et Jaafar, 2011).

A ce jour, plus d'un million de métabolites secondaires différents ont été découverts dans la biosphère (Demain, 2014). Ces métabolites représentent une source importante de nouveaux composés à fort intérêt industriel et thérapeutique de par la grande variabilité de leurs structures ainsi que de par leurs diverses et puissantes activités biologiques. Ils sont utilisés en tant qu'antibiotique, anticancéreux et immunosuppresseurs entre autres.

I.B. Essor et « âge d'or » des métabolites secondaires

I.B.1. Antibiotiques

Mis en évidence depuis le milieu du XIX^{ème} siècle, les métabolites secondaires ont connu un essor prodigieux depuis que le rôle antibiotique de certains a été découvert. La pénicilline est le premier antibiotique naturel mis en évidence en 1928 par Alexander Fleming. Cette molécule, un β -lactame, est un peptide non ribosomique (NRP) issu d'une NRPS trouvée chez certains champignons du genre *Penicillium*. En trente ans, la plupart des grandes familles d'antibiotiques, encore utilisées aujourd'hui, ont été découvertes par criblage phénotypique de bactéries et de champignons (Figure 2 page 3).

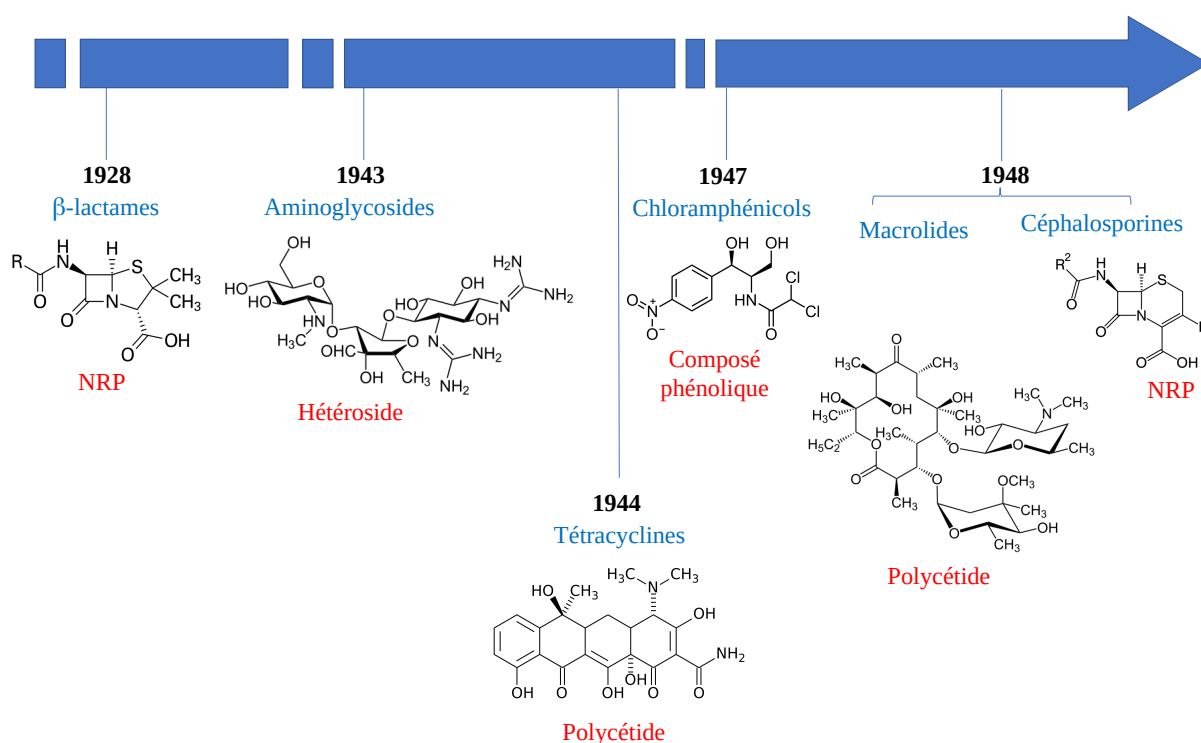


Figure 2 : Dates de découverte des grandes familles d'antibiotique (en bleu la famille d'antibiotique, en rouge leur famille chimique).

Il s'avère que les bactéries filamenteuses du sol, les *Actinomycètes*, et les champignons, sont des sources riches d'antibiotiques. A partir des années 1940, les employés des compagnies pharmaceutiques ont été encouragés à ramener des échantillons de sol de leur destination de vacances. Les *Actinomycètes* et les champignons très abondants dans ces échantillons glanés aux quatre coins du globe, ont été isolés et mis en fermentation dans différentes conditions. Les métabolites secondaires ont été extraits grâce à des solvants des milieux de culture et leur activité antibiotique contre des micro-organismes pathogènes, testée par bio-autographie. Il est

estimé que toutes les compagnies pharmaceutiques, les instituts de recherche et les laboratoires engagés dans ce processus de criblage phénotypique entre 1950 et 2000, ont criblé de 10 à 20 millions de colonies isolées sur boîtes (Katz et Baltz, 2016).

Le succès thérapeutique des métabolites secondaires antibactériens et de leurs dérivés (chimiques, synthétiques ou biosynthétiques) est tellement important que ce sont quasiment les seuls remèdes utilisés contre les bactéries pathogènes encore aujourd'hui (Demain, 2006). En effet, la découverte des antibiotiques a marqué un virage important dans l'histoire humaine : il est estimé que la multiplication par deux de notre espérance de vie au XXème siècle serait principalement liée à l'utilisation de ces métabolites secondaires bactériens et végétaux (Verdine, 1996). Les antibiotiques sont non seulement utilisés pour traiter les Hommes, mais aussi les animaux d'élevage (près de trente antibiotiques différents sont utilisés dans leur eau et alimentation) et les plantes. Ainsi des millions de tonnes d'antibiotiques ont été produites et employées depuis leur introduction (Davies et Davies, 2010).

I.B.2. Anticancéreux

Bien que les premiers efforts se soient concentrés sur la recherche d'activité antibiotique, les compagnies commencent à étendre leurs criblages dans les années 1950 vers des activités anticancéreuses. Les cribles consistent alors à observer soit la réduction des tumeurs transplantées ou induites expérimentalement chez des rongeurs, soit l'effet cytotoxique sur des lignées de cellules cancéreuses en culture. La bléomycine (Figure 3 page 5), découverte en 1966, est synthétisée par *Streptomyces verticillus*. Cette molécule est un mixte polycétide/NRP qui engendre des cassures de l'ADN. Elle est toujours utilisée actuellement pour traiter certains cancers ovariens et testiculaires. Les métabolites secondaires s'avèrent être une incroyable source d'anticancéreux, représentant 74% de ces molécules thérapeutiques mises sur le marché entre 1985 et 2005 (Demain, 2014).

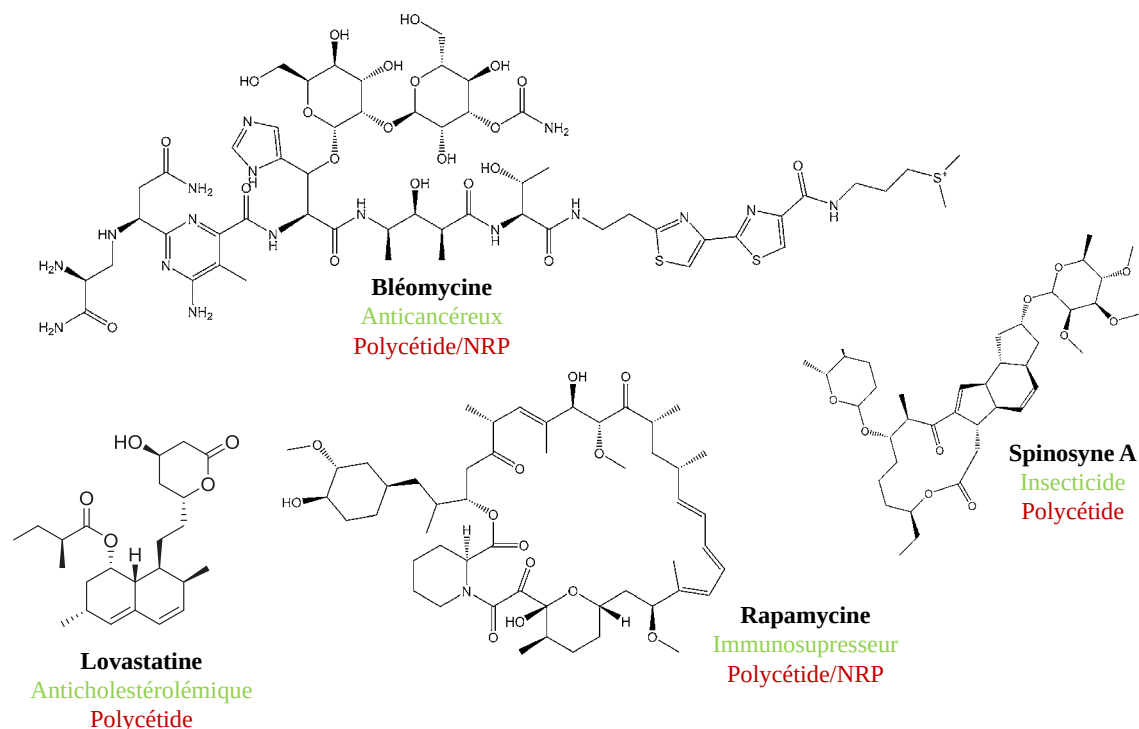


Figure 3 : Structure de métabolites secondaires d'intérêt thérapeutique (adaptée de Katz et Baltz, 2016).

I.B.3. Autres cibles thérapeutiques

Dans les années 1970, d'autres sources de métabolites secondaires, telles que les algues, les bactéries et invertébrés marins, ont été testées. Sans pour autant délaisser les criblages phénotypiques, d'autres cribles plus spécifiques et plus diversifiés commencent à faire leur apparition, c'est le début de la *target oriented strategy*. Le but est de découvrir de nouvelles molécules actives et de trouver une seconde vie aux « antibiotiques ratés » (Demain, 2006).

A titre d'exemple, une percée importante dans le traitement de l'hypercholestérolémie est permise par la découverte et le développement des molécules de la famille des statines, telles que la compactine ou la lovastatine (Figure 3 page 5). Ces molécules sont des polycétides synthétisés par des champignons du genre *Aspergillus* qui inhibent l'hydroxyméthylglutaryl-CoA réductase de la voie de biosynthèse du cholestérol (Alberts, 1988). Un hybride cyclique polycétide/NRP, la rapamycine, et un peptide cyclique, la cyclosporine A (Figure 3 page 5) sont des immunosuppresseurs largement utilisés après des transplantations d'organes afin de diminuer le risque de rejet. Un inhibiteur de glucosidase intestinale, l'acarbose (un pseudotetrasaccharide), a été découvert chez un actinomycète et permet de lutter contre le diabète de type 2. Des cribles pour des activités antagonistes d'autres organismes ont aussi été mis au point, permettant la découverte d'agents fongicides, herbicides et insecticides (Demain, 2006).

En 2003, le marché de ces composés pèse 40 milliards de dollars et plus de 50 000 tonnes de métabolites secondaires et dérivés sont produites par an dans le monde (Demain, 2014).

I.C.Apparition des résistances antimicrobiennes et besoin de nouvelles molécules thérapeutiques

Outres les intérêts industriels et économiques, l'intérêt de découvrir, synthétiser et modifier des métabolites secondaires reste considérable pour plusieurs raisons : obtenir des molécules thérapeutiques plus sûres et plus efficaces, améliorer leurs propriétés pharmacologiques, combattre les virus (Ebola, le virus de l'immunodéficience humaine responsable du SIDA par exemple), les bactéries (*Borrelia burgdorferi*, responsable de la maladie de Lyme), les cancers pour lesquels nous ne disposons pas encore de traitement efficace, et enfin combattre les virus et bactéries naturellement résistants ou ayant acquis une résistance.

Ce phénomène de résistance aux antibiotiques est l'un des problèmes majeurs de santé publique de notre siècle. En Europe, 25 000 personnes meurent tous les ans des suites d'infections par les bactéries multi-résistantes (Martens et Demain, 2017), tandis qu'aux États-Unis, d'après les *Centers for Disease Control and Prevention*, environ deux millions de personnes sont infectées par des bactéries résistantes aux antibiotiques et 23 000 en meurent par an (<https://www.cdc.gov/drugresistance/index.html>). Toujours aux États-Unis, rien que la souche de *Staphylococcus aureus* résistante à la méthicilline (MRSA) est responsable de 19 000 morts et 360 000 hospitalisations par an (Martens et Demain, 2017). Les seuls antibiotiques encore actifs face à cette souche sont la vancomycine et le linezolide. Malheureusement, des souches de *Staphylococcus aureus* résistantes à la vancomycine (VRSA) font de plus en plus leur apparition (Martens et Demain, 2017). Les mécanismes biologiques d'émergence des bactéries pathogènes multi-résistantes sont complexes et variés, mais il est clair que leur multiplication est en grande partie le résultat d'une pression de sélection incessante provenant de la sur-utilisation des antibiotiques et de leur dissémination dans l'environnement par l'activité humaine ces soixante dernières années (Davies et Davies, 2010).

Pourtant, cette crise n'est pas expliquée uniquement par la surutilisation des antibiotiques et l'apparition rapide des souches résistantes. Elle peut aussi l'être par une chute du nombre de nouvelles molécules découvertes et donc mises sur le marché. De 1960 à 1980, le taux de découverte de nouveaux antibiotiques chez les *Streptomyces* seuls est d'environ 75

par an. Ce taux diminue peu à peu jusqu'à 20 par an actuellement et il est prévu de tomber à 15 par an d'ici à 2060 (Watve et al., 2001). Deux problèmes majeurs sont apparus au cours des années 2000 : i) plus de dix millions d'*Actinomycètes* et de champignons avaient déjà été criblés et les mêmes souches productrices ont fini par ressortir des cribles, aucune nouvelle famille d'antibiotique n'a été découverte entre 1962 et 2000, période correspondant à un *gap* d'innovation dans l'industrie pharmaceutique (Fischbach et Walsh, 2009), ii) il était alors pensé que les antibiotiques potentiels à découvrir étaient produits par des micro-organismes beaucoup moins abondants que ceux déjà découverts et beaucoup plus difficiles à cultiver en conditions de laboratoire (Baltz, 2006) (cette dernière hypothèse est largement contredite depuis, grâce à l'émergence du *genome mining*, abordé par la suite).

Pour pallier ces difficultés, plusieurs stratégies prometteuses sont exploitées et développées dans les paragraphes suivants.

I.D. Différentes stratégies de découverte et de développement de nouvelles molécules

I.D.1. Hémi-synthèse et synthèse chimique

Historiquement, la première stratégie utilisée dans le but d'obtenir de nouvelles molécules thérapeutiques repose sur la synthèse chimique. Deux options sont à distinguer : la modification des charpentes moléculaires connues pour leur activité ou la synthèse chimique totale de nouveaux composés.

Dans le cas de l'hémi-synthèse, la charpente moléculaire est laissée intacte de manière à préserver l'activité antibiotique, et seuls les groupements périphériques sont modifiés dans le but d'améliorer les propriétés de la molécule. Dans le cas des antibiotiques, les générations successives ont bien souvent été conçues afin d'élargir le spectre d'activité de la molécule, améliorer leurs propriétés pharmacocinétiques et/ou pharmacodynamiques, mais aussi pour contourner un mécanisme de résistance apparu contre la génération précédente (Figure 4 page 8).

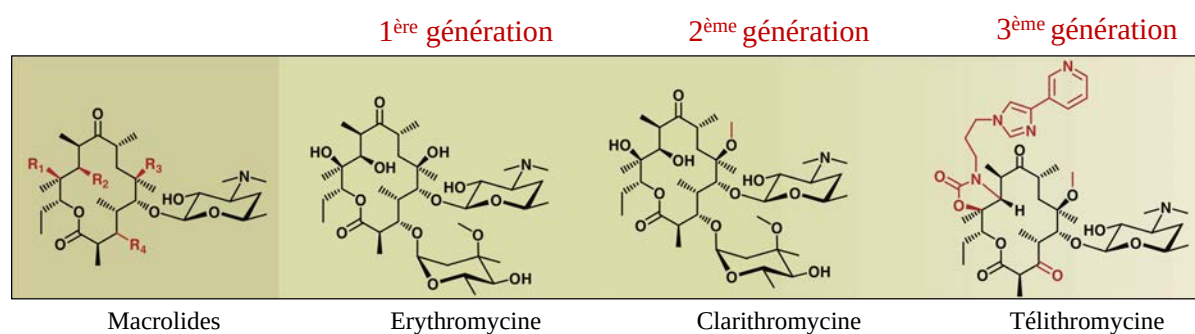


Figure 4 : Générations successives des macrolides obtenus par héli-synthèse (adaptée de Fischbach et Walsh, 2009).

Les charpentes moléculaires sont présentées en noir et les modifications chimiques apportées en rouge.

L'érythromycine est un antibiotique interagissant avec la sous-unité 50s du ribosome empêchant la libération du peptide en croissance. Plus précisément, les interactions se font entre des nucléotides de l'ARN 23s, la fonction hydroxyle du désosamine et les oxygènes portés par les chaînes latérales R_1 et R_3 (Schlünzen et al., 2001). La deuxième génération mise sur le marché a présenté des modifications au niveau de ces chaînes : les groupements hydroxyles originaux engendrent des attaques nucléophiles au niveau de la fonction cétone adjacente, rendant l'antibiotique inactif. La troisième génération, comme la télithromycine, donne naissance à des dérivés « cétolides » : le cladinose y est remplacé par une fonction cétone. Cette substitution ainsi que la modification de la chaîne R_2 permettent d'élargir le spectre d'activité de l'antibiotique, notamment vis à vis de souches résistantes, et améliorent son efficacité (Nilius, 2002).

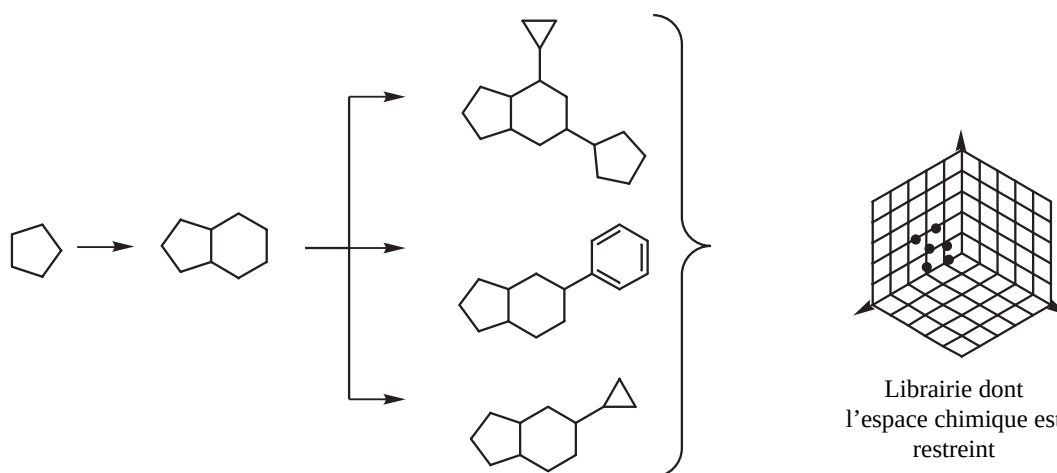
La seconde option, la synthèse totale *de novo*, est développée dans les années 1990 lorsque les cribles des métabolites secondaires commencent à perdre en efficacité. Les industries pharmaceutiques investissent alors massivement dans des banques de chimie combinatoire permettant la synthèse de milliers de nouveaux composés uniques à partir de quelque centaines de petites molécules différentes (Katz et Baltz, 2016). Le développement de la robotique et des cribles enzymatiques et ligand-récepteur, permet aux compagnies de tester ces bibliothèques avec des systèmes de criblage à haut débit. Entre 1995 et 2005, des millions de nouvelles molécules ont été testées contre des centaines de cibles thérapeutiques précises et présélectionnées, potentiellement impliquées dans toutes sortes de pathologies. Malheureusement, une seule molécule synthétisée *de novo* est ressortie de ces efforts de criblage : le sorafenib, un inhibiteur de kinases utilisé dans le traitement du cancer rénal (Newman et Cragg, 2012). En effet, les bibliothèques générées sont composées de petites molécules aux structures très proches, ne variant que par la nature des groupements pendants et

répondant aux règles contraignantes de Lipinski (Linpinski et al., 1997) qui permettent d'évaluer le potentiel médicamenteux d'une molécule. Selon ces règles, une molécule thérapeutique, prise par voie orale, ne doit pas présenter plus de cinq donneurs et plus de dix accepteurs de liaisons hydrogènes, sa masse ne doit pas excéder les 500 daltons et son coefficient de partition octanol-eau ($\log P$) doit être inférieur à cinq, reflétant sa bonne solubilité en milieu aqueux.

Les chimistes synthétiques changent alors de stratégie et créent des librairies avec plus de diversité et de complexité : multiples centres chiraux, groupements hétérocycliques, structures polycycliques, présence élevée d'hétéroatomes, propriétés normalement associées aux molécules naturelles (Newman, 2008). Ils se tournent vers la *diversity oriented strategy* dans le but de créer de plus petites librairies chimiques qui respectent ces aspects structuraux des métabolites secondaires. L'objectif est d'explorer l'espace des structures chimiques puis de déterminer les cibles potentielles, non connue *a priori* (Tan, 2005). Cette stratégie a l'avantage de créer une diversité au sein des squelettes même des molécules (Figure 5 page 10).

A

Librairies combinatoires restreintes



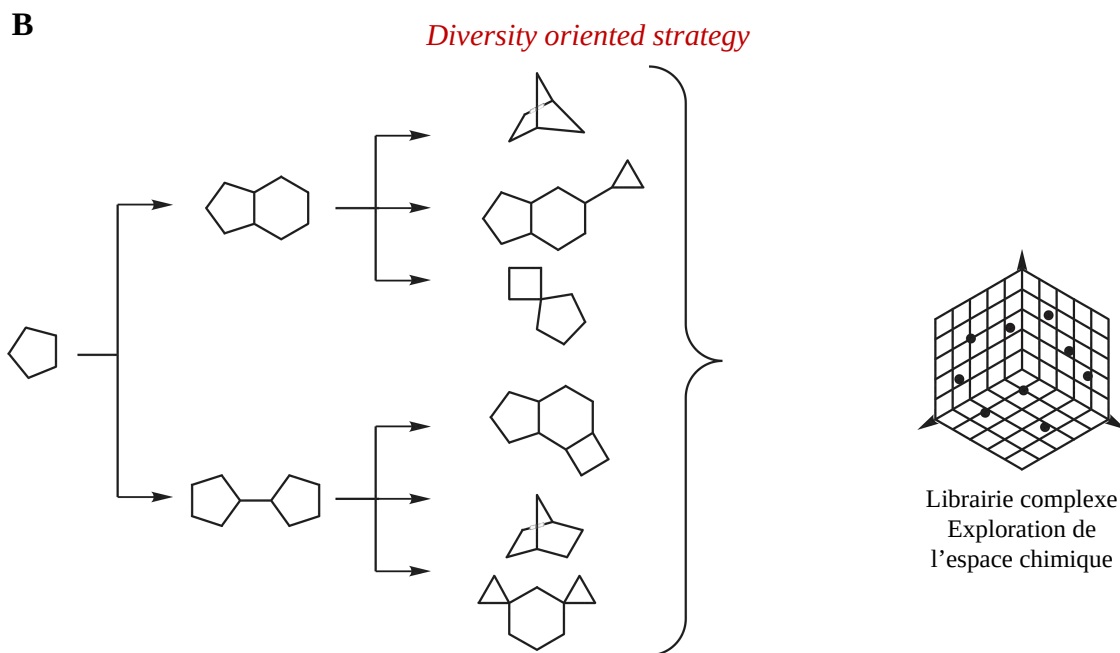


Figure 5 : Comparaison de la conception des bibliothèques et des molécules obtenues par les approches (A) de chimie combinatoire et (B) diversity oriented strategy (adaptée de O'Connell, Galloway, et Spring, 2013).

Ces dernières années, plusieurs molécules *hit* sont sorties de ces bibliothèques restreintes et certaines sont passées au stade d'essai clinique (Liu, Li, et Lam, 2017). Cependant le succès global des bibliothèques synthétiques est assez limité, c'est pourquoi de nombreuses équipes se sont refocalisées sur les organismes et les enzymes synthétisant les métabolites secondaires.

I.D.2. Genome mining

Cette autre stratégie, la plus récente, est permise grâce à la généralisation du séquençage et le développement de la génomique depuis les années 2000. Le potentiel codant des micro-organismes est ainsi révélé, entraînant une révolution dans le domaine : les bactéries possèdent plusieurs *clusters* de gènes permettant la synthèse de métabolites secondaires. Certains *Actinomycètes* possèdent jusqu'à 48 de ces *clusters*, ce qui représente 24% de leur génome total (Katz et Baltz, 2016). Les chercheurs n'auraient en réalité découvert que 10% des métabolites secondaires des souches criblées (Fischbach et Walsh, 2009). En effet, beaucoup des gènes menant à la synthèse de ces molécules ne s'expriment que dans des conditions de culture très particulières et les métabolites produits sont souvent trop faiblement concentrés pour être détectés. D'autre part, seul 1% des bactéries serait cultivable en laboratoire. Le nombre de sources potentielles de métabolites secondaires est donc immense : presque tous les micro-organismes en produisent au moins une douzaine, et le profil métabolique de la plupart des organismes vivants est inconnu (Breitling et al., 2013).

La majorité de ces nouveaux métabolites ne sont que prédits par des analyses bio-informatiques, mais dans certains cas, la production de métabolites secondaires « dormants » a pu être induite par manipulations génétiques (Foulston, 2019). De nombreux outils ont été mis au point ces dernières années afin de manipuler de manière optimale le matériel génétique des *Streptomyces* (Nafissi et Slavcev, 2014). Deux grandes stratégies sont à distinguer : l'expression homologue et l'expression hétérologue. La première stratégie consiste à activer les gènes cryptiques chez l'hôte natif en modifiant la transcription ou la traduction endogène par mutation. L'optimisation du milieu et des conditions de culture entre aussi dans cette première stratégie, car en effet, il est pensé que la production spécifique de certains métabolites secondaires est activée par un changement d'environnement de la bactérie productrice. Pour la seconde stratégie, les *clusters* entiers sont intégrés chez un hôte hétérologue dont les conditions de culture et d'activation de l'expression sont bien maîtrisées. Dans les deux cas les métabolites secondaires produits sont analysés par chromatographie en phase liquide ou gazeuse couplée à un spectromètre de masse (LC/MS ou GC/MS) et leur structure éventuellement résolue par résonance magnétique nucléaire (RMN) (Figure 6 page 11).

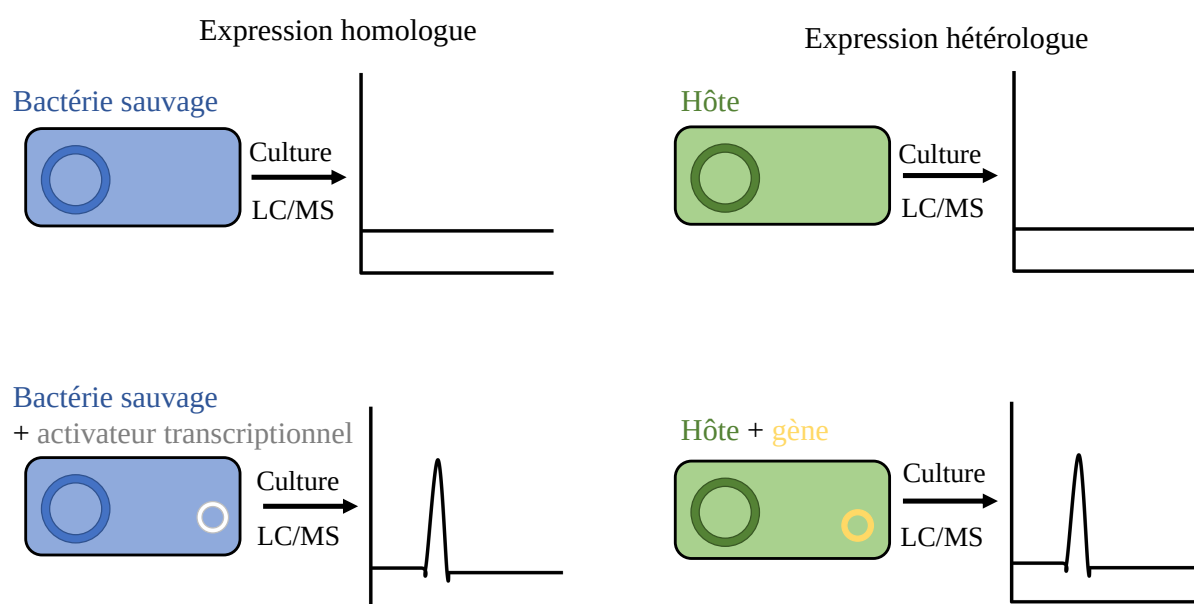


Figure 6 : Comparaison des approches par expression homologue ou hétérologue (adaptée de Scheffler et al., 2013).

Des centaines de nouvelles molécules ont déjà pu être découvertes grâce au *genome mining*, comme les stambomycines chez *Streptomyces ambofacines* (Laureti et al., 2011). Le séquençage du génome de cette bactérie a mis en évidence un *cluster* de 25 gènes dont 9 codent

des PKS. L'expression homologue des protéines a été déclenchée grâce à l'expression constitutive d'un activateur transcriptionnel. La structure des quatre stambomycines produites (divergentes de par la nature d'un groupement pendentif) a été résolue par RMN : ce sont des macrolides cycliques à 51 carbones et ces molécules présentent une activité prometteuse antiproliférative sur certaines cellules cancéreuses. Au vu de la grande richesse de *clusters* biosynthétiques de métabolites secondaires découverts par cette approche, le *genome mining* est et sera certainement pour encore de nombreuses années une source importante de nouvelles molécules (Foulston, 2019).

I.D.3. Biologie synthétique des PKS

La dernière approche possible est l'ingénierie des voies de synthèse naturelles de certains métabolites secondaires. Le but est d'obtenir de nouvelles molécules en modifiant directement ces voies enzymatiques de synthèse des métabolites secondaires. De par le large spectre d'activité des molécules qu'elles produisent et surtout leur architecture qui sera développée en détail par la suite, certaines familles de PKS sont tout particulièrement attractives pour cette stratégie.

Les polycétides sont utilisés en tant qu'antibiotique (érythromycine A, virginamycine M, enacyloxine), anticancéreux (lankacidine, épothilone), antifongique (amphotéricine), immunosuppresseur (rapamycine), antiparasitaire (avermectine B1b), ou même comme insecticide (spinosyne A) (Figure 7 page 12).

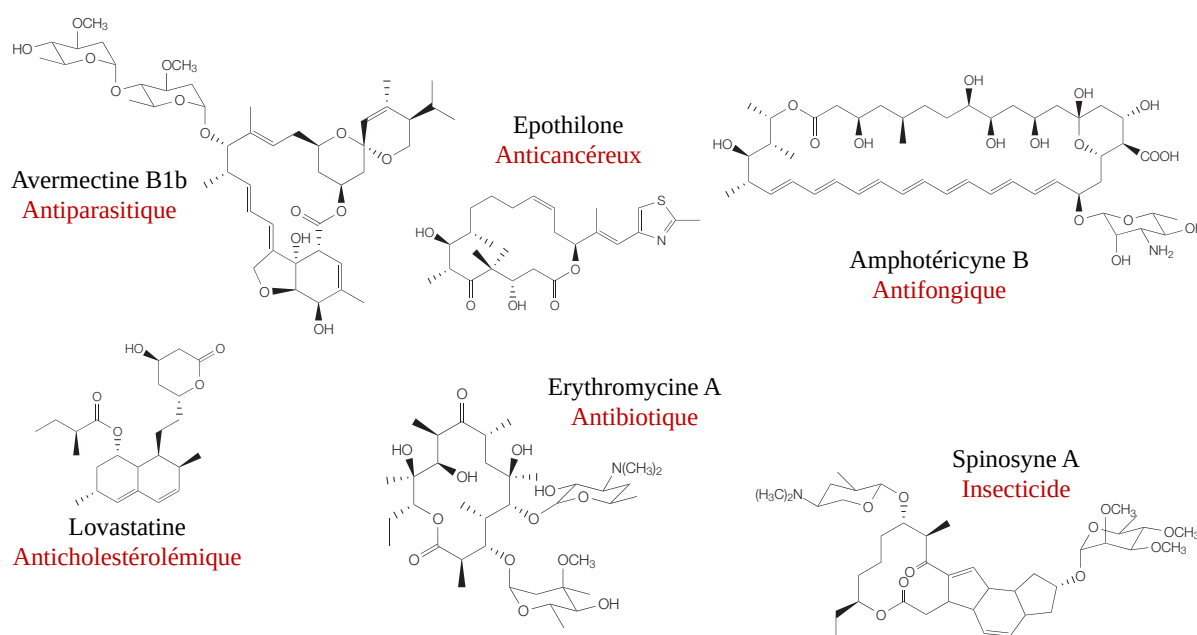


Figure 7 : Exemple de quelques polycétides d'intérêt (adaptée de Weissman et Leadlay, 2005).

En plus de leur grande diversité fonctionnelle, les polycétides présentent une grande variété de structures (Figure 7 page 12), bien que ces molécules soient toutes des dérivés de polycétones. 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 très nombreuses activités pharmacologiques, mais aussi un inconvénient majeur, puisque les polycétides sont des molécules complexes et leur synthèse chimique/hemi-synthétique se révèle laborieuse. La synthèse de ces composés et de dérivés reste donc un challenge pour le domaine médical. Une option de choix est alors d'utiliser la biologie synthétique, et certaines PKS sont d'excellentes candidates grâce à leur organisation modulaire.

II. Les polycétides synthèses

II.A. Découverte

Elles sont mises en évidence de manière indirecte dans les années 1950 grâce à Arthur Birch, à Oxford. Il a eu l'idée que les polycétides pouvaient être synthétisés à partir de plusieurs cycles de condensation d'unité acétates. Sa théorie (Figure 8 page 14) est confirmée en fournissant de l'acétate ^{14}C (marqués au niveau du C_1) (Birch, Massy-Westropp, et Moye, 1955) à *Penicillium patulum*, un champignon synthétisant l'acide 6-méthylsalicylique (6-MSA), un petit polycétide intermédiaire de la toxine patuline.

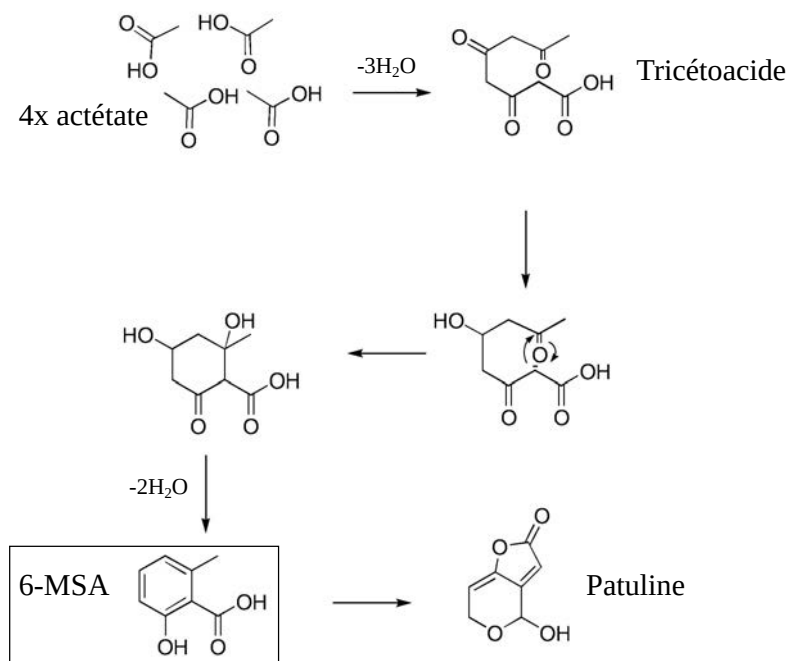


Figure 8 : Séquence de réactions proposée par Birch pour la synthèse de l'acide 6-méthylsalicylique 1 (adaptée de Staunton et Weissman, 2001).

L'intermédiaire 6-MSA, qu'il a pu à identifier grâce à une décomposition chimique, présente bien quatre carbones marqués comme proposé. Jusque dans les années 1960, la seule technique possible pour réaliser des études structurales de ces molécules était très proche de celle utilisée par Birch : elle consistait à dégrader chimiquement des polycétides en des fragments de taille identifiables par spectroscopie de masse (MS). Un travail fastidieux d'assemblage de ces fragments sur papier permettait ensuite d'obtenir une idée générale de la structure de la molécule. Fort heureusement, ce travail s'est trouvé grandement simplifié grâce à l'émergence de la RMN multi-dimensionnelle dans les années 1970. Des précurseurs marqués ^{13}C ont été fournis aux bactéries, les produits naturels ont été isolés et des spectres à plusieurs dimensions ^1H , ^{13}C ont été enregistrés. Cela a permis d'élucider la structure de nombreuses molécules et de décomposer en partie l'enchaînement des réactions chimiques nécessaires à leur synthèse.

II.B. Organisation en cluster

Ce n'est qu'à la fin des années 1970, avec l'émergence de la génétique, que les gènes codant les protéines responsables de la synthèse des polycétides commencent à être mis en évidence. En 1979, Rudd et Hopwood découvrent un *cluster* de gènes chez *Streptomyces coelicolor* permettant la synthèse de l'actinorhodine (Rudd et Hopwood, 1979). Ce polycétide

à activité antibiotique présente une couleur bleue caractéristique. Pour ce faire, les auteurs ont réalisé une série de mutations du génome de *Streptomyces coelicolor*. 76 mutations, engendrant une perte de la couleur bleue des colonies, ont été répertoriées. Les mutants ont été triés en classes phénotypiques distinguées par l'activité antibiotique du polycétide, l'accumulation de précurseurs colorés ou de produits abortifs. La zone du génome assurant la synthèse du polycétide a été précisée, et le *cluster* de gènes permettant la synthèse de l'actinorhodine a été alors séquencé. Le *cluster* ainsi identifié est composé de six gènes (Figure 9 page 15). Leur annotation a été réalisée en comparaison avec d'autres séquences de protéines connues, et notamment avec celles des synthèses d'acides gras (FAS). Les mécanismes de biosynthèse des acides gras et des polycétides sont très similaires : condensations et modifications successives de petites unités acyl-CoA, le tout corrélé à une forte homologie de séquence primaire entre les domaines enzymatiques responsables de ces activités. Ces indices laissent à penser que ces deux systèmes enzymatiques partagent un ancêtre commun.

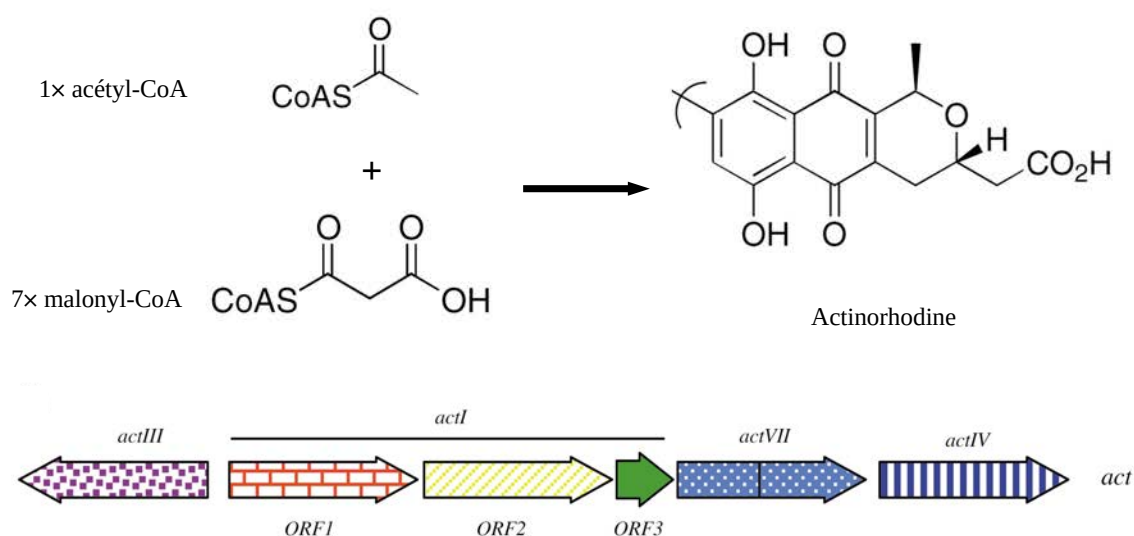


Figure 9 : Cluster de gènes responsables de la biosynthèse de l'actinorhodine mis en évidence par Rudd Hopwood (adaptée de Staunton et Weissman, 2001).

Les cadres ouverts de lecture (ORF) 1, 2 et 3 correspondant à *actI*, codent pour les enzymes catalysant les condensations des unités acétates et malonates, tandis que les gènes *actIII*, *actVII* et *actIV* codent pour des enzymes assurant la modification et la cyclisation du polycétide.

Plus tard, au début des années 1990, deux laboratoires découvrent indépendamment le *cluster* de gènes responsable de la synthèse de l'érythromycine A chez *Saccharopolyspora erythraea* (Bevitt et al., 1992 ; Donadio et al., 1991). Ce *cluster* comporte le gène *ermE*, gène de résistance à l'érythromycine, et les gènes *ery*. Parmi ces derniers, il y avait trois énormes ORF : *eryAI*, *eryAII* et *eryAIII*, faisant environ 10 kilo bases (kb) chacun et codant trois

polypeptides multi-enzymatiques gigantesques faisant environ 350 kilo Daltons (kDa) chacun : DEBS 1, DEBS 2 et DEBS 3 (pour 6-deoxyerythronolide B synthase) (Figure 10 page 16).

Bien d'autres *clusters* de PKS ont été mis en évidence par la suite, et ils comportent globalement les mêmes éléments : les gènes des PKS elles-mêmes, des gènes d'enzymes de modifications et de maturations des polycétides libérés par la PKS, ces protéines sont appelées enzymes post-PKS, mais aussi éventuellement le ou les gènes codant des protéines de résistance à la molécule synthétisée et des éléments de régulation de l'expression. Comme évoqué dans la section I.D.2., ces *clusters* peuvent occuper une part importante du génome des organismes. Par exemple, le *cluster* de l'érythromycine synthase représente près de 1% du génome de *Saccharopolyspora erythraea* (Oliynyk et al., 2007) (Figure 10 page 16), et au total, 3% du génome de cette bactérie est constitué de *clusters* de gènes consacrés à la synthèse de polycétides.

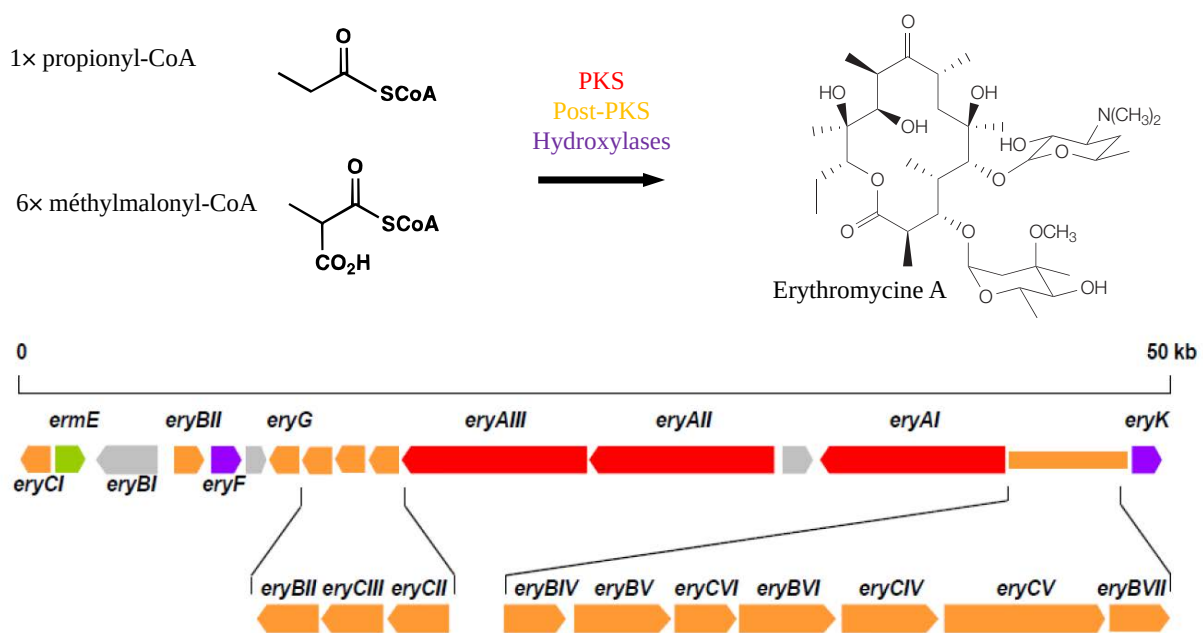


Figure 10 : Cluster de gènes responsables de la synthèse de l'érythromycine A (fournie par KJ Weissman). En rouge, les gènes codant les trois chaînes polypeptidiques de la PKS, en orange les gènes codant les enzymes post-PKS responsables de l'ajout des unités glucidiques et autres modifications, en violet les gènes codant les hydroxylases et en vert le gène conférant la résistance à l'érythromycine A.

L'organisation génique et protéique ainsi que le mécanisme synthétique des PKS permettent de les classer en trois familles distinctes. En comparant les *clusters* des PKS de l'actinorhodine et de l'érythromycine, les chercheurs mettent en évidence que la première est constituée de quelques protéines discrètes tandis que la seconde est composée d'énormes sous-unités multi-enzymatiques, différenciant ainsi les PKS de type II des PKS de type I. Il existe

aussi des PKS de type III, composées d'une seule enzyme synthétisant entièrement le squelette du polycétide.

II.C. Les trois types de PKS

La catalyse réalisée par ces enzymes complexes repose sur la présence de plusieurs activités. Selon le type de PKS, certaines enzymes ou domaines sont présents. Il est trouvé au minimum des cétoréductases (KS) qui catalysent la réaction de condensation entre les unités acyl-CoA chez les PKS de type III. Chez les PKS de type II est aussi présent un *acyl carrier protein* (ACP) qui lie covalamment les intermédiaires de réaction et les achemine au site actif de la KS. Enfin les PKS de type I sont des multi-enzymes présentant de nombreux domaines, notamment KS et ACP, mais aussi des domaines de modification du polycétide.

II.C.1. Les PKS de type III

Elles ont été principalement trouvées chez les plantes où elles sont aussi appelées chalcone ou stilbène synthases, et plus récemment chez les bactéries, les champignons et les algues. Leur composition est la plus simple : une cétosynthase-*like* homodimérique et multifonctionnelle (Yu et al., 2018). Elle sélectionne les blocs de construction de type acyl-CoA, les condense itérativement, et assure la cyclisation finale de la chaîne ainsi que sa libération (Figure 11 page 18). Pourtant, d'autres enzymes de modification peuvent intervenir par la suite (Weissman, 2009). La longueur finale du polycétide semble être dictée par la taille de la cavité accueillant le substrat au sein de la KS (Jez, Bowman, et Noel, 2001). La diversité des polycétides produits est la conséquence d'une grande variété de blocs de départ et d'extension utilisés, du nombre de condensation, du site de cyclisation intramoléculaire, ainsi que de la nature des nombreuses modifications subséquentes (Palmer et Alper, 2019).

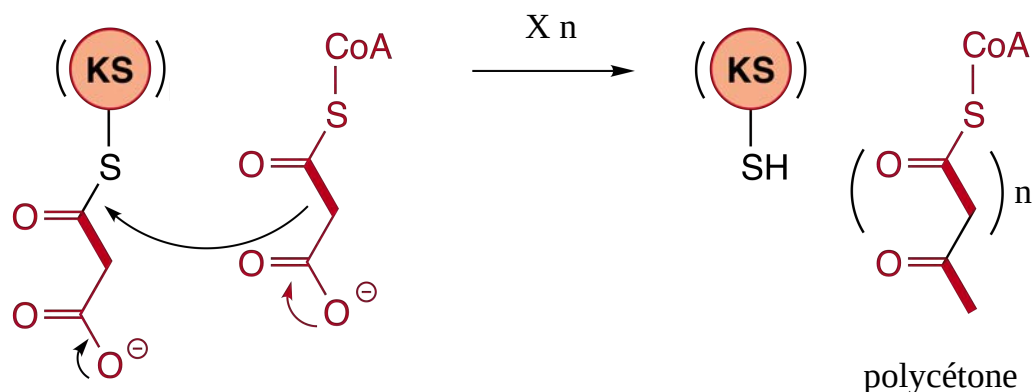


Figure 11 : Représentation schématique des PKS de type III (tirée de Shen, 2003).

Les PKS de type III synthétisent de nombreuses molécules d'intérêt industriel : chalcones, stilbènes, pyrones, curcuminoïdes, quinolones, résorcinols, et naphthalènes, entre autres. Bien que les PKS de type III soient restées des « boîtes noires » durant de nombreuses années, la résolution récente de structures atomiques de certaines d'entre elles a permis l'avancée de l'ingénierie rationnelle, et des dérivés d'intérêt thérapeutique ont été obtenus. Par exemple, chez la PKS de type III d'*Oryza sativa*, les résidus responsables de la sélection des blocs acyl-CoA ont été identifiés et mutés, dans le but d'augmenter la promiscuité de substrat de la KS, et de cette manière, deux nouveaux dérivés présentant des activités antibiotiques ont été identifiés (Go et al., 2015).

II.C.2. Les PKS de type II

Ces PKS qui sont responsables de la synthèse de polycétides aromatiques comme les tétracyclines, sont exclusivement procaryotes et majoritairement trouvées chez les *Actinomycètes*.

Elles sont composées de petites protéines monofonctionnelles codées par différents gènes, comme vu avec l'exemple du *cluster* codant l'actinorhodine synthase. Les enzymes interagissent entre elles au cours de la synthèse, ces interactions sont dites « en *trans* » puisque ce sont des enzymes discrètes et ne sont donc pas dans la même chaîne polypeptidique. Ces complexes sont constitués par l'association minimum de trois partenaires qui agissent de manière itérative. Les deux premiers consistent en une KS hétéro-dimérique : la KS_{α} est catalytique, et assure les condensations successives à partir des unités d'extension. Elle interagit avec une KS_{β} , non catalytique, aussi appelée *chain length factor* (CLF). Cette KS_{β} est dépourvue de cystéine catalytique et assure le contrôle de la longueur de la chaîne polycétidique, soit en orientant les réactions vers un nouveau cycle de polymérisation soit en déclenchant la libération

et la cyclisation du polycétide (Keatinge-Clay et al., 2004). De plus, l'hétérodimère KS_{α}/KS_{β} serait responsable du contrôle de la régiospécificité de la première réaction de cyclisation au cours de la synthèse du polycétide (Hertweck et al., 2007). Le troisième partenaire est un ACP qui sert comme point d'ancrage de la chaîne polycétidique en extension (Figure 12 page 19).

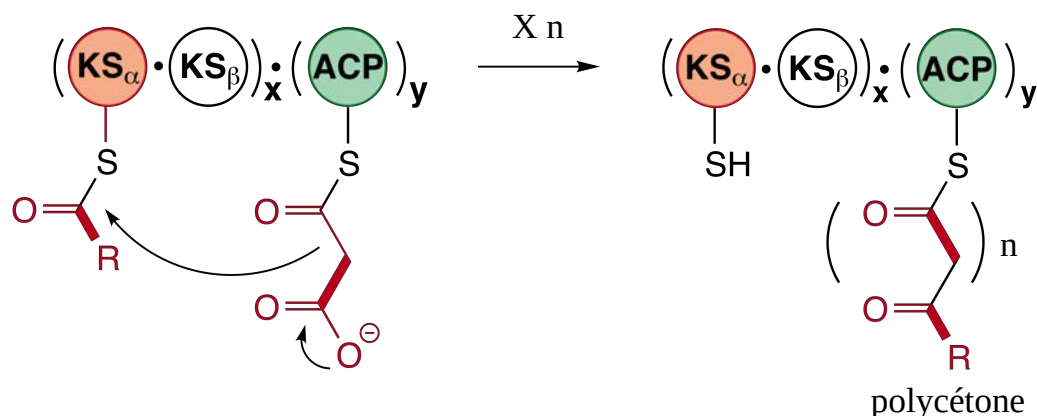


Figure 12 : Représentation schématique des PKS de type II (tirée de Shen, 2003).

Ces domaines ACP subissent une modification post-traductionnelle par l'ajout d'un bras phosphopantéthéine (Ppant) au niveau d'un résidu sérine conservé, cette réaction est catalysée par une enzyme, la phosphopantéthéinyle transférase (PPT). C'est en réalité ce groupement prosthétique qui porte la chaîne en élévation et les unités d'extensions tout au long de la catalyse.

Au contraire des PKS de type III, les PKS de type II n'utilisent pas un large panel de blocs de construction. Dans la majorité des cas, la synthèse débute avec une unité acétyl-CoA et les blocs d'extension sont quasiment exclusivement des unités malonyl-CoA. La sélection de ces unités transférées à l'ACP est réalisée par une malonyl-CoA : ACP acyltransférase (MCAT). De manière étonnante, le gène codant cette enzyme est absent dans la plupart des *clusters* de PKS de type II. Il est donc proposé que la MCAT intervenant dans la synthèse du polycétide soit aussi celle intervenant dans la biosynthèse des acides gras ou encore que les domaines ACP soient capables d'auto-malonylation (Hitchman et al., 1998).

D'autres enzymes optionnelles peuvent agir lors de l'élévation de la chaîne du polycétide, telles que des cétoréductases (KR) qui catalysent la réduction de la fonction cétone en hydroxyle, des cyclases ou des aromatasés (Hertweck et al., 2007). Enfin, le polycétide peut aussi être modifié par des enzymes post-PKS : oxygénases, méthyl- ou glycosyl-transférases, par exemple (Olano, Méndez, et Salas, 2010).

A ce jour, les efforts d'ingénierie de PKS de type II se sont portés sur les différents aspects clés de la synthèse. Par exemple, la longueur de la chaîne polycétidique a pu être modifiée grâce à des mutations ponctuelles de la CLF, et des dérivés de l'actinorhodine plus grands de deux unités malonate ont ainsi été obtenus (Tang, Tsai, et Khosla, 2003). Une autre stratégie consiste à modifier le polycétide une fois libéré par la PKS. Par exemple, des nouveaux dérivés de la chélocardine, une molécule appartenant à la famille des tétracyclines et efficace contre de nombreuses bactéries Gram négatives multi-résistantes, ont été générés en introduisant une aminotransférase et une acyltransférase trouvées chez un autre *Streptomyces* que celui produisant la chélocardine (Lešnik et al., 2015).

II.C.3. Les PKS de type I

Ces PKS sont les plus complexes et les plus variables. Ce sont des multi-enzymes : plusieurs domaines, possédant une structure globulaire propre et se repliant de manière indépendante, sont liés covalamment et chacun porte une activité enzymatique distincte. L'association de trois domaines minimums acyl-transférase (AT), KS et ACP constitue un module et est responsable d'une étape d'élongation de la chaîne acylée. Les domaines AT sont responsables de la sélection spécifique des blocs acyl-CoA. Trois domaines optionnels sont responsables de la modification de la fonction β -cétone : le domaine KR qui catalyse la réduction de la fonction cétone en hydroxyle, le domaine déshydratase (DH) qui élimine la fonction hydroxyle sous forme d' H_2O formant ainsi une double liaison α - β , et enfin le domaine énoyl-réductase (ER) responsable de la réduction de cette double liaison en liaison simple. Ces domaines agissent de manière séquentielle dans l'ordre cité précédemment et leur présence optionnelle assure la diversité de modifications de la fonction cétone (Fischbach et Walsh, 2006). Ainsi, si tous ces domaines optionnels sont présents, un acyl saturé est obtenu lors du cycle d'extension de la chaîne. Un domaine thioestérase (TE) assure le clivage de la liaison thioester entre l'ACP et le polycétide final, libérant ainsi ce dernier sous forme linéaire ou plus souvent, cyclique.

Il existe deux grandes familles de PKS de type I, les itératives et les modulaires (non-itératives) comme détaillées ci-dessous.

II.C.3.a. Les PKS de type I itératives

L'extension de la chaîne du polycétide se réalise au sein d'un seul et même module utilisé à répétition. Le degré de réduction peut varier pour chaque unité d'extension incorporée

: les domaines KR, DH, et ER, sont donc utilisés de manière optionnelle à chaque cycle d'élongation et les mécanismes de programmation de ces activités sont encore mal compris.

Ces PKS sont trouvées principalement chez les champignons mais également chez les bactéries. Chez les champignons, il est possible de diviser cette famille de PKS en trois catégories selon la présence ou l'absence de domaines réducteurs : non réductrices (nr), partiellement réductrices (pr) et hautement réductrices (hr), comme la lovastatine synthase par exemple (Figure 7 page 12). Nombre d'entre elles possèdent aussi un domaine C-méthyltransférase assurant la méthylation du carbone β du bloc de construction à partir d'S- adénosyl-méthionine (SAM) (Weissman, 2009). Chez les bactéries, ce type de PKS est responsable de la production d'énediynes, d'acides gras poly-insaturés (PUFA) à longue chaîne retrouvés dans les membranes de certaines espèces marines (Metz, 2001), et de petits polycétides aromatiques mono ou bicyclique qui servent de blocs de construction à d'autres métabolites secondaires (Zhang et al., 2013).

Des expériences d'ingénierie rationnelle ont été menées sur les PKS itératives fongiques notamment sur les benzenediol lactone (BDL) synthases (Xu et al., 2014). Les BDL présentent des activités agonistes aux œstrogènes, régulatrices à la réponse *heat shock* et régulatrices de la réponse immunitaire chez les mammifères, entre autres. Ils sont synthétisés grâce à un tandem de deux PKS, une première hrPKS qui synthétise la partie lipidique qui est ensuite transférée à une nrPKS qui lui greffe un fragment cyclique. La découverte de nombreuses voies de biosynthèse de BDL a permis l'association par ingénierie génétique de hrPKS et nrPKS de voies de synthèse de BDL différents, conduisant à de nouvelles molécules (Xu et al., 2014) (Figure 13 page 21). Ce résultat signifie que les hrPKS et les nrPKS de voies biosynthétiques distinctes peuvent communiquer entre elles de manière efficace.

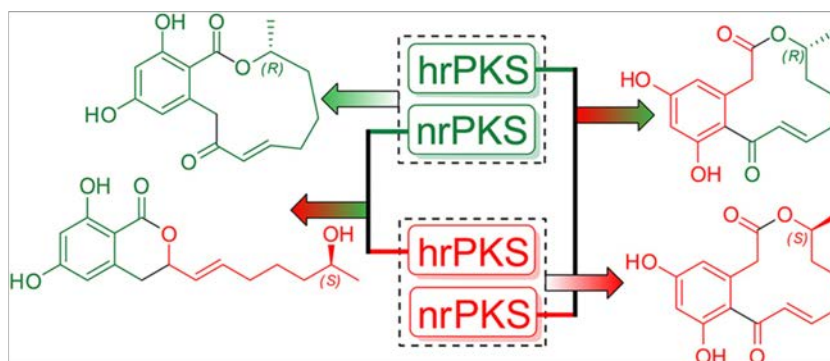


Figure 13 : Design de l'expérience d'échange de hrPKS et de nrPKS menant à la production de nouveaux produits (tirée de Agarwal et Moore, 2014).

En vert la hrPKS et la nrPKS nécessaires à la synthèse du trans-résorcylicide, et en rouge la hrPKS et la nrPKS nécessaires à la synthèse du 10,11-dehydrocurvularine. Différentes combinaisons ont été testées par expression hétérologue et les BDL non naturels obtenus sont présentés en rouge et vert.

Globalement les PKS de type III, II et I itératives sont difficilement maniables dans le cadre d'expériences d'ingénierie, notamment en raison du mode de synthèse itératif, impliquant un nombre restreint d'enzymes. Tous les aspects clefs de ces synthèses ne sont pas encore maîtrisés, même si les études structurales récentes commencent à donner des éléments de réponse. Ainsi les PKS de type modulaire sont plus attractives pour la biologie synthétique.

II.C.3.b. Les PKS de type I modulaires

Il existe d'autres PKS de type I constituées de plusieurs sous-unités protéiques, elles-mêmes comportant un ou plusieurs modules (Figure 14 page 22). Elles sont donc nommées PKS modulaires ou canoniques.

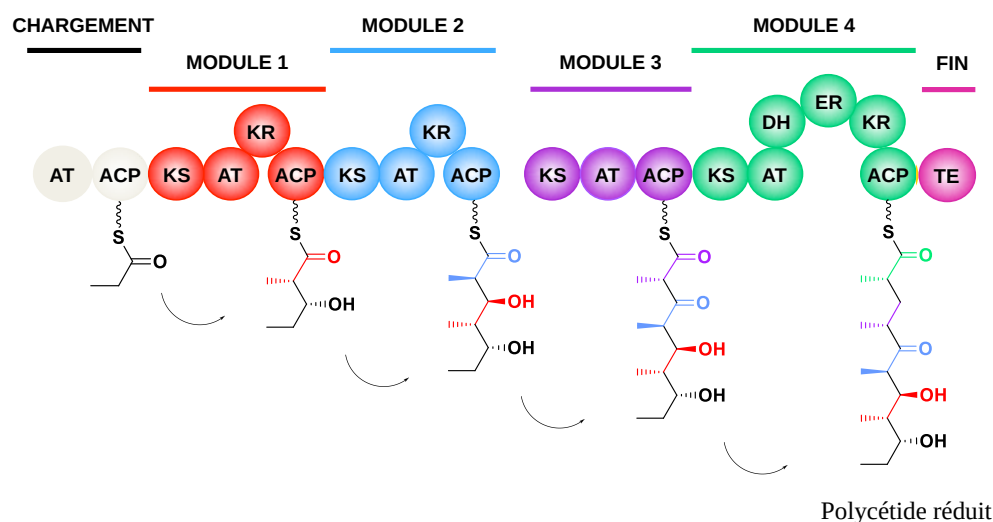


Figure 14 : Représentation schématique d'une PKS de type I non itératives ou modulaire (fournie par KJ Weissman).

Ces protéines font partie des plus grosses enzymes existantes dans le vivant. Par exemple, la PKS qui produit la mycolactone, facteur de virulence de *Mycobacterium ulcerans*, est composée de trois chaînes peptidiques de très haut poids moléculaire : MLSA1 d'une taille de 1,8 MDa comporte 9 modules, MLSA2 d'une taille de 0,26 MDa, est composée d'un seul module et MLSB qui fait 1,2 MDa comporte 8 modules (Stinear et al., 2004). Deux familles de PKS de type I modulaires d'origines évolutives différentes sont à distinguer, les *cis*-AT et les *trans*-AT.

II.C.3.b.i. Les PKS *cis*-AT

Chez ces enzymes, tous les domaines qui composent les modules sont présents au sein de la même sous-unité, c'est-à-dire de la même chaîne polypeptidique, que ce soient les domaines minimaux (AT, ACP et KS), les domaines optionnels (KR, DH et ER) ou encore le domaine de terminaison (TE).

La première PKS de ce genre mise en évidence est DEBS (évoquée en II.B.). Elle est responsable de la synthèse du cœur macrolactone de l'érythromycine (appelé 6-déoxyérythronolide B (6-dEB)) chez *Saccharopolyspora erythraea* (Cortes et al., 1990). Cette PKS est certainement la plus étudiée de toutes depuis la fin des années 1980, et sert donc largement de modèle quant au fonctionnement de ces enzymes (Figure 15 page 23).

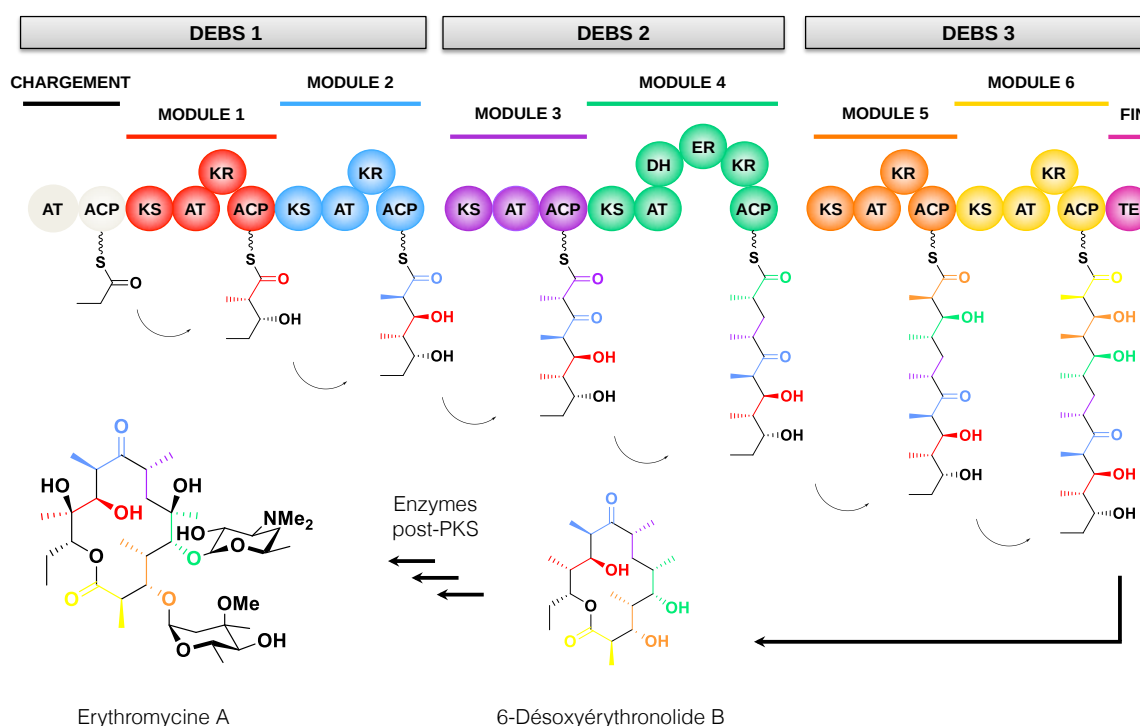


Figure 15 : Représentation schématique de la PKS synthétisant l'érythromycine A (fournie par KJ Weissman).

La PKS DEBS est composée de trois sous-unités protéiques : DEBS 1, DEBS 2 et DEBS 3, toutes composées de 2 ou 3 modules.

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. Cet ACP du module de chargement transfère alors l'unité propionyl à la KS du module 1 au niveau de sa cystéine catalytique. De la même manière, l'AT du module 1 sélectionne un méthylmalonyl-CoA comme unité d'extension et le transfère sur l'ACP du module 1. Cet ACP₁ présente l'unité d'extension à la KS₁, déjà modifiée

par l'unité de départ, et la KS₁ catalyse la réaction de condensation et forme ainsi un dicétide. L'ACP, chargé du dicétide, peut alors le présenter au domaine KR₁ qui catalyse la réaction de réduction de la fonction cétone en hydroxyle au niveau du carbone β. Le dicétide réduit est ensuite transféré à la KS₂ par l'ACP et un nouveau cycle d'extension et de modification peut avoir lieu au sein du module 2. Il y a ainsi six cycles d'extension de la chaîne, un par module, et de modifications chimiques, et en fin de synthèse, le domaine TE présent en C-terminale (C_{ter}) du module 6 catalyse la libération du polycétide et sa cyclisation et forme le 6-dEB. Cet intermédiaire est ensuite mûré par cinq enzymes post-PKS qui vont, entre autres, ajouter deux unités glycosidiques et former l'érythromycine A, la molécule antibiotique.

Au sein d'un même module, les différents domaines fonctionnels sont séparés par des séquences de liaison appelées *linkers*. Ces séquences peuvent être non structurées et très flexibles, riches en alanines, prolines et résidus chargés ou bien présenter une structure tertiaire et jouer un rôle structural important. C'est notamment le cas du *linker* entre les domaines KS et AT qui sert d'adaptateur structural entre ces domaines (Whicher et al., 2014).

Chez ce type de PKS, il existe une relation directe entre la séquence ADN, la séquence primaire de la PKS et donc la composition en domaine des modules, et la nature du composé synthétisé. Chaque module agit dans l'ordre selon lequel il est situé au sein de la séquence polypeptidique, et par conséquent, l'enchaînement en domaines des modules dicte la structure du polycétide synthétisé. Cette relation est décrite sous le terme de règle de la colinéarité (Minowa, Araki, et Kanehisa, 2007) et est propre aux PKS de type *cis*-AT. Ceci a été démontré pour la première fois 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 DEBS. Ces expériences ont mené à la formation des dérivés polycétidiques avec absence des modifications aux positions attendues. Dans ce cas, la modification d'un domaine ou d'un module entraîne un changement direct et prévisible de la structure du polycétide à la position voulue, d'où l'engouement que l'étude de ces enzymes suscite en vue d'approches d'ingénierie génétique.

II.C.3.b.ii. Les PKS *trans*-AT

Chez ces enzymes, certains domaines d'un même module et notamment les AT, sont absents (Piel, 2010). Les ACP sont acylés par l'intermédiaire de domaines AT discrets, situés en dehors de la PKS, et qui interagissent en *trans*. La bacillaene synthase est la première PKS *trans*-AT découverte en 1993 par analyse du génome de *Bacillus subtilis* (Scotti et al., 1993).

Malheureusement, aucune trace du polycétide n'a été détectée, ce qui aurait permis de prouver que cette enzyme était bien fonctionnelle, et le séquençage comportait de nombreuses erreurs. Il a fallu attendre 2002 pour que la première PKS *trans*-AT rattachée à la production d'un polycétide soit mise en évidence chez une bactérie symbiotique de *Paederus fuscipes*, une espèce de scarabée (Piel, 2002). Cette PKS responsable de la synthèse de pédérine, un anti-cancéreux, ne dispose pas d'AT au sein de la séquence des modules et il a été supposé que c'est une AT discrète, dont le gène est présent au sein du *cluster*, qui acyle les ACP. Ce type d'acylation a été confirmée *in vitro* par la suite chez la PKS responsable de la synthèse de leinamycine (Cheng, Tang, et Shen, 2003).

Bien que les *cis*-AT et les *trans*-AT soient similaires en termes de composition en domaines, ces dernières présentent de nombreuses particularités. Alors que les *cis*-AT ne présentent qu'un nombre limité de combinaisons de domaines (selon la présence ou non des domaines réducteurs), une panoplie très étendue de modules est rencontrée chez les *trans*-AT. A ce jour, il en a été découvert plus de 50 distincts (Piel, 2010) et il y a plusieurs raisons pour expliquer cela. D'abord il est possible de rencontrer des modules de NRPS permettant l'incorporation d'acides aminés dans la chaîne polyétidique en croissance au lieu d'unités acyles.

Les NRPS produisent aussi des composés à fort intérêt thérapeutique. Elles présentent une logique synthétique similaire à celle des PKS modulaires. En effet, elles sont aussi composées de plusieurs modules catalysant chacun l'incorporation d'un acide aminé au sein du peptide en extension. Trois domaines différents caractérisent un module minimal. Le domaine d'adénylation (A), comparable au domaine AT, sélectionne spécifiquement un acide aminé (il existe un domaine A différent pour chaque acide aminé) et l'active par adénylation au niveau du groupement carbonyle (Figure 16 page 26). Ce bloc de construction adénylé est ensuite transféré à l'analogue du domaine ACP : le *peptidyl carrier protein* (PCP), qui sert aussi de support à la chaîne en croissance *via* l'intermédiaire d'un groupement phosphopantéthéine. Enfin un domaine de condensation (C), « équivalent » du domaine KS, catalyse la réaction de condensation entre les acides aminés liés aux domaines PCP en formant une liaison peptidique.

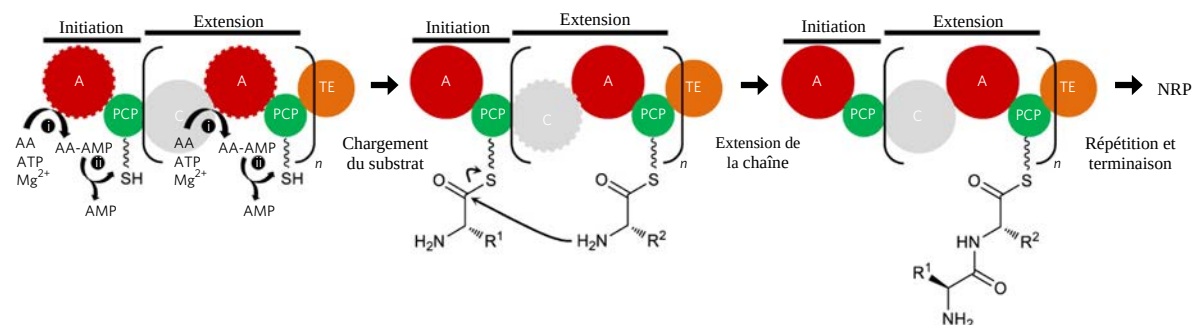


Figure 16 : Représentation schématique de la logique synthétique des NRPS (adaptée de Weissman, 2015). Le domaine actif est indiqué par des contours ondulés. La biosynthèse débute par i) l'adénylation, réalisée par le domaine A, d'un acide aminé spécifique en présence d'ATP et de Mg^{2+} , et ensuite ii) cet acide aminé activé est transféré au domaine PCP du même module servant de support. Le domaine C catalyse la formation de la liaison peptidique en utilisant les deux aminoacyles liés aux PCP comme substrat. Cette séquence de réactions est répétée jusqu'à ce que le peptide final soit libéré par le domaine TE terminal.

Le peptide en croissance peut être modifié par des *N*-méthyltransférases, des épimérases, permettant la présence d'acides aminés de la série D, et il peut aussi subir des hétérocyclisations grâce à des domaines hétérocyclases (HC), des oxydations et des réductions, entre autres modifications (Walsh et al., 2001). La présence de modules NRPS au sein des PKS *trans*-AT assure ainsi l'incorporation d'acides aminés au sein du polycétide, et ces hybrides PKS-NRPS permettent la génération de produits d'une grande variété structurale.

Une autre raison au nombre varié de modules différents rencontrés chez les *trans*-AT réside dans la présence non négligeable au sein de ces derniers de domaines de *C*-méthylation (MT) modifiant le carbone α de l'intermédiaire à partir de SAM. Il n'est pas rare non plus que des modules soient scindés sur deux chaînes polypeptidiques différentes ou qu'ils comprennent des domaines catalytiques disposés dans un ordre inhabituel voire des domaines inactifs. Enfin en cours de synthèse, d'autres enzymes discrètes de modification peuvent agir en *trans* pour assurer des méthylations sur la position β du polycétide en croissance. Un exemple de PKS *trans*-AT est celui de la virginiamycine M synthase de *S. virginiae* (Figure 17 page 27) dont le fonctionnement sera développé en V.A.

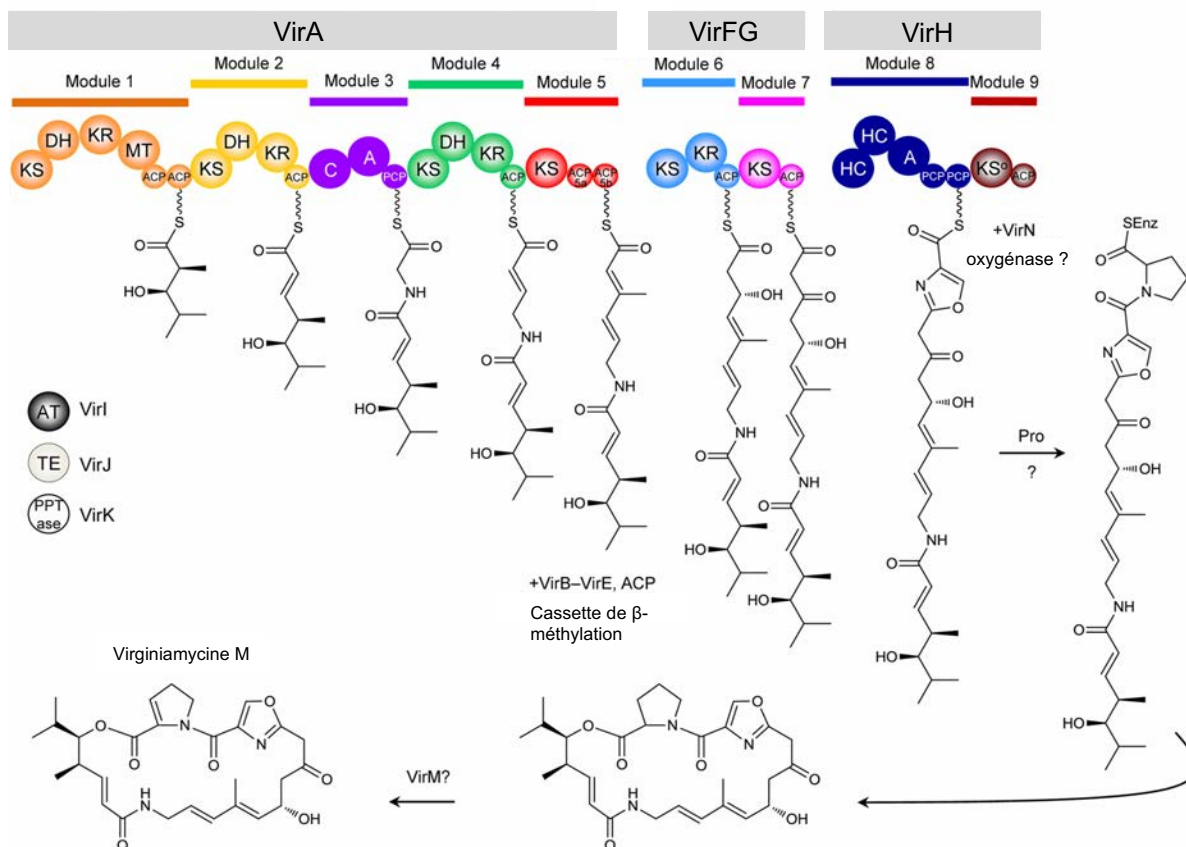


Figure 17 : Représentation schématique de la PKS *trans*-AT synthétisant la virginiamycine M (adaptée de Dorival et al., 2016).

Cette PKS présente une protéine acyltransférase discrète (VirI), un domaine MT au sein du module 1 qui permet l'ajout d'un méthyl sur le premier malonyl incorporé au sein du polycétide, trois modules NRPS (modules 3, 8 et 10) et une TE discrète (VirJ). Un module de chargement est également trouvé dans le cluster de gènes homologue chez *S. pristinaspiraelis*.

II.C.3.b.iii. Différence évolutive entre PKS *cis*- et *trans*-AT

Chez les *trans*-AT, la règle de la colinéarité existante entre la composition en domaines des modules et le polycétide obtenu n'est pas aussi évidente que chez les *cis*-AT, rendant la prédiction de la structure du produit final difficile. Cependant cette prédiction s'est vue améliorée grâce à des travaux phylogénétiques et notamment grâce à la classification des domaines KS en différents clades (Nguyen et al., 2008). En partant de 138 modules de PKS *trans*-AT, des alignements de séquences ont permis de classer les domaines KS en 16 clades distincts. Puis, ces domaines répartis ont été analysés vis-à-vis de leur substrat respectif, prédit *via* la structure du polycétide obtenu et connu. Ainsi, il s'est avéré que la spécificité de substrat des domaines KS est la même au sein de chaque clade. Par conséquent, la nature de la structure

générée aux positions α et β du polycétide en croissance par le module en amont peut être prédite suivant la catégorisation des domaines KS au sein des 16 clades. Cette étude a permis aussi de mettre en évidence des clades pour les domaines MT, alors qu'aucun clade n'est identifié pour les domaines ACP, KR et DH. Enfin, les domaines KS issus d'un même système PKS ne se retrouvent pas nécessairement au sein d'un même clade, suggérant que les *trans*-AT ont émergé au cours de l'évolution par transferts horizontaux de gènes (Nguyen et al., 2008). Au contraire, les *cis*-AT proviendraient de duplications de gènes et de recombinaisons entre modules d'un même système au sein de l'hôte (Jenke-Kodama et Dittmann, 2009), les domaines KS d'un même système se situant dans un même clade.

Malgré leurs modes d'évolution distincts, les *cis*-AT et les *trans*-AT emploient la même logique biosynthétique pour l'assemblage des polycétides. Il n'y a donc *a priori* aucun frein à la coexistence de modules provenant des deux systèmes au sein d'une même synthase. Il est donc surprenant que seulement deux hybrides *cis*-AT/*trans*-AT ne soient caractérisés à ce jour et ce, très récemment. La première mise en évidence, la kirromycine synthase, est majoritairement composée de modules provenant de PKS *trans*-AT et de modules NRPS, et seul le dernier module PKS est de type *cis*-AT (Weber et al., 2008). Au contraire l'enacyloxine synthase est majoritairement composée de modules *cis*-AT et seul le dernier module PKS est d'origine *trans*-AT (Mahenthiralingam et al., 2011). De manière étonnante, ces deux molécules inhibent la synthèse protéique en se fixant dans la même poche du facteur EF-Tu, et bloquent ainsi la protéine dans sa conformation liée au GTP, même si ce dernier ou le GDP ne sont pas fixés. La libération de l'aminoacyl-ARNt du ribosome est ainsi empêchée (Parmeggiani et Nissen, 2006). Trois autres *clusters* de ce type ont été identifiés mais pas encore caractérisés dans une souche de *Clostridium* (Behnken et Hertweck, 2012).

III. Différentes stratégies d'ingénierie des PKS de type I modulaire

Du fait de leur organisation en sous-unités multi-modulaires, ces PKS de type I sont d'excellents modèles pour la biologie synthétique notamment pour le sous-type *cis*-AT. Ces dernières sont maintenant connues et étudiées depuis presque 30 ans et la règle de la colinéarité permet de concevoir de manière très précise les expériences d'ingénierie : en insérant ou en supprimant des domaines par exemple, de manière rationnelle, il est théoriquement possible d'obtenir des dérivés de structure prédite et inversement. La biologie synthétique des PKS modulaires utilise différentes stratégies dans le but de créer de la diversité moléculaire. La

première utilisée et *a priori* la plus évidente, est la modification des domaines de façon individuelle.

III.A. Ingénierie des domaines

A ce jour, la modification des domaines est la plus fréquemment utilisée ; tous les domaines peuvent être ciblés, engendrant des modifications structurales différentes au niveau du polycétide (Figure 18 page 29).

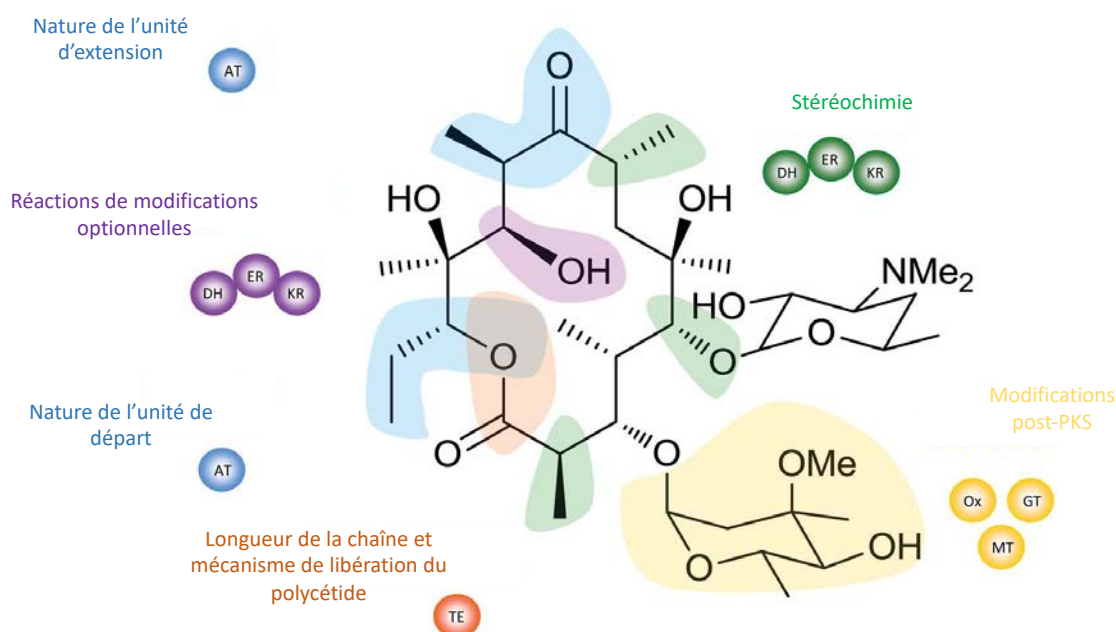


Figure 18 : Différentes possibilités de modification des domaines de PKS pour la génération de nouveaux polycétides, illustrées avec l'exemple de l'érythromycine A (adaptée de Weissman, 2016).
Le type de modification de domaines et les changements structuraux engendrés sur le polycétide sont corrélés selon un même code couleur. Toutes ces modifications sont théoriquement cumulables.

Plusieurs approches sont utilisées. La première consiste à muter des résidus spécifiques du site actif de manière à altérer la spécificité de substrat, la stéréochimie ou inactiver le domaine.

III.A.1. Quelques exemples de mutations spécifiques

III.A.1.a. Inactivation des domaines de modification optionnels

Les premières expériences de ce genre ont été menées très tôt, dans les années 1990 : comme discuté précédemment, des domaines KR et ER de DEBS ont été inactivés en mutants des résidus responsables de la fixation du cofacteur NADPH. Les dérivés de l'érythromycine A, présentant des degrés d'oxydation plus importants aux positions attendues, ont été obtenus (Donadio et al., 1991; Donadio et al., 1993). Le succès de ces expériences peut être expliqué

par le fait que ces mutations sont très peu invasives et ont peu de chance d'engendrer des modifications structurales drastiques des multi-enzymes. Pourtant, une expérience similaire a été menée en mutant l'histidine catalytique du domaine DH₄ de DEBS 2, et seuls des produits abortifs ont été obtenus en très petite quantité, le module suivant la modification introduite ne prendrait plus en charge le polycétide en croissance (Bevitt, Staunton, et Leadlay, 1993). Plus récemment, plusieurs domaines de modification KR, DH et ER de la monensine synthase, ont été inactivés de manière plus extensive. Sur les 32 produits attendus, 22 ont été obtenus, mais avec un rendement maximum sept fois inférieur à celui de la PKS sauvage (Kushnir et al., 2012).

III.A.1.b. Changement de la spécificité de substrat des domaines AT

Pour les domaines AT, des mutations ponctuelles des résidus impliqués dans la spécificité de sélection des blocs de construction, permettent l'intégration de nouvelles unités d'extension au sein du polycétide. En comparant les séquences primaires des AT sélectionnant le malonate à ceux sélectionnant le méthylmalonate, plusieurs signatures ont été identifiées et corrélées à la préférence de substrat. Parmi ces signatures, un motif du site actif a été pointé : YASH pour les AT spécifique du méthylmalonate contre HAFH pour le malonate (Del Vecchio et al., 2003) (Figure 19 page 31). Cependant, le changement de ce motif *in vivo* pour inverser la spécificité du méthylmalonate vers le malonate n'a engendré qu'un inversement partiel du choix de l'unité d'extension ainsi qu'une baisse de 10 à 20% du rendement.

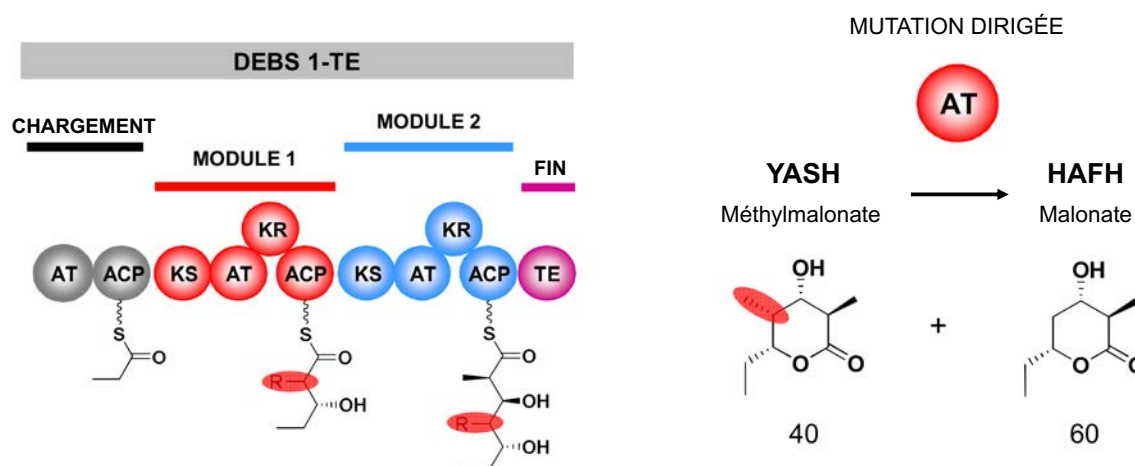


Figure 19 : Design de l'expérience de mutation in vivo du domaine AT dans le but d'inverser sa spécificité de substrat (adaptée de KJ Weissman).

Cette expérience a été menée sur une construction minimale de DEBS comprenant uniquement DEBS 1 et le groupement TE à son extrémité C_{ter}. Le domaine AT du module 1 porte la séquence signature YASH, spécifique de la sélection du méthylmalonate. Cette séquence a été mutée en HAFH spécifique de la sélection du malonate. Les tricétides produits par cette PKS minimum mutée ont été analysés par GC/MS, et deux types de molécules ont été produites. La première population, présente à 60%, ne présente pas le groupement méthyl en position 4. Ceci est attendu suite à la mutation introduite. La seconde population, présente à 40%, correspond au produit natif pour lequel l'AT sélectionne un groupement méthylmalonate malgré la mutation introduite.

Ces résultats suggèrent qu'il existe des déterminants de spécificité de substrat supplémentaires chez les AT. D'autres équipes travaillant directement sur des AT connus pour leur promiscuité de substrat, ont réussi à inverser de manière efficace la spécificité grâce à une seule mutation ponctuelle. Par exemple l'AT₆ de DEBS, portant la signature YASH, sélectionne du méthylmalonate, la mutation de cette signature en RASH a assuré 88% d'efficacité de sélection d'un dérivé alkynyl non naturel (Koryakina et al., 2017). De manière surprenante, cette même mutation appliquée à l'AT₅ de la pikromycine synthase a entraîné une inactivation du domaine (Kalkreuter et al., 2019).

III.A.1.c. Changement de la stéréochimie des domaines KR

Cette stratégie de mutation ponctuelle est aussi utilisée sur les domaines KR dans le but de modifier la stéréochimie du polycétide. Ces domaines, lorsqu'ils réduisent la fonction céto en hydroxyle, peuvent générer des intermédiaires L-3-hydroxyacyl ou D-3-hydroxyacyl. L'obtention de l'un de ces deux stéréoisomères permet de différencier respectivement les KR de type A (production du L-3-hydroxyacyl) de celles de type B (production du D-3-hydroxyacyl) (Caffrey, 2003 ; Reid et al., 2003). De même que pour les domaines AT, des analyses de séquences mettent en relation des motifs signatures spécifiques aux deux catégories : le motif LDD pour les KR de type B, tandis que celles de type A ne le présentent

pas mais possèdent un tryptophane conservé situé à l'opposé du site actif. Les domaines KR sont aussi capables d'une activité épimérase au niveau du centre 2-méthyl adjacent à la fonction céto, et ce même si l'activité cétoréductase est inexistante (Garg, Khosla, et Cane, 2013 ; Garg et al., 2014). Ceci mène à la distinction de six types de domaines KR : celles catalysant la réduction mais sans épimérisation, A1 et B1, celles qui catalysent les deux, A2 et B2, celles qui catalysent seulement l'épimérisation, C2, et enfin celles qui sont totalement inactives, C1 (Keatinge-Clay, 2007). Cependant, aucune signature de l'activité épimérase dans la séquence primaire n'est mise en évidence à ce jour (Weissman, 2016)

Des tentatives d'altération de la stéréochimie ont été donc menées par mutagenèse : soit par changement du motif (Baerga-Ortiz et al., 2006), soit par des mutagenèses à saturation sur ces mêmes positions clés (O'Hare et al., 2006). Les KR mutées et isolées ont bien engendré des produits à la stéréochimie inversée *in vitro* même si une perte d'activité significative a été observée.

III.A.1.d. Mutasynthèse

Cette approche fait appel à l'ingénierie génétique et protéique, et à la synthèse organique (Weissman, 2007). Elle repose sur l'inactivation de la biosynthèse à des stades précoces en ciblant des enzymes impliquées dans la génération des précurseurs, le module de chargement ou le premier domaine KS, par exemple. La biosynthèse du polycétide est court-circuitée en fournissant à la souche productrice des analogues des blocs de construction manquants ou des intermédiaires réactionnels directement. Par cette approche, des dérivés de la rapamycine ont été obtenus en bloquant la voie de biosynthèse de l'unité de départ naturelle et en fournissant à l'hôte producteur *S. hygroscopicus* des unités de départ alternatives (Gregory et al., 2005). Les rendements obtenus ont été jusqu'à vingt fois supérieurs à ceux obtenus sans muter la voie de synthèse du précurseur naturel.

III.A.2. Échange et insertion de domaines

En raison des difficultés à préciser les signatures spécifiques responsables de la sélection des substrats et du mode d'action des domaines enzymatiques, une autre stratégie voit le jour : l'échange et l'insertion de domaines. L'échange de domaines est énormément utilisé pour les AT. Le premier hybride de PKS a été d'ailleurs obtenu en échangeant un AT de DEBS par un AT de la rapamycine synthase (RAPS) (Oliynyk et al., 1996). Par cette approche, les six unités d'extension incorporées à l'érythromycine A ont été modifiées au cours de différentes expériences (Dunn et Khosla, 2013). Il faut tout de même noter que ces expériences ne sont pas un total succès, puisque dans tous les cas, le rendement de production du polycétide final était

largement réduit. Plusieurs hypothèses ont été émises quant à la raison de cette diminution de rendement : une diminution de l'activité du domaine AT hétérologue, l'impossibilité de ce domaine à reconnaître le domaine ACP du module hôte, une non reconnaissance de l'unité d'extension par le domaine KS au moment de la formation de la liaison carbone-carbone, ou encore une perturbation de l'interaction entre les domaines KS et ACP du module hôte lors de la formation de la liaison carbone-carbone (Hans et al., 2003). Des études récentes tendent à montrer que l'interaction entre un domaine AT et son ACP seraient dix fois plus spécifiques qu'avec un ACP hétérologue et que les *linkers* en amont et en aval de l'AT contribuent grandement à l'efficacité et à la spécificité de transacylation de l'ACP (Wong et al., 2010).

Afin de faciliter l'échange des domaines KR, une construction modèle a été mise au point (Kellenberger et al., 2008) : seule la première sous-unité de DEBS 1 a été conservée, le domaine TE étant relocalisé à l'extrémité C_{ter} du module 2, et un *polylinker* nucléotidique contenant plusieurs sites de restriction a été inséré à la place de la séquence codant la KR₂ qui est de type A1 (Figure 20 page 34). L'insertion de domaines KR de type A1 ou B1 (réducteurs et non épimérisants) a permis l'obtention de tricétides attendus dans des rendements proches du modèle sauvage (Kellenberger et al., 2008) ; cependant ces résultats sont fonctions du domaine KR ainsi que des sites de restrictions utilisés lors du clonage. L'insertion de certains domaines KR de type A2 ont permis l'obtention de produits réduits et épimérisés, même si des traces de produits non réduits ou non épimérisés ont été également détectés (Annaval et al., 2015). Enfin, les domaines B2 semblent beaucoup plus récalcitrants à la relocalisation puisque seulement deux d'entre eux sur les six testés ont permis la détection du produit épimérisés (Annaval et al., 2015). Ces résultats suggèrent que l'échange du module entier est requis afin d'obtenir à la fois une cétoréduction et un changement de la stéréochimie du groupement 2-méthyl efficaces.

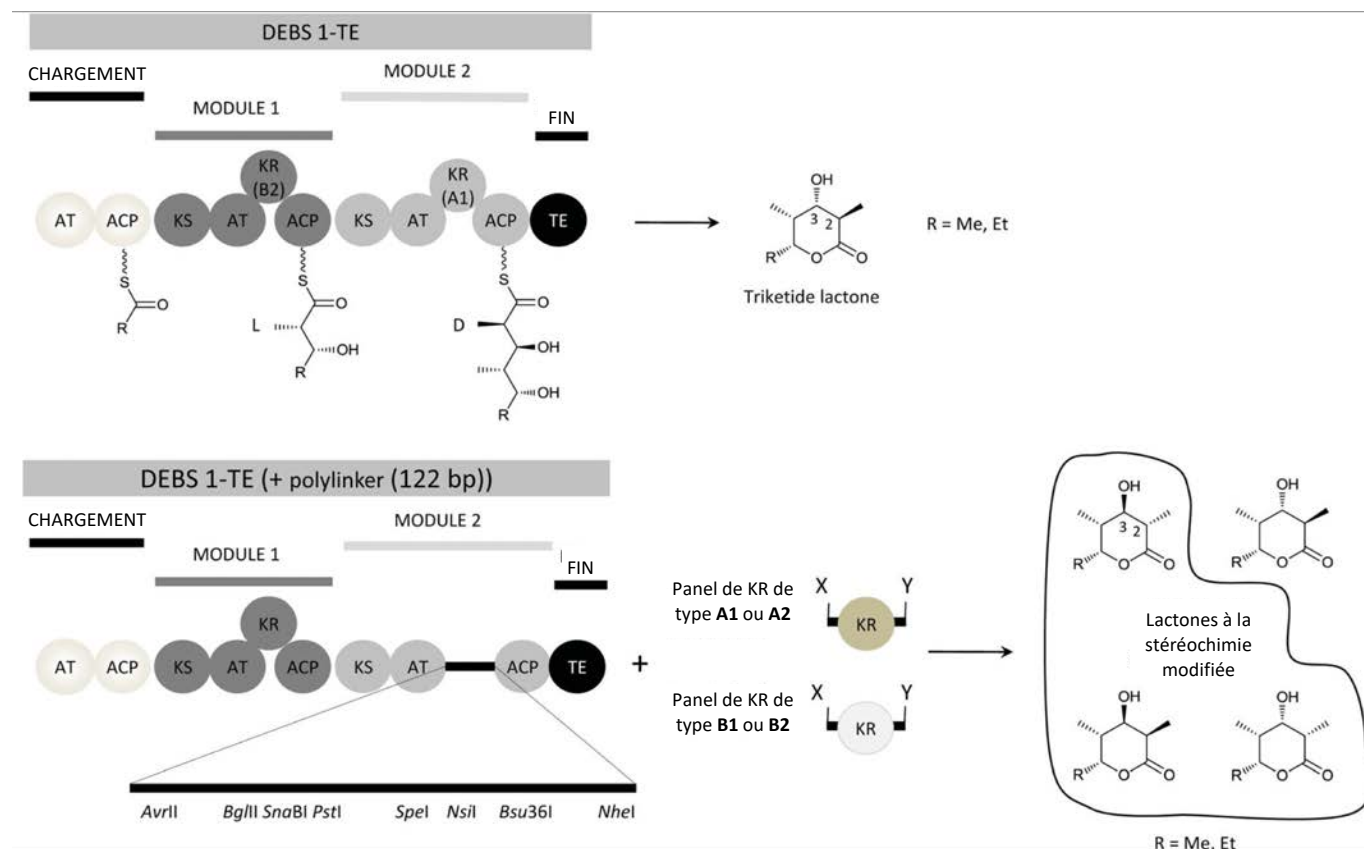


Figure 20 : Schéma de la construction DEBS 1-TE contenant un polylinker remplaçant la séquence de la KR₂ et facilitant ainsi l'insertion d'autres domaines de modifications (adaptée de Weissman, 2016).

La KR native présente dans le module 1 est de type B2 et celle présente dans le module 2 de type A2. Dans ces expériences, cette dernière a été remplacée par une séquence polylinker permettant l'insertion de plusieurs sites de clonage. L'influence des sites de fusion sur le rendement et la stéréochimie a pu ainsi être évaluée. Les structures des quatre lactones générées lors de ces expériences sont présentées, et les trois structures présentant des changements de stéréochimie par rapport à la construction native DEBS 1-TE sont encadrées.

Le domaine KR₂ de DEBS a aussi été échangé par un ensemble de domaines de modification : DH-KR-ER (Kao et al., 1997) au sein d'un modèle trimodulaire dérivé de DEBS (Figure 21 page 35). L'introduction de ces domaines, issus de RAPS, a entraîné une réduction complète de la fonction cétone et donc l'élimination de la fonction nucléophile hydroxyle. Le domaine TE a cyclisé alors exclusivement le polycétide final en formant une lactone à huit membres au lieu de six, avec des rendements très proches de la construction mère.

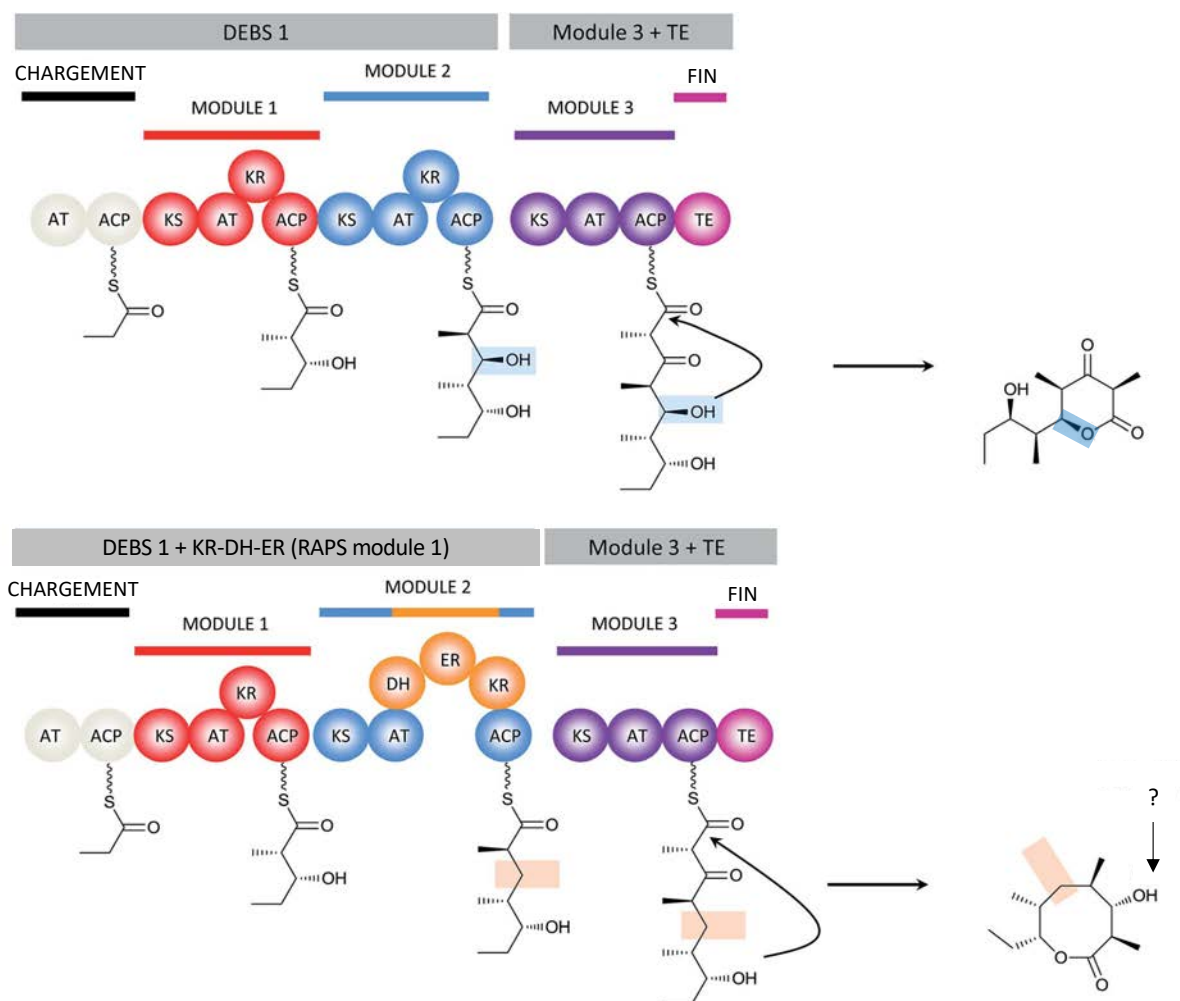


Figure 21 : Génération d'un tétracétide capable de former une octolactone (adaptée de Weissman, 2016).
 La construction modèle initiale comprenait DEBS 1 et le module 3 de DEBS 2 auquel a été accolé le domaine TE. La cyclisation préférentielle du tétracétide a été réalisée par l'hydroxyle porté par le carbone 5, résultant en une lactone à six membres. En intégrant la « boucle de modification » KR-DH-ER au module 2, le domaine TE a été forcé à utiliser l'hydroxyle terminal, formant une lactone à huit membres. Une KR non identifiée serait responsable de la réduction de la fonction céto du carbone 3 (noté ? dans le schéma).

Des modifications cumulatives ont été aussi réalisées sur DEBS (McDaniel et al., 1999) au sein des modules 2, 5 et 6. Les auteurs ont combiné des échanges de domaines AT et KR, avec la substitution de domaines KR par des associations DH-KR ou DH-ER-KR, voire par un fragment synthétique de dix-huit acides aminés entraînant une perte de l'activité. Une librairie de plus de cinquante dérivés a été obtenue ; elle présenterait des modifications du degré de réduction sur un ou plusieurs centres du macrolide, des modifications de stéréochimie ainsi que des modifications de sa structure cœur. Cependant, la vérification structurale des molécules a été effectuée en corrélant les données de MS avec les structures prédites, donc ni la localisation des modifications introduites ni la stéréochimie des produits n'ont été clairement démontrées. De plus, une forte corrélation entre le nombre de changements introduits et les rendements a

été constatée : moins de 0,1 mg.L⁻¹ de rendement pour les doubles et triples mutants contre 20 mg.L⁻¹ pour la souche sauvage.

Au vu des problèmes de rendement et de sous-produits engendrés par les échanges et insertions de domaines, des modules entiers sont aussi échangés dans les PKS. Cette stratégie permettrait de pallier au moins en principe les problèmes de structuration et de communication entre domaines au sein de ces ensembles.

III.B. Ingénierie des modules

L'échange du module de chargement de DEBS avec celui de l'avermectine synthase a permis l'obtention d'une PKS chimère dont l'unité de départ a été bien modifiée (Marsden et al., 1998). Des échanges de modules d'extension ont été aussi réalisés au sein de cette PKS : le module 3 ou le module 6 ont été relocalisés à la suite du module 1. Les molécules attendues ont été bien produites à condition de conserver le *linker* reliant naturellement les modules 1 et 2 (Gokhale, 1999).

Une autre expérience a consisté à insérer un module entier de RAPS entre les modules 1 et 2 de la construction DEBS 1-TE (Figure 22 page 37). Cette PKS allongée a produit le tétracétide attendu mais le produit majeur de synthèse était un tricétide résultant de « l'évitement » du module hétérologue introduit (Rowe et al., 2001 ; Thomas et al., 2002). Une étude précédente avait mis en évidence l'importance du maintien de l'interface naturelle entre l'ACP_n et la KS du module n+1, assurant ainsi le bon transfert de l'intermédiaire d'un module à l'autre (Ranganathan et al., 1999).

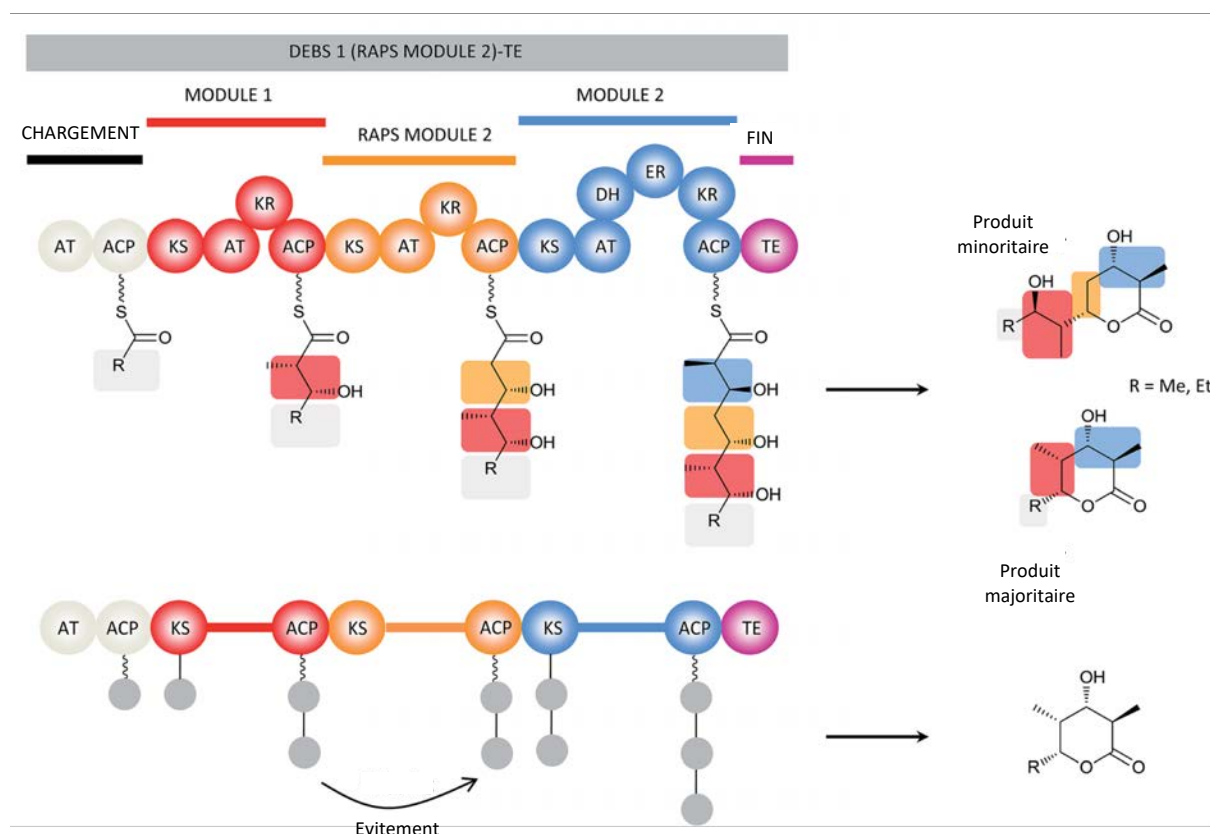


Figure 22 : Tentative d'allongement de la construction DEBS 1-TE par l'insertion d'un module de RAPS (adaptée de Weissman, 2016).

Le module 2 de RAPS a été inséré entre les modules 1 et 2 de la construction DEBS 1-TE. Le tétracétide attendu a été synthétisé de façon minoritaire par rapport au tricétide. L'explication de ce résultat est donnée par le second schéma. La chaîne polycétidique en croissance passe directement de l'ACP du module 1 au module 2 de DEBS. Le module hétérologue de RAPS est ainsi évité et donc ne peut pas procéder à son cycle d'élongation et de modification.

Cette stratégie est employée plus tard sur une autre PKS particulière, la neoauréothine synthase (Traitcheva et al., 2007). Chez cette dernière, le premier module est capable de catalyser entre deux et quatre cycles d'élongation, synthétisant majoritairement la neoauréothine dans le cas de deux itérations et l'homoneoauréothine dans le cas de trois itérations (Figure 23 page 38). La première expérience d'ingénierie sur ce système a consisté à supprimer les modules 2 et 3, seul la KS du module 2 étant conservée et le module 4 y étant accolé. Ainsi la jonction native entre l'ACP du module 1 et la KS du module 2 a été maintenue. Malheureusement, cette chimère était inactive. L'hypothèse qui a été émise est la non prise en charge du substrat par le module 4. La seconde construction a impliqué une délétion moins drastique : seul le module 2 a été supprimé à l'exception de la KS, afin de maintenir la jonction ACP₁-KS₂. Cette PKS était active : les deux produits plus courts de trois carbones ont été bien synthétisés.

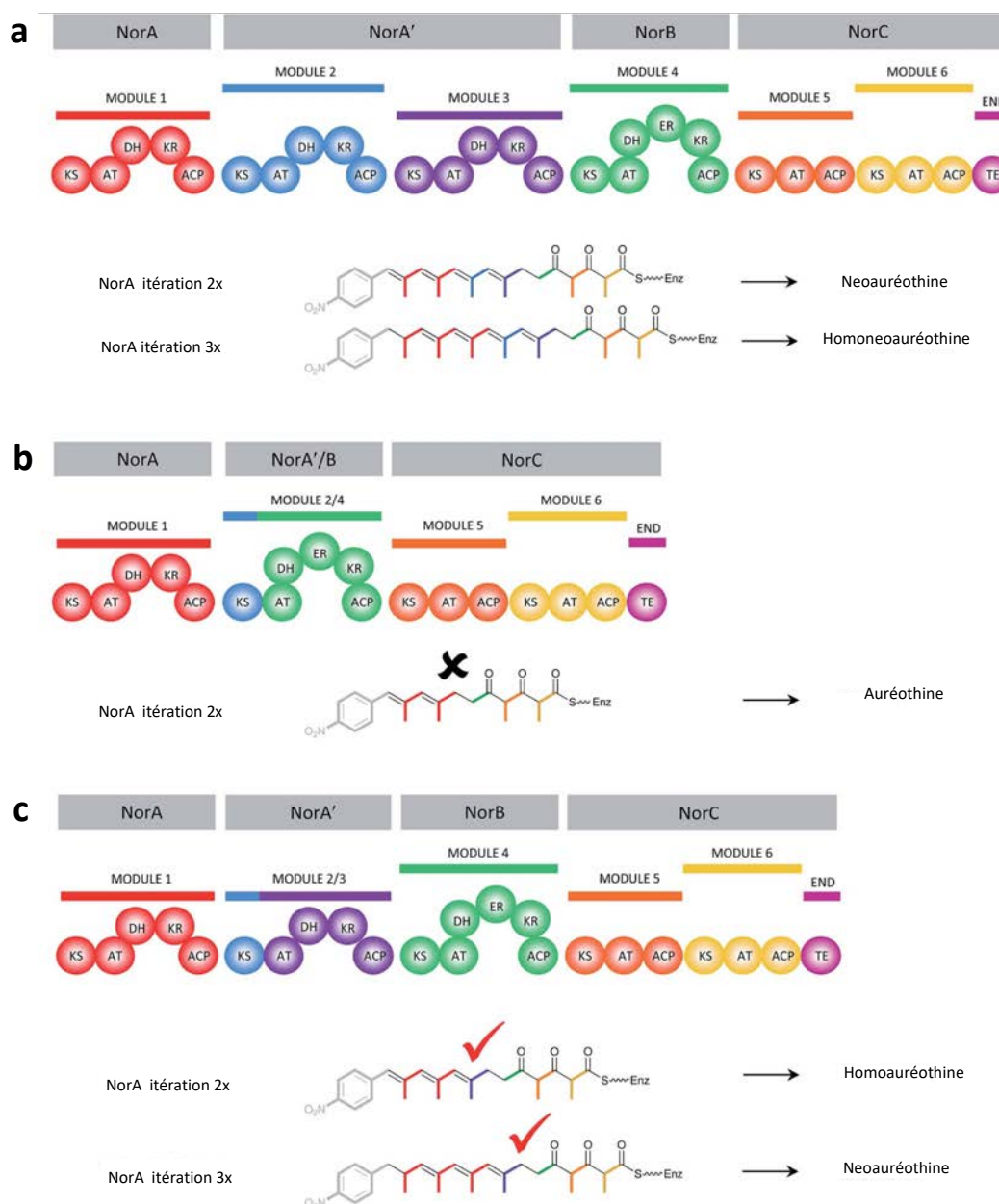


Figure 23 : Suppression de modules au sein de la neoaureothine synthase (adaptée de Weissman, 2016).
a : La PKS native synthétise la neoaureothine grâce à deux itérations du premier module NorA. Si trois itérations ont lieu, alors l'homoneoaureothine est synthétisée.
b : Les modules 2 et 3 de NorA' sont supprimés, et seule la KS₂ est conservée afin de maintenir l'interface naturelle avec l'ACP₁. Pourtant, cette suppression donne naissance à une PKS non fonctionnelle.
c : L'interface naturelle ACP₁-KS₂ est maintenue, et seuls les autres domaines du module 2 sont supprimés. Cette construction a permis la synthèse des deux produits attendus, plus courts de trois carbones par rapport au polycétide naturel.

III.C. Le cas des PKS *trans*-AT

Pour plusieurs raisons, l'approche d'ingénierie est à ce jour principalement menée sur les PKS de type *cis*-AT. La première est historique puisque ces dernières sont mises en évidence depuis le début des années 1990, alors que l'existence des *trans*-AT n'est confirmée que dix ans plus tard. La seconde est d'un ordre plus pratique, puisqu'en raison de la règle de la colinéarité les *cis*-AT sont plus simples à étudier. Pour les *trans*-AT, le lien entre la composition en domaines et le produit obtenu est, au moment de leur découverte, beaucoup plus flou et reste toujours délicat. Pourtant, du fait de leur origine évolutive qui consiste en de multiples transferts horizontaux de gènes, et de leur architecture mosaïque en découlant, elles conviendraient plus aux manipulations génétiques et à la reprogrammation de la synthèse (Till et Race, 2014).

La première caractéristique attrayante et évidente de ces systèmes est la présence d'un domaine AT indépendant puisque ce sont des protéines discrètes, elles semblent mieux convenir à l'expression recombinante, à la purification et donc à la caractérisation. Ensuite les séquences motifs relatives à la spécificité de substrat et le repliement de la protéine sont semblables entre *cis*- et *trans*-AT, rendant possible dans certains cas la prédiction du substrat utilisé (Wong et al., 2011 ; Till et Race, 2014). Il est aussi mis en évidence que certains AT discrets sont capables de sélectionner des substrats atypiques, différents du malonyl-CoA qui est le plus souvent sélectionné chez les *trans*-AT. C'est le cas, par exemple, au sein de la kirromycine synthase évoquée en II.C.3.b.iii. : cette hybride *cis*-AT/*trans*-AT présente la particularité de posséder deux AT discrets au sein de son *cluster* de gènes (Weber et al., 2008). Le premier, KirCI, sélectionne du malonyl-CoA et le délivre à tous les ACP du système (Musiol et al., 2013) à l'exception de l'ACP₅, qui est chargé avec de l'éthylmalonate par KirCII (Musiol et al., 2011).

Des expériences d'inactivation des AT de DEBS montrent que ces dernières peuvent être complémentées par des AT discrètes et notamment KirCII (Dunn et al., 2014). Les auteurs ont inactivé l'AT₁ de DEBS en mutant la sérine catalytique en alanine, et ont complémenté le système par KirCII. Le dérivé du 6-dEB présentant un éthyl en position 12, au lieu du méthyl, a bien été obtenu, malgré une perte d'efficacité du système. La même expérience a été menée avec l'AT de la disorazole synthase qui sélectionne le malonyl-CoA, et de manière étonnante, l'intégration de cette nouvelle unité au sein du 6-dEB a été même plus efficace que l'intégration naturelle de méthylmalonyl-CoA par l'AT₁ de DEBS (Dunn et al., 2014). Plus récemment, une autre étude s'est basée sur la large spécificité de substrat de KirCII afin d'introduire des dérivés alkynes dans la kirromycine (Musiol-Kroll et al., 2017). Cette nouvelle fonction rend accessible

un large spectre de modification par *click chemistry* (Kolb, Finn, et Sharpless, 2001). La *click chemistry* correspond à un ensemble de méthodes permettant la conjugaison d'une sonde ou d'un substrat à une molécule d'intérêt biologique. Classiquement, elle assure la cyclisation de fonctions azides et alkynes en utilisant le cuivre comme catalyseur (cyclo-addition de Huisgen), mais bien d'autres réactions ont été mises au point depuis.

D'autre part, les modules NRPS présents au sein des PKS *trans*-AT peuvent aussi être la cible de l'ingénierie. De la même manière que chez les PKS, des précurseurs alternatifs peuvent être intégrés à la molécule, des enzymes de modifications hétérologues peuvent être ajoutées, halogénases et glycosyltransférases par exemple, de manière à augmenter les modifications possibles. Les domaines et les modules peuvent aussi être échangés, insérés ou supprimés. De plus, des modifications ponctuelles des domaines d'adénylation permettent un changement de sélection d'acide aminé (Winn et al., 2016). Il faut cependant noter que de la même manière que chez les PKS, plus le nombre de modifications est important, moins le rendement final est bon.

III.D. Questions en suspens

De manière générale, le bilan des expériences d'ingénierie menées sur les PKS modulaires est mitigé. Dans la vaste majorité des cas, les rendements de production sont grandement impactés par rapport à celui de la PKS native et lorsque la molécule attendue est synthétisée, elle n'est souvent pas la seule, puisque des sous-produits sont présents. L'approche la moins invasive, l'inactivation de domaines, donne les meilleurs résultats, mais le potentiel de modifications introduites de cette manière est assez limité (Weissman, 2016). Une autre approche, la mutation de résidus clés notamment pour les domaines AT et KR, est freinée par notre compréhension incomplète des mécanismes de reconnaissance du substrat, de la catalyse, et dans le cas des KR, du contrôle de la stéréochimie (Del Vecchio et al., 2003 ; Baerga-Ortiz et al., 2006). Il existe un manque important de structures des domaines à haute résolution en complexe avec leur substrat ou un analogue. D'autre part, le faible succès de ces expériences peut aussi être expliqué par la non reconnaissance du nouvel intermédiaire par les domaines en aval de ceux modifiés (Weissman, 2016). La spécificité envers des substrats non naturels doit donc être plus extensivement étudiée dans le but d'être optimisée.

Quant aux modifications plus importantes, telles que l'échange ou l'insertion de domaines, des problèmes de communications sont engendrés au sein de la PKS. Ces derniers peuvent être issus du changement d'interface entre domaines hétérologues et domaines natifs

du module, impliquant par exemple le domaine ACP (Wong et al., 2010), ou en raison de la création d'un *linker* hybride entre domaines échangés. L'intégrité structurale du module semble être un critère à respecter impérativement dans le but de modifier ces systèmes (Weissman, 2016). La stratégie d'insertion ou d'échange de modules voire de sous-unités est donc la plus prometteuse. Pour ce faire, il est crucial d'augmenter nos connaissances sur la structure et la dynamique des modules à plusieurs stades de la catalyse, afin de déterminer les mécanismes de communication entre domaines. Un autre prérequis est de mieux caractériser les déterminants clés de la communication entre modules consécutifs, au sein d'une même sous-unité ou situés sur deux sous-unités successives. Enfin, un aspect peu exploité jusqu'à présent, est l'utilisation et l'ingénierie d'enzymes post-PKS qui pourraient être utilisées avec l'objectif d'introduire des modifications extensives sur les polycétides. Dans ce but, des études structure/fonction de ces enzymes sont nécessaires.

III.E. Les enzymes post-PKS, une cible de choix pour l'ingénierie ?

Généralement, suite à la synthèse du polycétide par la PKS, celui-ci est mûré par des enzymes qui le modifient chimiquement et lui confèrent sa structure finale et son activité biologique. Dans le cas de l'érythromycine A, cinq enzymes post-PKS interviennent de manière séquentielle pour modifier le polycétide, le 6-dEB, et lui conférer son activité (Figure 24 page 42) : EryF et EryK hydroxylent les carbones 6 et 12 respectivement, EryBV est une mycarosyltransférase assurant le transfert de ce sucre en position 3, EryCIII est une désosaminyltransférase agissant sur le carbone 5, et EryG, une O-méthyltransférase, modifie le désosamine.

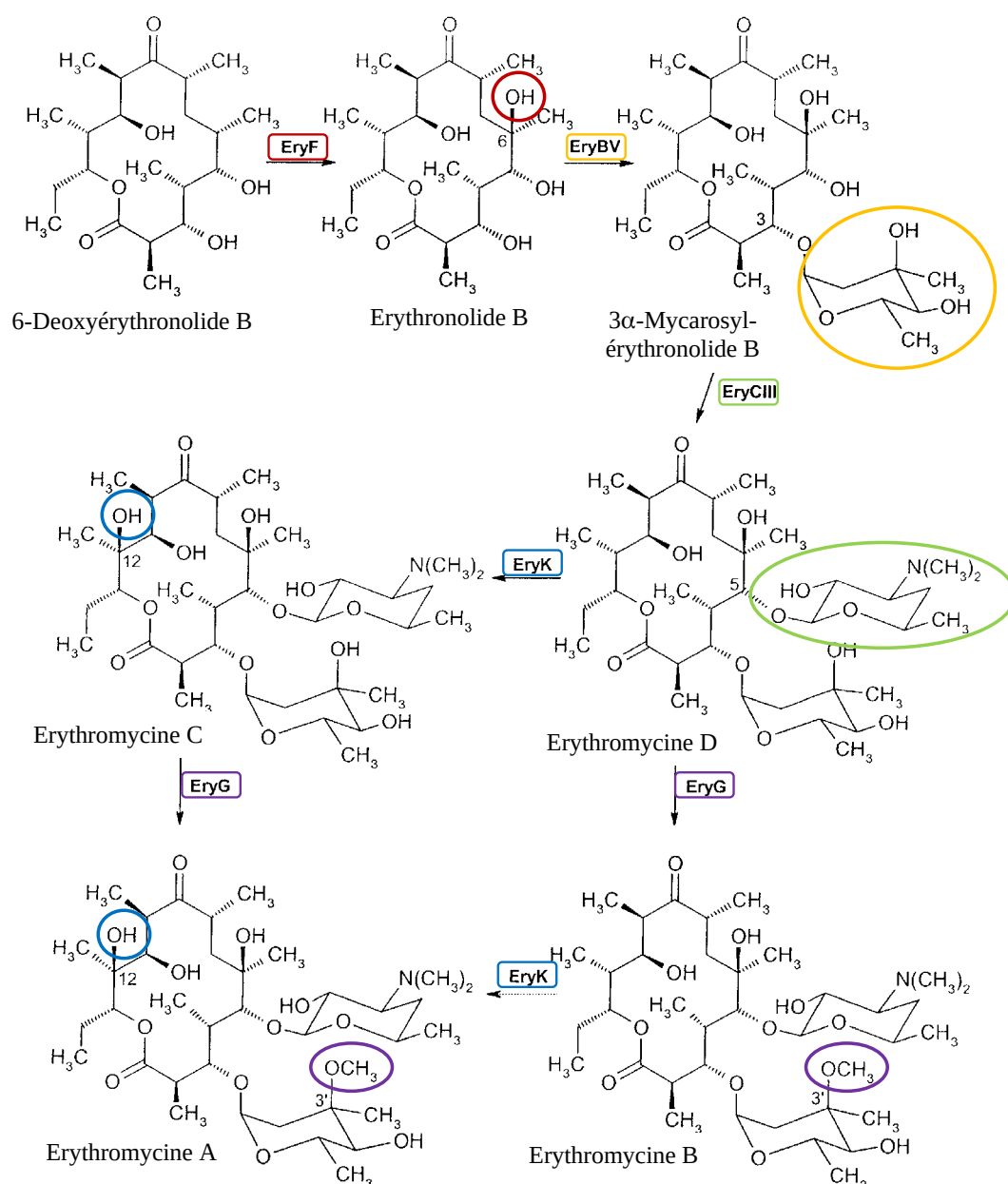


Figure 24 : Etapes de modification de l'intermédiaire 6-dEB par les enzymes post-PKS aboutissant à l'érythromycine A (adaptée de Rix et al., 2002).

Les gènes codant ces enzymes post-PKS sont pour la plupart localisés au sein du *cluster* de gène codant la PKS. Au vu de leur importance cruciale pour la bio-activité des polycétides, 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).

Parmi ces enzymes se trouvent de nombreuses oxyréductases. Elles sont responsables de l'ajout de groupements hydroxyle, aldéhyde ou cétone 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 glycosyl-transférases (Rix et al., 2002). Dans cette catégorie, des cytochromes P450 mono-oxygénases sont particulièrement attractifs pour l'approche d'ingénierie, puisqu'ils agissent dans les ultimes étapes de la biosynthèse des polycétides. De plus, il existe des P450 de spécificité large, telle que 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 polycétides sont également glycosylés. L'ajout des unités osidiques confère leur activité à certaines molécules comme l'érythromycine A ou l'amphotéricine B, puisque ces sucres participent à l'interaction avec les cibles de ces polycétides (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 sept unités osidiques soient greffés au polycétide. L'ingénierie des glycosyl-transférases 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 se trouvent également diverses transférases : méthyltransférases, acyltransférases, carbamoyltransférases et aminotransférases, entre autres, ou bien encore des enzymes catalysant des réactions d'halogénéation, soit autant de cibles potentielles pour l'ingénierie et la génération de nouveaux polycétides (Olano, Méndez, et Salas, 2010).

Bien que les domaines TE libèrent généralement le polycétide en catalysant sa macrocyclisation, quelques cyclases sont impliquées dans la dernière étape de synthèse du polycétide. C'est notamment le cas de la lankacidine, libérée sous forme d'un précurseur linéaire par le domaine TE qui n'entraîne que la formation d'une lactone à six membres. C'est ensuite une amide oxydase, LkcE, qui est proposée pour catalyser la macrocyclisation du polycétide. Dans d'autres cas, plusieurs types d'oxygénases sont décrites pour intervenir dans la cyclisation du polycétide. Par exemple la cyclisation de la rifamycine B est réalisée par une hydroxylase catalysant la formation du noyau naphthalène, même si ici, l'enzyme agit entre la deuxième et la quatrième étape d'élongation (Xu, 2005). De même, un cytochrome P450 mono-

oxygénase 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 (Figure 25 page 44) (He, Müller, et Hertweck, 2004).

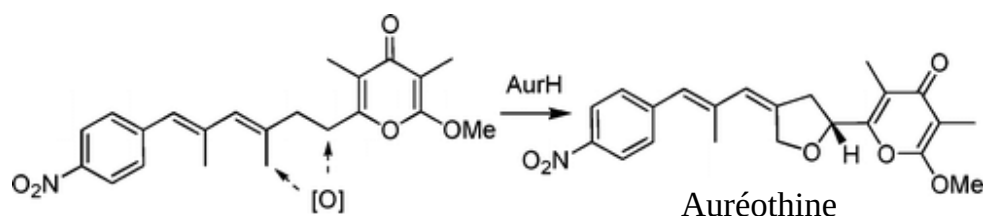


Figure 25 : Réaction de formation du cycle tétrahydrofurane catalysée par AurH (adaptée de He, Müller, et Hertweck, 2004).

Ces enzymes représentent donc un large vivier pour l'approche de biologie synthétique. Leur utilisation est pertinente étant donné qu'elles interviennent au stade final de la synthèse du polycétide. Une modification de ces enzymes n'entraîne pas de problème de spécificité de substrat, à l'exception des enzymes post-PKS intervenant en aval de la modification, comme cela peut être le cas avec les domaines en aval d'une modification de la PKS.

L'un des objectifs de ma thèse était de poursuivre le travail de Jonathan DORIVAL, un doctorant ayant soutenu sa thèse au sein de notre équipe, qui consistait à caractériser de LkcE, l'enzyme post-PKS catalysant la cyclisation de la lankacidine.

IV. Structures et dynamique des modules et des sous-unités

Grâce aux expériences d'ingénierie défailantes, il est devenu évident que les PKS modulaires sont extrêmement sensibles aux modifications introduites par les expériences d'ingénierie. Pour mieux comprendre les raisons de ces échecs, il est essentiel de mieux caractériser ces systèmes d'un point de vue structural et dynamique. En effet, il n'existe toujours pas de structure à haute résolution d'un module complet de PKS modulaire, que ce soit chez les *cis*-AT ou les *trans*-AT. La raison majeure de cette absence réside probablement dans la trop grande flexibilité entre domaines d'un même module. Ces informations structurales sont pourtant capitales afin de mieux comprendre comment les domaines au sein d'un module communiquent entre eux au cours du cycle d'élongation/réduction. Ces interactions spécifiques sont essentielles au bon acheminement de l'intermédiaire polycétidique d'un site actif en amont,

à l'autre en aval, sans être libéré dans la solution. Les blocs de construction sont transférés de l'AT à l'ACP et l'intermédiaire réactionnel de l'ACP vers la KS. Ce dernier est ensuite présenté aux sites actifs des différents domaines optionnels par l'intermédiaire de l'ACP dans un ordre bien précis. L'intermédiaire est par la suite transféré d'un module à l'autre et d'une sous-unité à l'autre. Cette stratégie biosynthétique offre l'avantage d'améliorer la cinétique de la voie de synthèse en évitant que les intermédiaires ne diffusent à distance des sous-unités, de les protéger de réactions non productives et d'empêcher leur entrée dans d'autres voies enzymatiques en compétition (Wu et al., 2001). De par le nombre élevé de réactions séquentielles qui ont lieu au sein des sous-unités, cette stratégie est particulièrement intéressante pour les PKS. Dans un premier temps, le modèle d'organisation structurale des modules s'est basé sur des structures cristallographiques de la FAS porcine (mFAS).

IV.A. La mFAS comme modèle structural d'un module de PKS modulaire

En raison des fortes homologies de séquences et de mécanismes synthétiques très similaires, les chercheurs se sont longtemps basés sur la structure cristallographique de la mFAS (Figure 26 page 46) (Maier, Leibundgut, et Ban, 2008) pour proposer des modèles d'organisation de modules de PKS de type I comprenant les domaines KR, DH et ER.

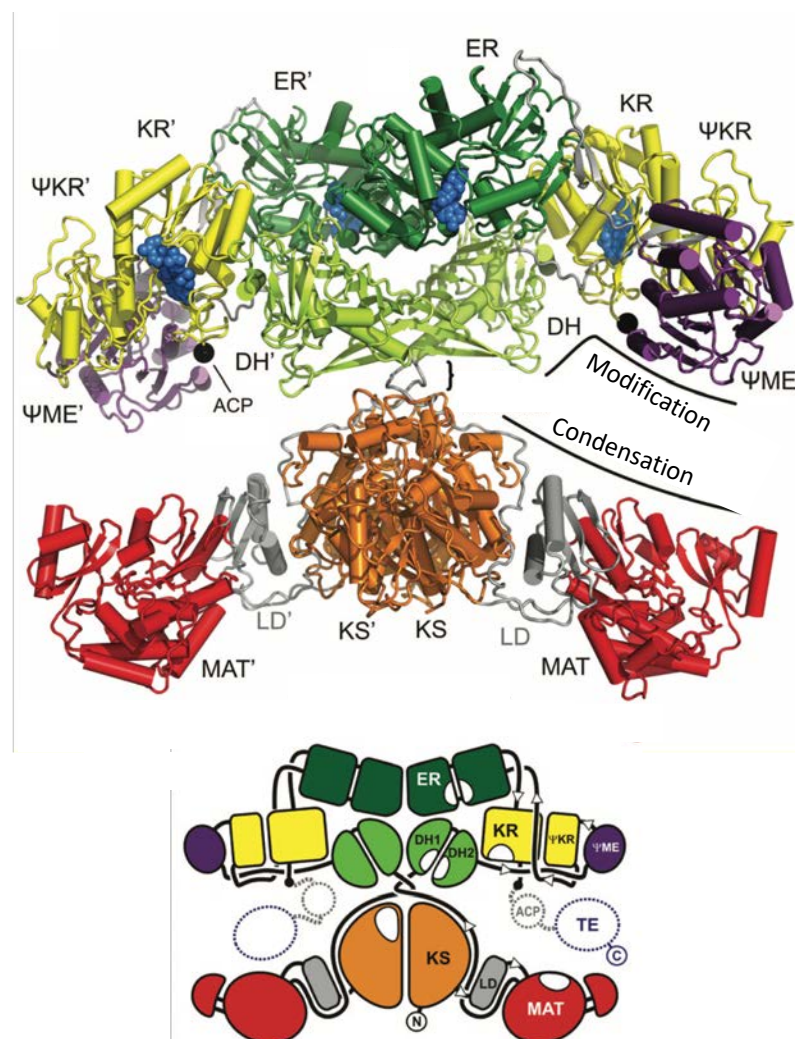


Figure 26 : Représentation cartoon de la structure cristallographique de la mFAS, obtenue à 3,2 Å de résolution et représentation schématique de la protéine (tirée de Maier, Leibundgut, et Ban, 2008).

Les domaines sont représentés par un code couleur, et les linkers séparant les domaines sont indiqués en gris. Les deux portions (condensation et modification) et le linker central séparant ces portions sont indiqués sur la structure, de même que la position du dernier résidu du linker entre la KR et l'ACP. 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 domaine malonyl/acétyl-transférase (MAT).

Globalement la mFAS adopte une structure dimérique entrecroisée en forme de X, via interactions des domaines KS, DH et ER qui sont homodimériques. Deux chambres réactionnelles sont présentes et elles accueilleraient les domaines ACP et TE, ces deux domaines ne sont pas visibles sur la structure, ceci étant probablement dû à leur trop grande flexibilité. Deux portions peuvent être distinguées : la première assure la condensation avec les domaines KS et malonyl/acétyl-transférase (MAT) et la seconde, la modification des intermédiaires avec les domaines KR, DH et ER (Figure 26 page 46). D'après les distances séparant les différents sites actifs, il est pensé que l'ACP est hautement mobile, passant d'un côté à l'autre de la chambre réactionnelle afin d'atteindre les différents sites actifs ; cela peut

également expliquer pourquoi ce domaine est absent de la densité électronique. Il est montré par cryo-microscopie électronique (cryo-EM) que la portion de condensation est capable d'une rotation de 180° par rapport à la portion réductrice (Brignole, Smith, et Asturias, 2009). Ces mouvements sont dus à la présence d'un *linker* d'une vingtaine de résidus non structurés présent entre les domaines MAT et DH, et rendraient asynchrone la synthèse dans les deux chambres. L'élongation a lieu dans la chambre réactionnelle relativement close, alors que la réduction a lieu dans celle qui est plus ouverte. De plus, la distribution des conformations pourrait être influencée par la présence du substrat ou par l'inactivation de certains domaines catalytiques, ce qui suggère une corrélation directe entre la configuration du module et l'étape de biosynthèse.

D'autres domaines non actifs se situent aussi dans la portion réductrice de la mFAS. Le premier est un domaine pseudo cétoréductase (Ψ KR), analogue au domaine KR mais dont le *Rossmann fold* est tronqué, et qui assure une pseudo-homodimérisation. Le second est un domaine pseudo-méthyltransférase (Ψ MT), dont le site de fixation au cofacteur SAM est muté, rendant sa prise en charge impossible.

Quant aux PKS, des études préliminaires de protéolyse ménagée couplée au *cross-linking* réalisées sur des sous-unités de DEBS, ont permis d'émettre l'hypothèse de modules et de sous-unités homodimériques (Staunton et al., 1996). Par la suite des structures cristallographiques et RMN de tous les domaines d'un module ont été obtenues, soit sous la forme de domaines isolés, soit sous la forme de didomaines (Weissman, 2017) (Figure 27 page 48).

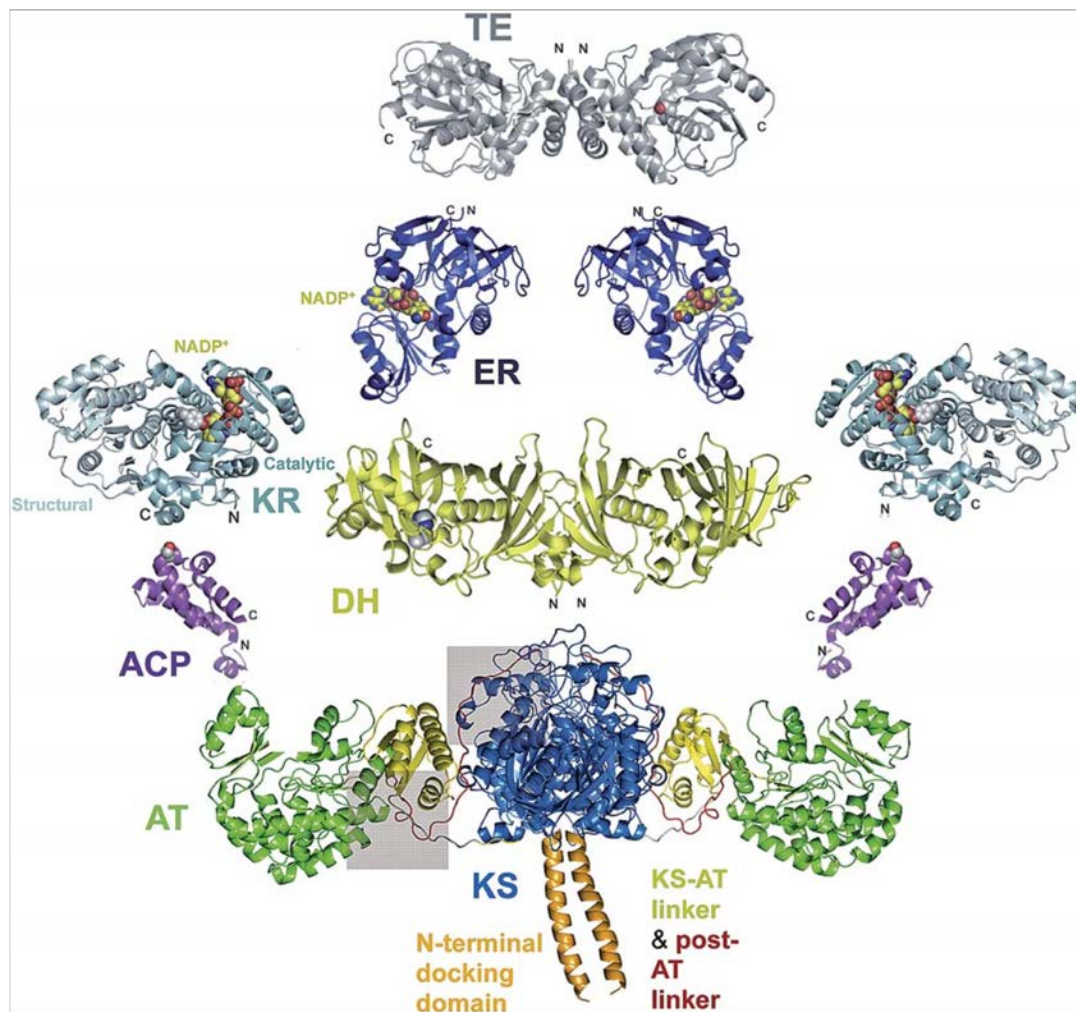


Figure 27 : Structures haute résolution de domaines individuels et de didomaines de PKS modulaires (tirée de Weissman, 2017).

Sont présentés : la structure cristallographique du didomaine KS_5-AT_5 de DEBS (orange : domaine de docking N-terminal, bleu : KS, vert : AT, jaune : linker de la KS vers l'AT, rouge : linker post-AT ; la structure RMN de l'ACP₂ de DEBS, la sérine phosphopanthéthéinylée est présentée en sphères ; la structure cristallographique de la DH₄ de DEBS, l'histidine catalytique est présentée en sphères ; la structure cristallographique de la KR₂ de l'amphotéricine synthase, de type A1, les domaines catalytiques et structuraux sont indiqués, la Tyr catalytique et le cofacteur NADP⁺ sont présentés en sphère ; la structure cristallographique du domaine ER monomérique de la spinosyne synthase, le cofacteur NADP⁺ est présenté en sphères ; la structure cristallographique du domaine TE homodimérique de DEBS.

Ces structures permettent de proposer une organisation homodimérique des PKS *cis*-AT, puisque les domaines KS, DH et TE sont homodimériques. De manière globale, ces domaines isolés présentent des structures très similaires à leur équivalent chez la mFAS (Keatinge-Clay et Stroud, 2006 ; Tang et al., 2006 ; Tang et al., 2007 ; Alekseyev et al., 2007). Par exemple, la disposition du didomaine KS-AT est identique entre les deux systèmes, la position de chaque domaine étant fixée par rapport à l'autre grâce à la présence d'un *linker*

structuré reliant la KS à l'AT. Sur cette base, plusieurs modèles d'organisation structurale des modules de PKS ont été proposés (Figure 28 page 49).

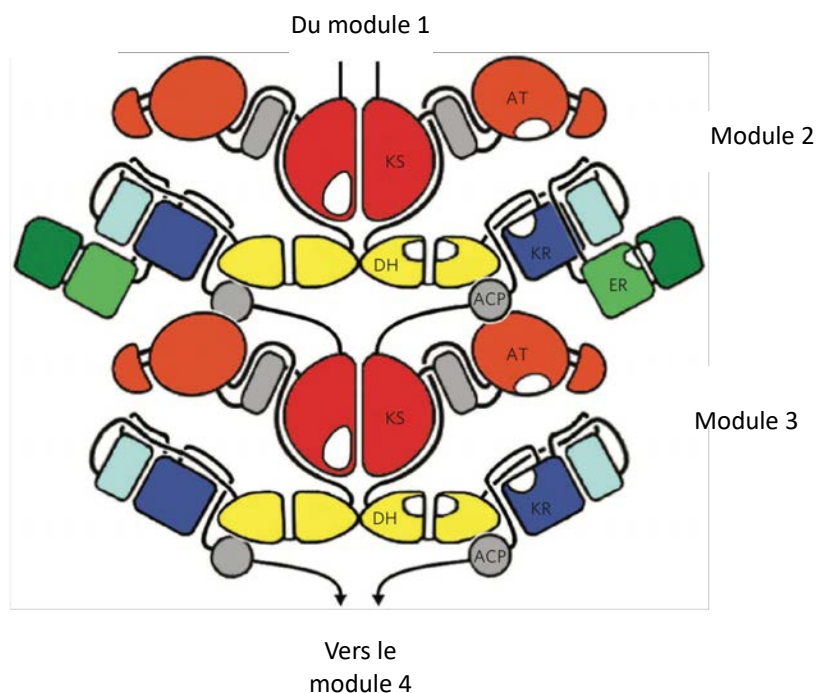


Figure 28 : Proposition de structuration d'une construction multi-modulaire de PKS (tirée de Zheng et al., 2012).

Ce modèle est proposé pour les modules de la mycolactone synthase, incluant des domaines KR et DH (module 3) ou tous les domaines de modification (module 2).

Cependant des divergences notables émergent au cours de ces études : en effet chez les PKS modulaires, les domaines ER sont monomériques et les domaines TE homodimériques (Zheng et al., 2012 ; Tsai et al., 2001), l'interface entre les domaines ER et DH sont très différentes, et les domaines DH ne s'homodimérisent pas de la même manière (Keatinge-Clay, 2008). Ces différences non négligeables ont posé alors la question de la validité du modèle de la structuration modulaire des PKS basée sur la mFAS.

IV.B. Structure de modules de PKS à moyenne résolution

IV.B.1. Structure d'un module entier par cryo-EM

La structure cryo-EM de la sous-unité PikAIII de la PKS synthétisant la pikromycine chez *S. venezuelae* est obtenue en 2014. Cette sous-unité ne comporte qu'un seul module, le module 5. Il est composé des domaines de condensation : KS, AT et ACP, ainsi que d'un domaine KR et de deux interfaces de *docking* assurant l'interaction en *trans* avec les sous-unités partenaires.

Plusieurs structures correspondant à différents stades de la catalyse ont été obtenues entre 7 et 11 Å de résolution. Les structures à haute résolution de domaines homologues ont été remplacées dans les cartes obtenues. La première étude (Dutta et al., 2014) a été réalisée alors que l'ACP du module était sous forme *holo*, c'est-à-dire modifié par le bras phosphopantéthéine (Figure 29 page 50).

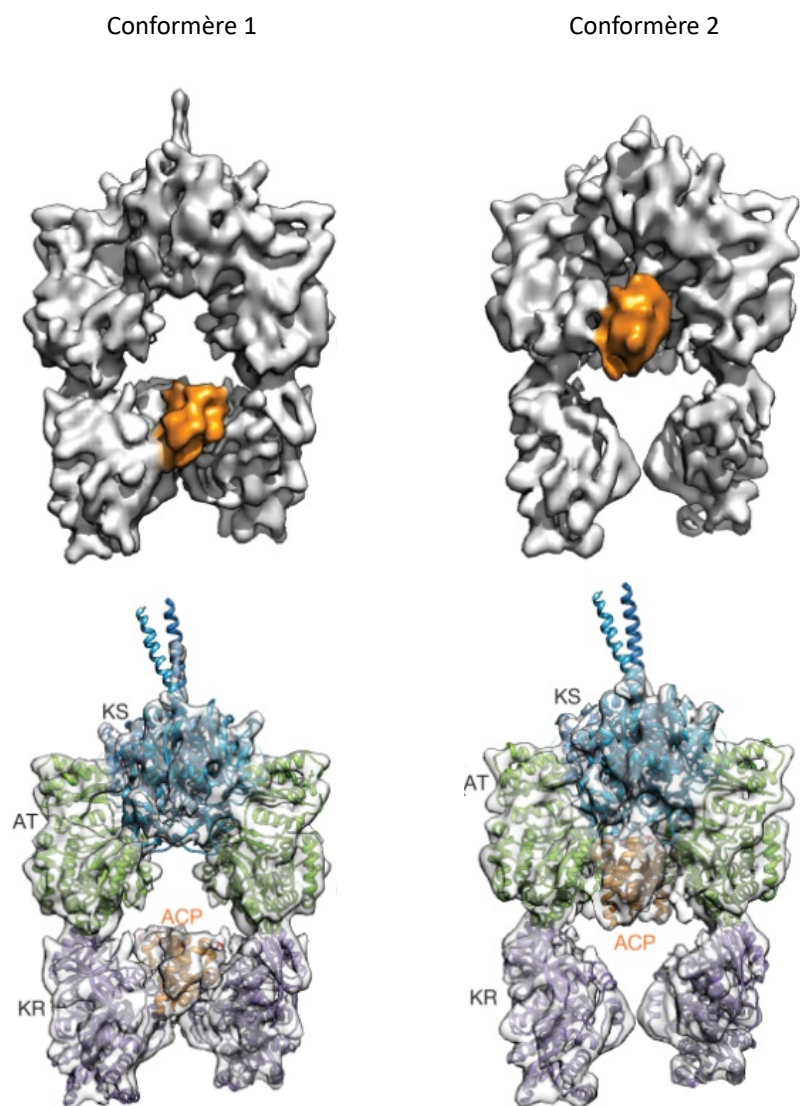


Figure 29 : Structures cryo-EM de PikAIII sous forme holo (tirée de Dutta et al., 2014).
Représentation de la structure cryo-EM de PikAIII en solide ou en transparent avec les structures à haute résolution positionnées au sein de la carte EM. L'ensemble s'agence en une forme d'arche, le domaine KS homodimérique 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).

Contrairement à ce qui était pensé jusqu'alors, ce module présente une forme d'arche symétrique ne comportant donc qu'une unique chambre réactionnelle, les sites actifs des domaines catalytiques faisant face à cette chambre. Les domaines AT et KR sont positionnés à

la suite, en dessous de la KS homodimérique. Cette disposition entre les domaines KS et AT est totalement différente de celle obtenue dans les structures cristallographiques de didomains issus de DEBS (Figure 27 page 48) (Tang et al., 2006 ; Tang et al., 2007). Les domaines ACP sont clairement visibles, et occupent deux positions différentes : proches du domaine KR (conformère 1) ou proches du domaine AT (conformère 2). De plus, ils occupent ces positions équivalentes simultanément, ce qui suggère qu'ils se déplacent ensemble au sein de l'arche, contrairement à ce qui est supposé pour les ACP de la mFAS (Brignole, Smith, et Asturias, 2009).

Les auteurs se sont aussi intéressés aux changements conformationnels du module nécessaires à la réalisation du cycle d'élongation/réduction (Whicher et al., 2014). Pour cela, des domaines ou des combinaisons de domaines ont été tour à tour modifiés de manière à mimer chaque étape du cycle catalytique. Des cartes des modules ainsi bloqués ont été ensuite obtenues par cryo-EM (Figure 30 page 52).

La première étape du cycle d'élongation/réduction de PikAIII consiste à charger la KS₅ par l'ACP du module précédent, l'ACP₄. Pour comprendre comment la KS₅ discrimine entre l'ACP₄, avec lequel elle interagit en *trans*, de l'ACP₅, avec lequel elle interagit en *cis*, une protéine de fusion ACP₄-PikAIII(Δ ACP₅) a été analysée. Cette protéine ne contient pas l'ACP₅ qui pourrait interférer dans l'interaction entre l'ACP₄, qui lui était modifié par un substrat pentacétidique, son ligand naturel. De manière à bloquer le transfert de ce pentacétide, la cystéine catalytique de la KS₅ a été mutée. Les cartes cryo-EM obtenues ont révélées que l'ACP₄ interagit avec la KS₅ à l'entrée de son site actif, dans une configuration telle qu'il demeure à l'extérieur de l'arche formée par le module (conformation 1-2, Figure 30 page 52). De cette manière, la KS serait en mesure de discriminer entre les ACP et d'éviter les mauvais transferts. En l'absence du pentacétide fixé à l'ACP₄, le domaine *holo* ne contacte pas la KS₅, ce qui suggère que la présence du substrat est nécessaire à l'établissement d'une interaction stable. Pour mimer l'étape suivante, la KS₅ a été acylée spécifiquement par l'intermédiaire pentacétidique, et l'ACP₅ maintenu sous forme *apo*. Cet état serait loin d'être révélateur du phénomène physiologique, puisque de nombreuses données démontrent que les domaines ACP et AT sont constamment chargés par des unités d'extension *in vivo* (Dunn et Khosla, 2013 ; Hong, Leadlay, et Staunton, 2009). En effet la structure rapportée montre l'ACP₅ en complexe avec le domaine AT₅, sa sérine catalytique pointée vers le site actif de l'AT₅ (conformation 2 Figure 30 page 52), ce qui pourrait correspondre à une réponse de détresse du module favorisant le chargement rapide de l'unité d'extension. De plus, il est étonnant que la KR₅ procède à une rotation de presque 180° par rapport à la conformation précédente.

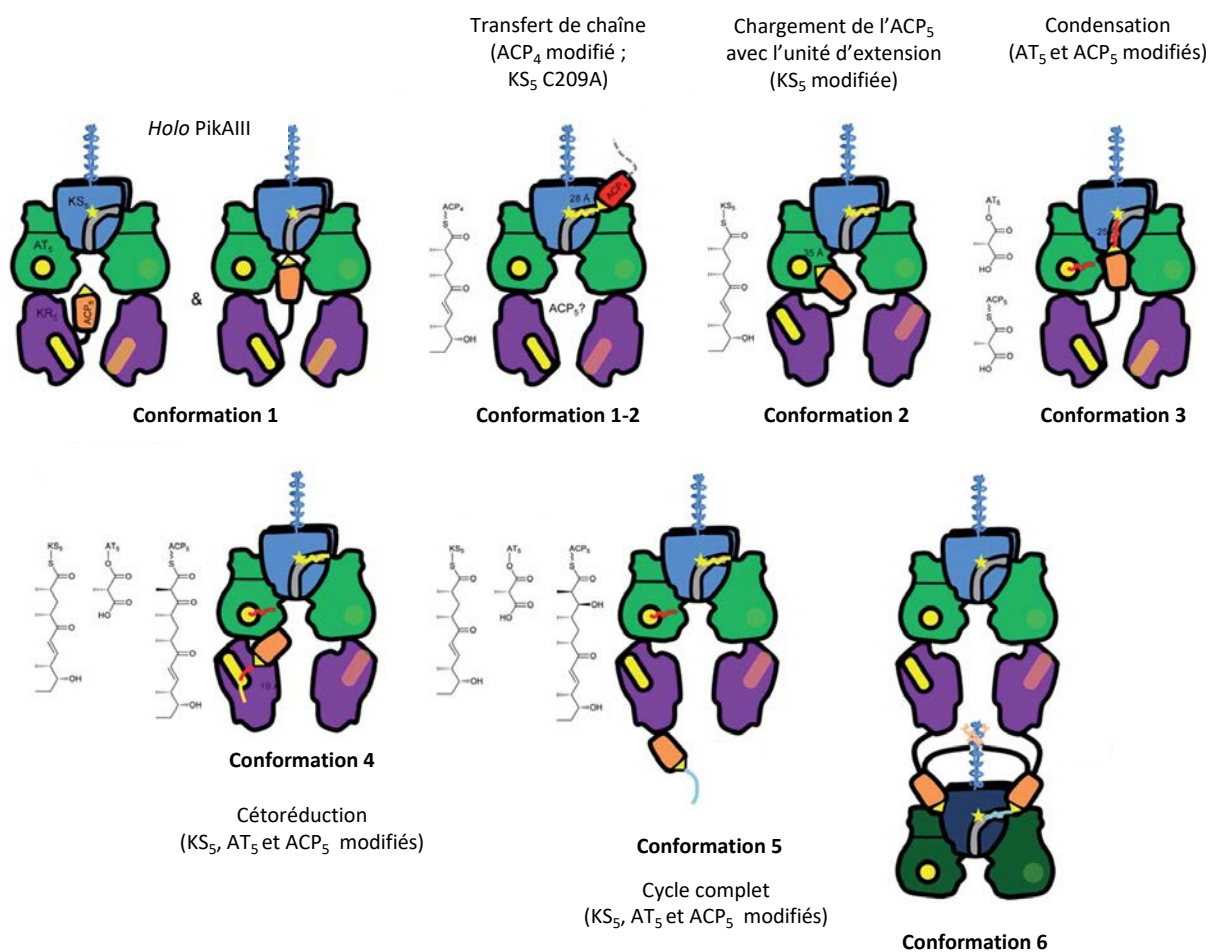


Figure 30 : Conformations adoptées par PikAIII au cours du cycle d'élongation/réduction (adaptée de Whicher et al., 2014).

Le domaine KS₅ homodimérique est représenté en bleu (sa cystéine catalytique est symbolisée par une étoile), l'AT₅ en vert, la KR₅ en violet, l'ACP₄ en rouge et l'ACP₅ en orange. Par soucis de clarté, seul un ACP₅ est représenté.

Ensuite les auteurs ont assuré le chargement de l'ACP₅ avec l'unité d'extension, le méthylmalonate, grâce à l'activité naturelle de l'AT₅. Dans la structure résultante, l'ACP₅ *holo* pointe le substrat vers le site actif de la KS₅ (conformation 3 Figure 30 page 52) et son site d'interaction est à l'opposé de celui de l'ACP₄. Au vu des surfaces d'interaction de ces deux ACP, aucune gêne stérique ne serait engendrée par la présence de l'un ou l'autre, rendant possible une interaction simultanée des deux domaines sur la KS₅. Ceci pourrait expliquer les transferts directs ACP-ACP observés dans des systèmes PKS modifiés, résultant en un « évitement » de module (Figure 22 page 37) (Thomas et al., 2002).

L'étape suivante de la catalyse étant la condensation du méthylmalonate et du pentacétide, les auteurs ont laissé incuber l'ACP₅ *holo* avec la KS₅ modifiée au niveau de sa cystéine catalytique par le pentacétide. Ceci a mené à l'obtention d'un ACP₅ chargé par un β-cétohexacétide (conformation 4 Figure 30 page 52). Dans cette configuration, l'ACP₅ interagit

avec la KR₅ qui a subi une rotation d'environ 180° par rapport au conformère 3, positionnant son site actif à portée du bras phosphopanthéine.

En incubant la construction précédente avec du NADPH, les auteurs ont obtenu le produit final synthétisé par le module. Une fois la cétone réduite par le domaine KR₅, l'ACP₅ se place directement sous la KR₅ afin de faciliter l'échange de la chaîne en élongation avec le module suivant (conformations 5 et 6 Figure 30 page 52).

Cette étude permet une contribution majeure à la compréhension des interactions entre le domaine ACP et ses différents partenaires. L'ACP d'un module sonderait de manière aléatoire un partenaire potentiel *via* des interactions charge-charge complémentaires. Le partenaire spécifique serait sélectionné lorsque son domaine catalytique reconnaîtrait son substrat. Ces interactions supplémentaires permettraient la stabilisation du complexe qui ne serait que transitoire dans le cas contraire (Weissman, 2017).

IV.B.2. Structures d'une construction modulaire et bimodulaire par SAXS

Deux multi-enzymes ont été analysées au cours de cette étude : la première comprend le module 3 de DEBS fusionné au domaine TE du même système (M3-TE) et la seconde correspond à la sous-unité DEBS 3 comprenant les modules 5 et 6, ainsi que le domaine TE. Tous les modules étudiés comprennent les mêmes domaines : KS, AT, KR et ACP, mais diffèrent selon la construction : l'ACP est suivi soit du domaine TE pour les modules 3 et 6, soit d'une KS homodimérique pour le module 5 (Edwards et al., 2014).

Dans un premier temps, des analyses SAXS ont été menées sur des fragments mono ou didomains du M3-TE (KS-AT, KR, ACP et TE) pour lesquels des structures à haute résolution sont connues. Puis des courbes de diffusion ont été enregistrées sur la construction M3-TE complète. Par modélisation en corps rigides, trois modèles en adéquation avec les données SAXS ont été proposés, suggérant l'existence de plusieurs états conformationnels en solution (Figure 31 page 54).

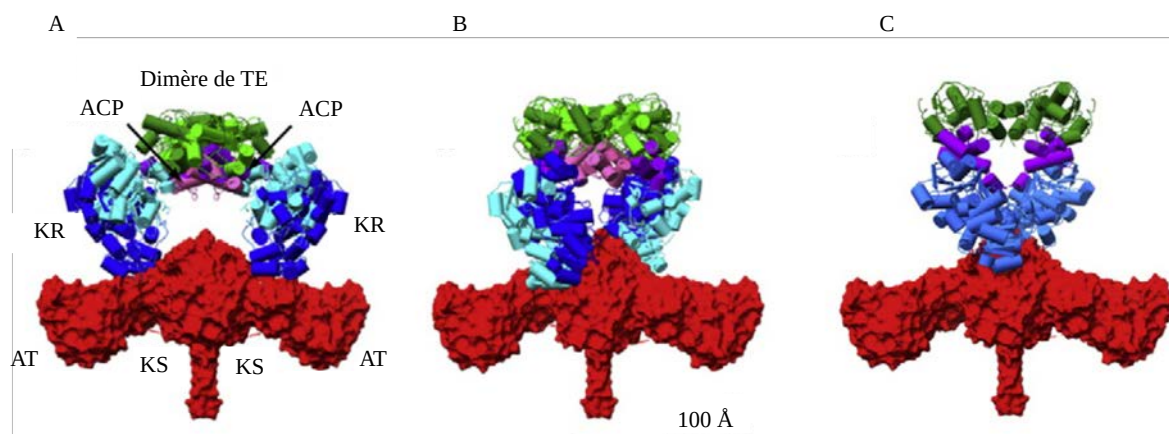


Figure 31 : Modélisations en corps rigides de la construction M3-TE (adaptée de Edwards et al., 2014). Le didomaine KS-AT (en rouge) et le pseudo dimère de KR (en bleu) ont été traités comme des corps rigides durant la modélisation. Les domaines ACP (en violet) sont localisés à proximité du dimère de TE (en vert), quel que soit le modèle généré.

Évidemment, l'organisation générale de ce module est différente de celle du module PikAIII obtenue par cryo-EM. Tout d'abord le didomaine KS-AT présente une conformation étendue ; ceci est certainement dû au fait qu'il est considéré comme un corps rigide durant la modélisation, ainsi la structure de cette région du module est la même que celle du didomaine isolé (Tang et al., 2006 ; Tang et al., 2007). De plus, les domaines ACP sont localisés proches des domaines TE et donc à distance trop importante des sites actifs des domaines KR, KS et AT. Ces divergences d'organisation spatiale peuvent être expliquées par une troncature des données de diffusion à des q faibles ($0,15 \text{ \AA}^{-1}$), ce qui limite la résolution à $40 \text{ \AA}$.

Les données enregistrées sur la construction correspondant à DEBS 3 sont en accord avec une protéine allongée et homodimérique. De manière analogue à la première construction, une modélisation en corps rigides a été menée. Les auteurs ont disposé ainsi les domaines les uns par rapports aux autres en accord avec les données. Globalement, l'organisation des domaines au sein des modules 5 et 6 est la même que celle des domaines de M3-TE (Figure 32 page 55), mais les deux modules, bien que liés covalamment, seraient capables d'une rotation l'un par rapport à l'autre allant jusqu'à 70° le long d'un plan passant par le didomaine KS-AT et les domaines ACP.

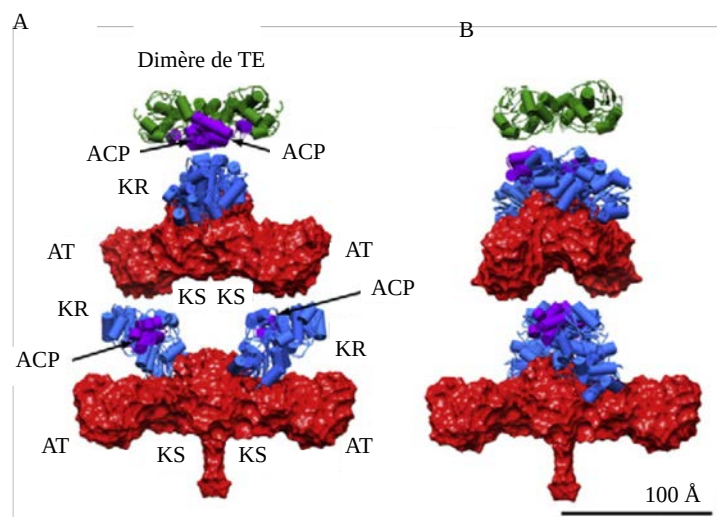


Figure 32 : Modélisations en corps rigides de la construction DEBS 3 (tirée de Edwards et al., 2014). Le didomaine KS-AT (en rouge), les pseudo dimères de KR (en bleu), les ACP (en violet) et le dimère de TE, ont été traités comme des corps rigides.

Ensemble, ces études structurales permettent de mieux appréhender les interactions spécifiques entre domaines et le transfert de l'intermédiaire d'un site actif à l'autre au sein d'un même module grâce aux domaines ACP. Cependant, bien des questions restent en suspens. D'autres études doivent être menées sur des modules comprenant les autres domaines de modifications : DH et ER. A mon arrivée au laboratoire, aucune structure de modules de PKS *trans*-AT n'était disponible, or ces derniers présentent des compositions en domaines atypiques et très variables. Il était donc intéressant de déterminer l'impact d'une composition alternative en domaines sur l'architecture modulaire.

De plus, les précédents travaux (Dutta et al., 2014 ; Whicher et al., 2014) révèlent l'interdépendance et la grande communication entre domaines d'un même module au cours du cycle d'élongation/modification, ce qui peut expliquer en grande partie le faible taux de succès des expériences d'ingénierie. Il semble donc que l'approche la plus adaptée pour l'ingénierie serait d'échanger directement des modules voire des sous-unités entières de PKS de manière à garantir l'intégrité structurale des modules et une communication efficace entre domaines. Pour ce faire, pourtant, il est impératif de déterminer comment les sous-unités de PKS communiquent ensemble de manière spécifique et sélective afin d'assurer le bon transfert de l'intermédiaire réactionnel.

IV.C. Les domaines de docking

Les PKS sont composées de plusieurs sous-unités et les organismes synthétisant des polycétides présentent bien souvent plusieurs *clusters* de PKS différents au sein de leur génome (Katz et Baltz, 2016), or l'intermédiaire polycétidique ne doit pas être transféré à une sous-unité portant des modules antérieurs déjà intervenus dans la synthèse, ni à une sous-unité appartenant à un autre système PKS. Cette spécificité d'interaction est garantie par des petits éléments structuraux situés aux extrémités N-terminales (N_{ter}) et C-terminales (C_{ter}) des sous-unités et nommés domaines de *docking* (DD).

IV.C.1. Mise en évidence

Les DD ont été mis en évidence de manière indirecte dès 1992 par alignement des séquences des sous-unités de l'érythromycine synthase. Les auteurs ont découvert, sur DEBS 2 et DEBS 3, des séquences en N_{ter} des sous-unités qui n'étaient pas assignables aux domaines KS et en C_{ter} de DEBS 1 et DEBS 2, des séquences qui n'étaient pas attribuables aux ACP. Il a été alors émis comme hypothèse que ces séquences jouent un rôle dans l'assemblage et l'ordonnement des sous-unités (Donadio et Katz, 1992). Plus tard, des alignements de séquences entre plusieurs PKS *cis*-AT : l'érythromycine, la rapamycine et l'oléomycine synthase, ont confirmé la présence répandue de ces séquences en N_{ter} alors baptisées *linker* (Aparicio et al., 1996). Elles présentent la particularité de contenir des répétitions en heptades d'acides aminés aliphatiques et ont été prédites comme se structurant en hélices α . Il a été supposé que ces séquences puissent s'homodimériser en formant un *coiled-coil* et permettent d'initier la dimérisation des sous-unités multi-enzymatiques en partant de leur extrémité N_{ter} (Aparicio et al., 1996). D'ailleurs des constructions comprenant le module 2 et le domaine TE (M2-TE) ou le module 6 et le domaine TE (M6-TE) de DEBS, modules situés à l'extrémité C_{ter} de leur sous-unité, sont incapables de prendre en charge un substrat dicétidique. Au contraire, les constructions M3-TE et M5-TE sont capables de produire le tricétide à partir du même dicétide ; ce sont des modules situés en N_{ter} de leurs sous-unités respectives qui comprennent donc la séquence d'homodimérisation supposée. En effet, l'activité d'extension du M2-TE et du M6-TE a été restaurée en insérant la séquence de trente-neuf acides aminés situés en N_{ter} du module 5 (Gokhale, 1999).

Il a été très vite montré que ces *linkers*, rebaptisés DD par la suite, naturellement présents entre sous-unités sont d'un grand intérêt pour les expériences d'ingénierie. Par exemple, une PKS hybride a été créée par génie génétique, qui comprenait les trois sous-unités

de DEBS, mais le module 2 de DEBS 1 a été remplacé par le module 5 de la rifamycine synthase tout en conservant le DD en C_{ter} de DEBS 1. Le polycétide attendu a bien été synthétisé sans perte de rendements (Figure 33 page 57).

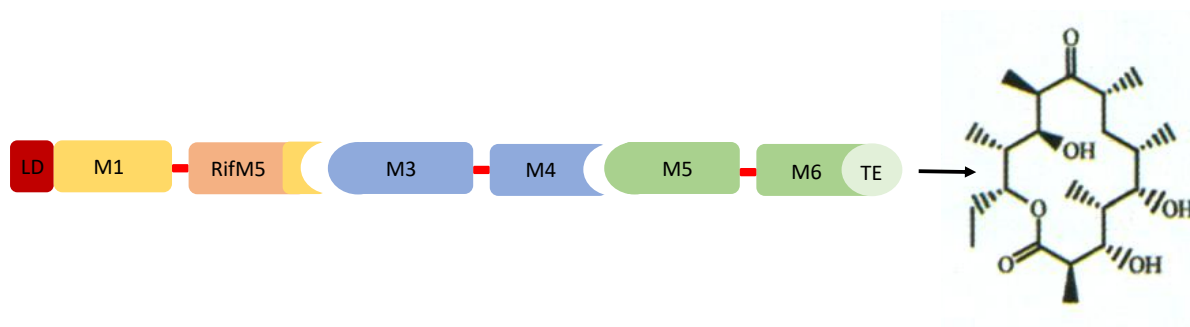


Figure 33 : Représentation schématique de la PKS hybride dans laquelle le module 2 de DEBS a été remplacé par le module 5 de la rifamycine synthase (adaptée de Gokhale, et al. 1999).

Le module 5 de la rifamycine synthase (en orange) a été inséré à la place du module 2, de manière à préserver les linkers entre modules d'une même sous-unité (en rouge), ainsi que les DD entre modules situés sur des sous-unités distinctes (en jaune, en C_{ter} du module 5 nouvellement inséré).

Par la suite, il a été démontré que ces DD facilitent en partie le transfert de la chaîne polycétidique en croissance d'une sous-unité à l'autre (Tsuji, Cane, et Khosla, 2001). En prenant DEBS pour modèle, le transfert de l'intermédiaire entre les modules 2 et 3, donc entre DEBS 1 et DEBS 2, a été suivi. Des paires non natives de DD ont été testées : celui en C_{ter} de DEBS 1 avec celui en N_{ter} de DEBS 3 ou celui en C_{ter} de DEBS 2 avec celui en N_{ter} de DEBS 2. Des transferts d'intermédiaires ont été observés, mais extrêmement ralentis par rapport aux constructions présentant des paires natives. En effet, lorsque les DD natifs présents entre DEBS 2 et DEBS 3 ont été utilisés entre DEBS 1 et DEBS 2, le transfert de l'intermédiaire a été réalisé à une vitesse comparable à celle de la construction sauvage.

Une étude similaire a été menée afin d'évaluer la contribution des domaines ACP_{n-1} et KS_n dans l'efficacité de transfert de l'intermédiaire (Wu, Cane, et Khosla, 2002). Les données obtenues au cours de cette étude ont montré que bien que les DD ne soient pas appariés, si l'ACP_{n-1} et la KS_n le sont, le transfert de l'intermédiaire est possible. Ces résultats sont confirmés par d'autres études qui tendent à prouver que l'ACP et la KS interagissent bien ensemble et que les DD ne sont pas les seuls domaines impliqués dans la reconnaissance entre sous-unités (Weissman, 2006a ; Buchholz et al., 2009 ; Whicher et al., 2014).

Ces expériences démontrent que même si les DD ne sont pas les seuls déterminants du bon transfert de l'intermédiaire, ils jouent un rôle capital dans les interactions spécifiques entre sous-unités. De plus ce sont des éléments très intéressants pour l'ingénierie puisqu'ils sont portatifs, c'est-à-dire qu'ils peuvent fonctionner de manière indépendante de leur sous-unité, et

l'insertion d'une paire native peut permettre la communication entre des sous-unités de différents systèmes (Tang, Fu, et McDaniel, 2000 ; Menzella, Carney, et Santi, 2007 ; Kim et al., 2017).

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., 2013). Ces interactions seraient aussi transitoires au sein de la cellule, ce qui faciliterait le renouvellement de sous-unités défectueuses en évitant une néo-synthèse totale d'une nouvelle PKS entière, ce qui, au vu de la taille totale de ces systèmes, serait extrêmement coûteux à la cellule d'un point de vue énergétique.

IV.C.2. Différentes classes de domaines de docking

La caractérisation structurale ainsi que des alignements de séquences plus poussés permettent de distinguer clairement plusieurs classes de DD.

IV.C.2.a. Domaines de docking de classe 1

IV.C.2.a.i. Classe 1a

La première classe mise en évidence chez les PKS *cis*-AT et plus particulièrement entre les sous-unités DEBS 2 et DEBS 3 (Broadhurst et al., 2003), a été la classe 1a. La structure des deux partenaires en complexe a été résolue par RMN. Pour faciliter l'étude et favoriser l'interaction plutôt faible entre les deux partenaires, les extrémités C_{ter} de DEBS 2 et N_{ter} de DEBS 3 ont été fusionnées par l'intermédiaire d'un *linker* artificiel de deux acides aminés. Il a été montré qu'une telle fusion de sous-unité n'altérerait en rien l'activité de la PKS *in vivo* et au contraire favorisait son repliement et sa stabilité (Liang, Sandberg, et Terwilliger, 1993 ; Squire et al., 2003).

La structure RMN présente un complexe homodimérique qui peut être divisé en deux domaines distincts, A et B, connectés entre eux par des boucles d'environ dix résidus, comprenant le *linker* artificiel (Figure 34 page 59). Le domaine A correspond aux hélices α 1 et 2 du DD situé en C_{ter} de DEBS2 (^CDD) ; en s'homodimérisant par des interactions hydrophobes, un fagot de quatre hélices α est formé.

Le domaine B comprend l'interface de *docking*. Comme prédit par des études précédentes, le DD situé en N_{ter} de DEBS 3 (^NDD), se replie en hélice α et s'homodimérise par des interactions hydrophobes formant ainsi un *coiled-coil*. Au niveau de ce *coiled-coil*, les

hélices $\alpha 3$ du C DD viennent s'insérer dans le sillon. L'ensemble forme un fagot de quatre hélices α , les DD interagissent ensemble *via* des interactions hydrophobes mais aussi des interactions électrostatiques.

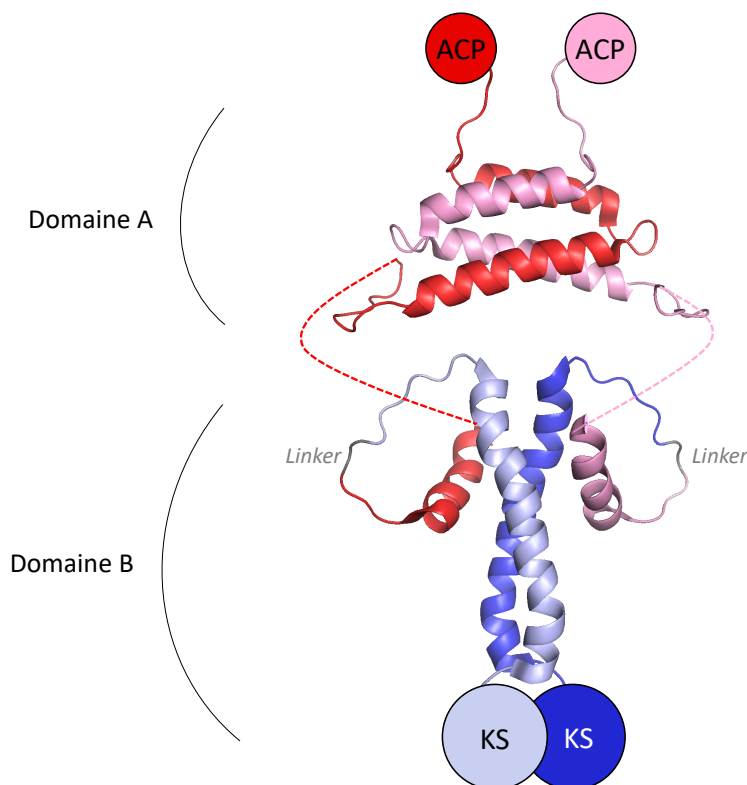


Figure 34 : Structures RMN des domaines de docking situés entre les sous-unités DEBS 2 et DEBS 3 (PDB : 1PZR et 1PZQ).

Le CDD, situé en aval du dernier ACP de la sous-unité DEBS 2 (en rouge et en rose), comprend trois hélices α . Les deux premières assurent l'homodimérisation du domaine A et la troisième interagit avec le coiled-coil formé par le NDD homodimérique (en bleu clair et en bleu foncé), situé en amont de la première KS de la sous-unité DEBS 3. Le linker artificiel permettant la fusion des deux partenaires est présenté en gris et les boucles reliant des domaines A et B en pointillés. Ces deux domaines ne semblent pas interagir ensemble, c'est pourquoi deux fichiers PDB ont été déposés (domaines A et B), ils sont ici reliés par les lignes en pointillés.

Cette classe de DD est largement répandue au sein des PKS *cis*-AT, et peut être présente plusieurs fois au sein du même système. C'est le cas, par exemple, chez DEBS : les DD de l'interface DEBS 1/DEBS 2 en feraient aussi partie, mais aussi chez l'amphotéricine synthase qui est composée de six sous-unités, où des DD de cette classe seraient présents à chaque interface. Des études ont donc été menées afin de préciser les déterminants de la spécificité d'interaction au sein des DD de classe 1a. Par alignements de séquences, des résidus hydrophobes et chargés situés à des positions précises de l'hélice $\alpha 3$ du C DD et de l'hélice α du N DD, ont été mis en évidence. Ces résidus seraient à la base d'un « code de spécificité » : les résidus chargés, grâce à des interactions charge-charge compatibles, et les résidus

hydrophobes *via* une surface d'interaction compatible (Weissman, 2006b). Des études bio-informatiques plus poussées ont même permis de distinguer des classes phylogéniques mutuellement exclusives (Thattai, Burak, et Shraiman, 2007).

IV.C.2.a.ii. Classe 1b

Cette classe a été identifiée pour la première fois par des études des DD situés entre les sous-unités 3 et 4 de la pikromycine synthase (Buchholz et al., 2009). Dans un premier temps, les auteurs ont tenté d'obtenir des cristaux d'une protéine de fusion comprenant les trois hélices α prédites du C DD et l'hélice α du N DD. Cependant, aucun cristal n'a été obtenu, ceci étant certainement dû à la présence de la boucle flexible entre les hélices $\alpha 2$ et 3 du C DD. Une construction tronquée a été donc produite, qui ne comprenait que les hélices α impliquées dans le *docking* proprement dit, fusionnées directement sans *linker*. La structure cristallographique obtenue à 1,75 Å de résolution est très proche de celle des DD de classe 1a (Figure 35 page 60). L'absence de *linker* ne semble pas fournir une flexibilité suffisante pour que les DD d'une même chaîne polypeptidique interagissent ensemble ; la structure obtenue est telle que le C DD d'une protéine de fusion interagit avec le N DD d'une seconde protéine de fusion.

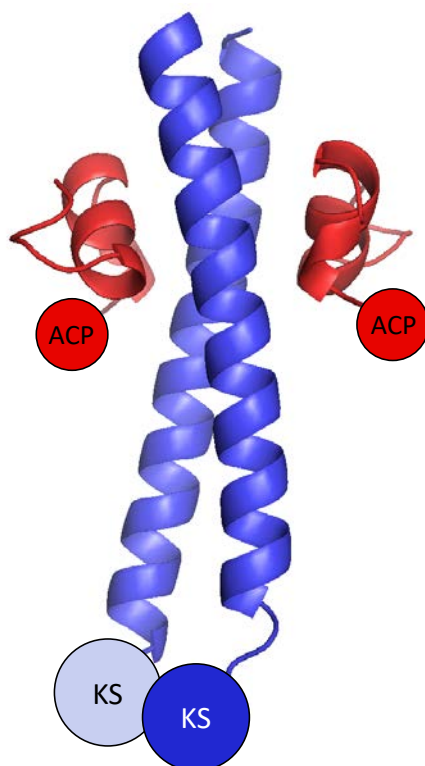


Figure 35 : Structure cristallographique à 1,75 Å de résolution de l'interface de docking entre les sous-unités PikAIII et PikAIV (PDB : 3F5H).

Tout comme dans la classe 1a, le ^NDD comporte une hélice α unique qui s'homodimérise en formant un *coiled-coil*, et l'hélice α terminale du ^CDD s'insère dans son sillon. Quelques détails permettent cependant de les différencier de la classe 1a et rendent ces deux classes mutuellement exclusives : tout d'abord l'hélice α_3 du ^CDD est bien plus courte dans la classe 1b (neuf résidus contre quinze), et ensuite, une courte hélice α supplémentaire est présente, qui se place dans la jonction du *coiled-coil*. De plus, la position des résidus chargés et des résidus hydrophobes intervenant dans l'interaction et sa spécificité ne sont pas les mêmes.

IV.C.2.b. Domaines de docking de classe 2

Les domaines de docking de classe 1 ont été mis en évidence chez les actinobactéries : *Saccharopolyspora erythraea* pour DEBS et *Streptomyces venezuelae* pour Pik. L'alignement de séquences de DD de systèmes PKS trouvés d'une part chez des Actinobactéries, et d'autre part chez des cyanobactéries et des myxobactéries, montre que ces deux clades se distinguent dans leur organisation structurale (Whicher et al., 2013) : les ^CDD sont plus courts de quarante résidus par rapport à la classe 1 et ne contiendraient donc pas la région d'homodimérisation ; quant aux ^NDD, ils sont de taille similaire à la classe 1 mais des résidus polaires sont présents en N_{ter} du domaine, et ne pourraient donc pas former un seul long *coiled-coil*. Les auteurs se sont intéressés à deux interfaces de la curacine A synthase, trouvée chez une cyanobactérie, *Moorea producens*. Les structures de paires de DD à ces deux interfaces ont été résolues par cristallographie sur des protéines de fusion : les deux partenaires ont été liés par un *linker* de huit résidus (Gly₃Ser)₂. (Figure 36 page 62).

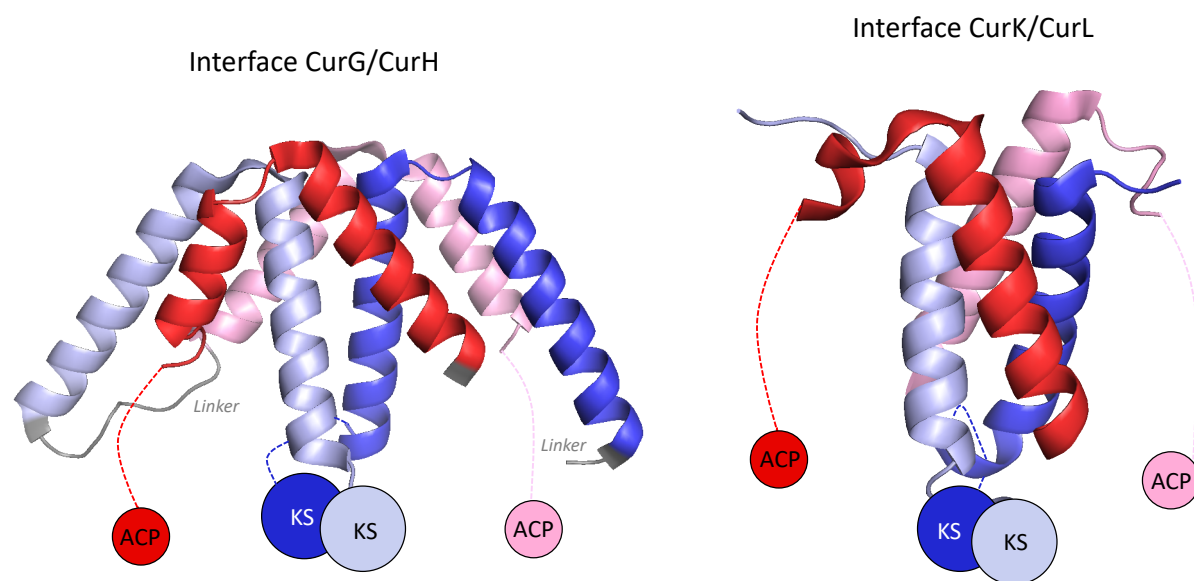


Figure 36 : Structures cristallographiques de deux interfaces de docking trouvées chez la curacine A synthase (PDB : 4MYY, 4MYZ).

La structure des DD de l'interface CurG/CurH a été résolue à 1,7 Å de résolution. Les C^oDD (en rouge et en rose) sont monomériques et constitués de deux hélices α . Les N^oDD, en bleu (foncé et clair), sont homodimériques, et comprennent aussi deux hélices α . Un coiled-coil est formé par les deux hélices en amont des KS. La structure des DD de l'interface CurK/CurL a été résolue à 1,5 Å de résolution. Les C^oDD (en rouge) sont aussi monomériques et constitués d'une hélice α , connectée par une boucle flexible à un tour d'hélice. Les N^oDD (en bleu) sont homodimériques en formant un coiled-coil et comprennent une hélice α unique.

Le couple de DD trouvé à l'interface CurG/CurH forme un fagot de huit hélices α . Chaque domaine est constitué de deux hélices α reliées par une boucle flexible. Au contraire des DD de classe 1, les C^oDD ne disposent pas du fagot de dimérisation et sont monomériques. Les N^oDD sont homodimériques, car les hélice α (bleues) en amont des KS forment un coiled-coil. La première hélice α (rouge foncé) des C^oDD interagit avec les deux hélice α (bleu clair) d'un monomère de N^oDD, tandis que la seconde (rouge foncé) contacte le coiled-coil (bleu clair et foncé).

L'interface entre les sous-unités CurK et CurL présente la même topologie. Il faut cependant noter que les N^oDD ne comportent qu'une seule hélice α et que le C^oDD est constitué d'une hélice α suivie d'un tour d'hélice, mais l'angle formé entre ces deux structures est le même que celui trouvé chez le C^oDD de CurG. De la même manière, la boucle entre les hélices α des monomères de N^oDD de CurH et celle trouvée en amont des monomères de N^oDD de CurL adoptent la même conformation.

La topologie de ces deux complexes permet de positionner directement les ACP aux côtés des KS alors que chez les DD de classe 1, les boucles reliant les C^oDD à l'ACP sont à

l'opposé. De manière intéressante, la spécificité d'interaction chez les DD de classe 2 semble aussi dictée par la présence d'une large surface hydrophobe complémentaire ainsi que des interactions charge-charge entre résidus clés. Une simple mutation d'un résidu glutamate en arginine sur le ^CDD de CurK a permis d'abolir son interaction avec son partenaire natif, mais a créé une communication avec le ^NDD de CurH, qui est impossible avec la protéine sauvage. L'affinité d'interactions entre DD sauvages est de l'ordre de la dizaine de micro molaires (Whicher et al., 2013).

IV.C.2.c. Domaines de *docking* particuliers chez les PKS *trans*-AT

Les DD de classes 1 et 2 ont été mis en évidence chez les PKS de type *cis*-AT. Jusqu'à récemment, peu d'informations structurales étaient disponibles quant aux DD des PKS *trans*-AT. Une différence importante chez ces systèmes réside dans la présence de modules « splités », où des domaines d'un même module sont localisés sur deux sous-unités distinctes. Plusieurs jonctions de ce type sont trouvées : KS/KR, KS/DH, KR/MT et DH/KR, entre autres.

Un mécanisme d'interaction entre des domaines KS et DH situés aux extrémités C_{ter} et N_{ter} de leur sous-unité respective a été mis en évidence (Jenner et al., 2018). Des alignements de séquences des extrémités C_{ter} de KS et N_{ter} de DH impliquées dans de telles interfaces montrent que ces dernières ne présentent pas de séquence supplémentaire non assignée, alors qu'il existe une séquence de longueur variable non annotée en aval des KS. Les auteurs se sont penchés plus particulièrement sur l'interface située entre les sous-unités 4 et 5 de la gladioline synthase (GbnD4 et GbnD5). Une analyse bio-informatique indique que les séquences en C_{ter} des KS présentent un haut degré de désordre avec une propension à se structurer en brin β. Le caractère désordonné de cette séquence située en C_{ter} de GbnD4, baptisée DH *docking* (DHD), a été confirmé par RMN et par dichroïsme circulaire (CD). En assignant les spectres RMN, les auteurs ont mis en évidence que ce domaine peut adopter plusieurs conformations en solution et que certaines régions ont tendance à se structurer en hélice α ou en brin β, mais le DHD n'adopte pas de structure tertiaire stable. Des expériences de titration RMN du DHD marqué ¹³C, ¹⁵N avec son partenaire DH-ACP non marqué, ont démontré que deux régions du DHD, d'une dizaine de résidus chacun, seraient impliquées dans l'interaction et se fixeraient en surface du domaine DH (Figure 37 page 64), avec une constante de dissociation (*K_d*) de 40 μM.

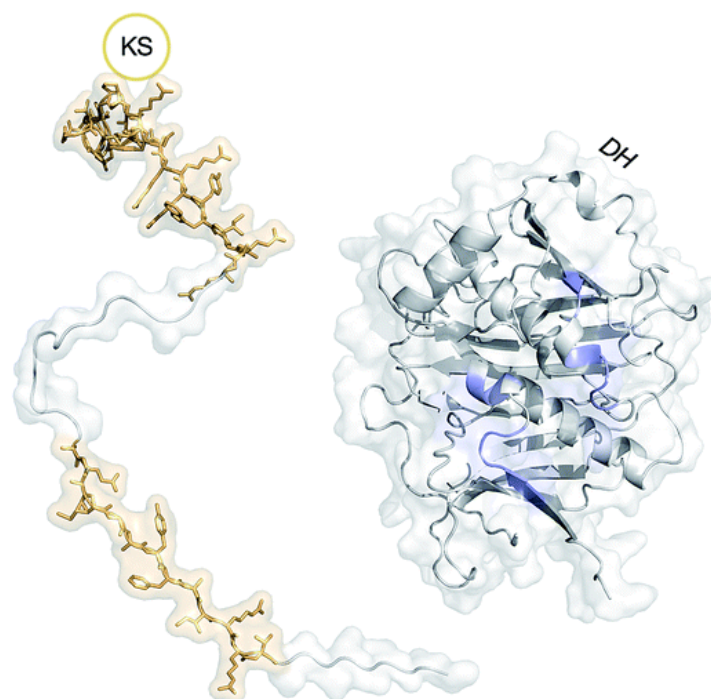


Figure 37 : Modèle de l'interaction entre le DHD et le domaine DH situés à l'interface des sous-unités 4 et 5 de la gladioline synthase (tirée de Kosol et al., 2018).

Le DHD déstructuré (à gauche), situé en aval de la KS, présente deux zones distinctes (en jaune) qui assurent son interaction avec des résidus en surface du domaine DH (en bleu).

Ces interactions sont spécifiques, car le DHD de GbnD4 ne forme pas de complexes stables avec d'autres DH de modules « splités » testés (Jenner et al., 2018), mais aucun code de spécificité n'est déterminé pour le moment. Cette étude a montré pour la première fois une interaction médiée directement entre un DD et un domaine catalytique.

IV.C.2.d. Domaines de docking de classe 3

Une autre classe de DD a été caractérisée chez le système hybride PKS/NRPS responsable de la synthèse de la tubulysine (Richter et al., 2008). Ce couple est situé à l'interface entre les sous-unités TubB et TubC, et plus spécifiquement, entre deux modules NRPS. Par alignement de séquences, les auteurs ont mis en évidence que cette classe est largement répandue chez les hybrides, résultats confirmés par des travaux récents (Hacker et al., 2018 ; Kosol et al., 2019).

La structure du DD en N_{ter} de TubC, nommé TubCdd, a été résolue par RMN. Ce DD de soixante-dix résidus se structure en trois hélices α et deux brins β qui adoptent une topologie

$\alpha\beta\alpha\alpha$. Ce DD s'homodimérise par l'intermédiaire des deux brins β , mais également par la boucle reliant le deuxième brin à l'hélice α_2 (Figure 38 page 65).

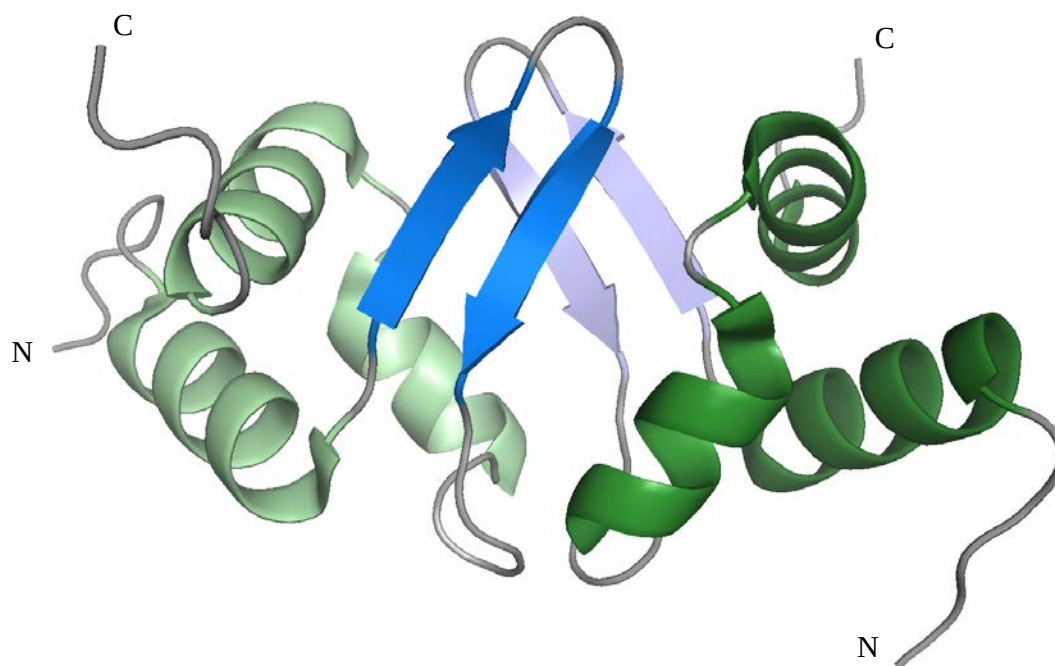


Figure 38 : Structure RMN de TubCdd (PDB : 2JUG).

Au centre, les feuillet β (en bleu) et la boucle (en gris) assurant la formation de l'homodimère. Chaque monomère comporte aussi trois hélices α (en vert).

Pour résoudre la structure des deux DD en interaction, les deux partenaires ont été fusionnés par l'intermédiaire d'un *linker* de deux acides aminés. Malheureusement, une analyse de la protéine par spectrométrie de masse a mis en évidence une protéolyse en N_{ter} de la construction, au niveau de TubBdd. Cette dernière, constituée de vingt-cinq résidus, est par la suite produite par synthèse chimique, et se révèle non structurée par RMN et CD. Elle permet tout de même de tester l'implication dans l'interaction de résidus chargés à la surface de TubCdd. De la même manière que chez les DD de classes 1 et 2, ces résidus chargés constitueraient un code de spécificité au sein de la classe 3.

Plus tard, la structure d'une protéine de fusion de DD de type 3 a été résolue par RMN (Hacker et al., 2018). L'interface est située entre les deux dernières sous-unités d'une NRPS : une rhabdopeptide synthase. A la différence de TubCdd, issue d'une hybride PKS/NRPS, le ^NDD de Kj12C est monomérique, mais la même conformation $\alpha\beta\alpha\alpha$ est adoptée. Il s'avère que le ^CDD de Kj12B est largement déstructuré en l'absence de son partenaire. Il présente cependant une hélice α et un brin β qui s'insère dans le feuillet formé par les deux brins du ^NDD dans la structure du complexe (Figure 39 page 66).

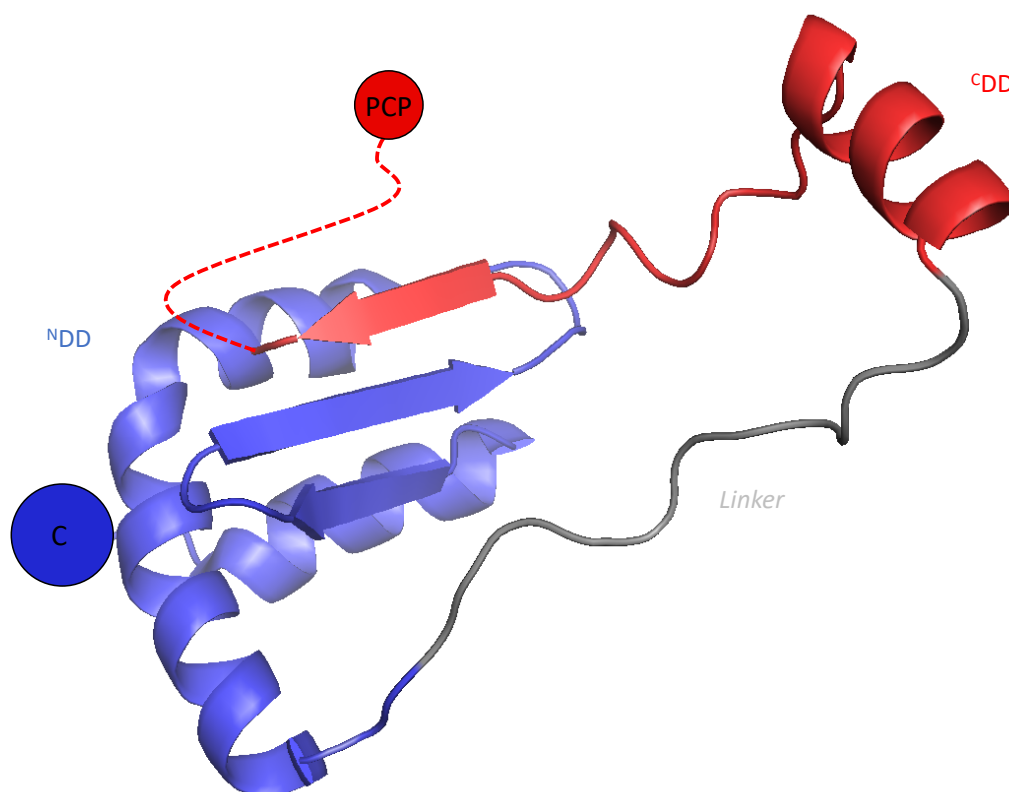


Figure 39 : Structure RMN d'une protéine de fusion de DD de type 3 trouvée chez une NRPS (PDB : 5EWV). Le ^CDD de la sous-unité KJ12B (en rouge) situé en aval d'un PCP, est composé d'une hélice α et d'un brin β qui se structure en s'insérant au niveau du feuillet β du ^NDD de KJ12C (en bleu). Ce dernier est situé en amont d'un domaine de condensation (C) monomérique. Le linker de douze résidus (en gris) a permis la fusion covalente de l'extrémité C_{ter} du ^NDD à l'extrémité N_{ter} du ^CDD.

Il semble donc que ce brin β est formé lorsque les deux partenaires interagissent, puisque l'hélice α n'intervient pas dans l'interaction. Une titration du ^NDD marqué par le ^CDD a démontré que le site de fixation du ^CDD sur son partenaire fait partie de la surface d'homodimérisation équivalente chez TubCdd. Le mécanisme d'interaction entre les deux DD de la PKS/NRPS doit donc être différent. En effet les NRPS sont monomériques tandis que les modules NRPS présents au sein d'hybrides PKS/NRPS sont homodimériques, ceci étant certainement dû à la présence des PKS.

IV.C.3. Un code de spécificité hiérarchique

La résolution de plusieurs structures de DD en complexes et les alignements de séquences permettent de hiérarchiser le code de spécificité de ces DD. Le premier niveau correspond à la classe structurale à laquelle ils appartiennent, et le second aux sous-classes : les DD d'une même classe structurale ne divergent que de quelques résidus clés. Ainsi, il est possible que des DD d'une même sous-classe puissent interagir ensemble grâce à quelques mutations ponctuelles, tandis que des DD de deux classes différentes ne pourront surement

jamais interagir ensemble. Ces informations nous donnent des indices quant à la pression de sélection qui opère au sein des systèmes PKS. S'il était possible qu'un DD puisse changer de classe par simples mutations, des interactions indésirables entre sous-unités de PKS seraient fréquentes. A l'extrême opposé, si les DD étaient très différents les uns des autres, il serait extrêmement difficile de « reprogrammer » l'ordonnancement des sous-unités, et les bactéries seraient moins adaptables à un changement de leur environnement. Cette organisation hiérarchique en classes et sous-classes représente certainement une stratégie intermédiaire optimale entre les exigences compétitives et la flexibilité (Thattai, Burak, et Shraiman, 2007).

L'objectif principal de ma thèse était de caractériser plus en détails les DD d'un point de vue biochimique mais aussi structural. Pour ce faire, plusieurs couples ont été étudiés au sein de deux systèmes PKS différents dont les molécules synthétisées présentent un fort potentiel pharmacologique.

IV.D. Organisation macromoléculaire de PKS *trans*-AT

Peu d'études se sont intéressées à l'organisation et à la localisation cellulaire adoptées par les PKS. Au vu des constantes de dissociations relativement faibles présentées par les DD, il est fort probable que les sous-unités au sein des systèmes PKS interagissent ensemble de manière transitoire. Ceci pourrait grandement ralentir la synthèse du polycétide en croissance dans le cas où les différentes sous-unités seraient libres de diffuser au sein de la cellule. Ces interactions transitoires auraient leur avantage : comme discuté précédemment, cela favoriserait le renouvellement des sous-unités défectueuses mais cela pourrait aussi refléter la nécessité de l'ACP_n d'accéder aux domaines de son module mais aussi à la KS_{n+1} située sur la sous-unité suivante. En effet si l'ACP était immobilisé de manière permanente par une interaction trop forte entre les DD, la chaîne d'assemblage serait bloquée.

La localisation cellulaire d'une seule PKS a été étudiée à ce jour. Il s'agit de la bacillaene synthase, une PKS *trans*-AT présente chez *Bacillus subtilis* (Straight et al., 2007). Ce système comprend 16 gènes au sein de son *cluster*, ce qui représente près de 2% du génome de la bactérie. Si l'on considère une seule copie de l'ensemble des sous-unités protéiques nécessaires à la synthèse du polycétide, la chaîne d'assemblage fait environ 2,5 MDa. Dans un premier temps, les auteurs ont fusionné différentes sous-unités de la PKS ainsi que l'une des *trans*-AT à des protéines fluorescentes : *cyan fluorescent protein* (CFP) ou *yellow fluorescent protein* (YFP) (Figure 40 page 68).

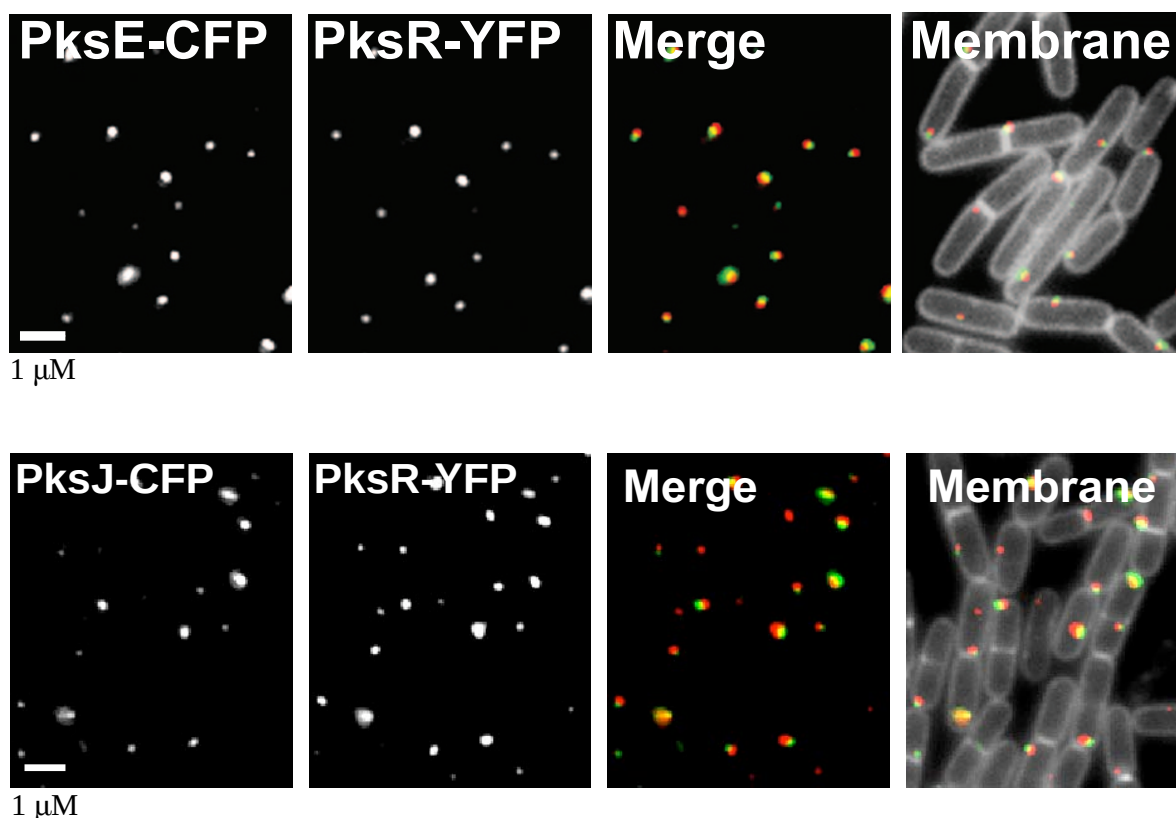


Figure 40 : Co-localisation au sein des cellules de différentes protéines impliquées dans la synthèse du bacillaène (adaptée de Straight et al., 2007).

PksE, l'une des *trans*-AT du système, a été fusionnée à la cyan fluorescent protein. *PksR* est la dernière sous-unité multi-modulaire du système, elle est fusionnée en *C_{ter}* à la yellow fluorescent protein. *PksJ* est la première sous-unité multi-modulaire du système, elle est fusionnée en *N_{ter}* à la cyan fluorescent protein.

A gauche : fluorescence des protéines indiquées. Les images ont été ensuite fusionnées (merge) et la co-localisation des deux protéines a été mise en évidence en points jaunes correspondant à une superposition des spots de fluorescence de la CFP et de l'YFP. Les membranes ont été ensuite révélées grâce à un marquage spécifique.

Ils ont mis ainsi en évidence que la dernière sous-unité du système (*PksR*) co-localise avec l'une des *trans*-AT (*PksE*) mais aussi avec la première sous-unité (*PksJ*). De plus, dans les deux cas, les protéines se juxtaposent à la membrane cellulaire et leur position au sein de la cellule varie de l'équateur au pôle.

Pour commencer à mettre en évidence les déterminants nécessaires à la formation des mega complexes de PKS, les auteurs ont aussi travaillé sur une souche dont les différentes sous-unités multi-modulaires sont supprimées, et ceci mène à la diffusion totale de la *trans*-AT *PksE* au sein de la cellule, ce qui suggère que c'est la localisation des sous-unités géantes qui détermine la localisation des enzymes agissant en *trans* (Straight et al., 2007).

Ces résultats ont été ensuite confirmés par cryo-microscopie électronique à transmission (Figure 41 page 69). Dans un premier temps, des masses riches en électrons ont été observées

proches des membranes. Elles diffèrent en forme et en taille et ne semblent pas présenter de limite discrète, ce qui suggère qu'elles ne sont pas enfermées par la membrane. Pour déterminer si ces masses observées dans la souche *wild type* (wt) correspondent bien au mega complexe de bacillaene synthase, les auteurs ont procédé à une détection immuno-EM dans les souches PksR-YFP et PksE-CFP, en utilisant un anticorps primaire anti-*green fluorescent protein*, ainsi que des billes d'or couplées à un anticorps secondaire. Il a été alors observé que les protéines de fusion (par l'intermédiaire des billes d'or) se localisaient bien avec les masses denses en électrons situées dans le cytoplasme.

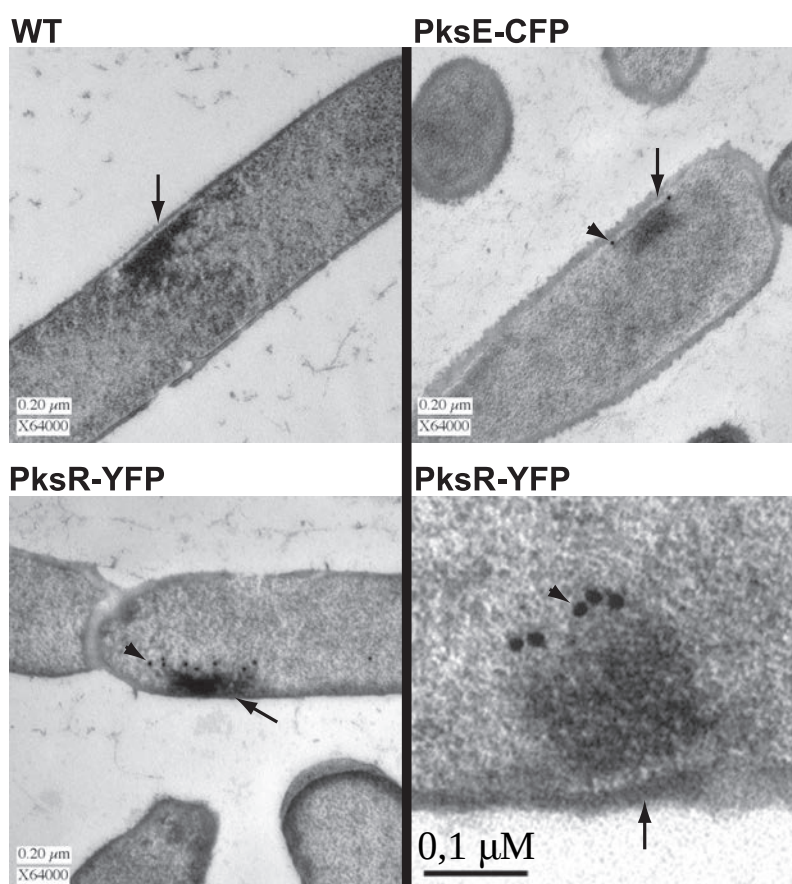


Figure 41 : Mise en évidence de l'assemblage d'un mega complexe de PKS par microscopie électronique à transmission (Straight et al., 2007).

Des masses denses en électrons sont localisées proches de la membrane pour la souche wt non modifiée, mais aussi chez les cellules produisant les protéines de fusion PksE-CFP et PksR-YFP (indiquées par des flèches). La localisation immunoEM de ces protéines de fusion a été visualisée par la présence de particules d'or de 20 nm conjuguées à des anticorps secondaires (indiquées par des têtes de flèche). En bas à droite, un zoom plus poussé de l'image enregistrée pour la construction PksR-YFP est présenté. La flèche pointe vers la membrane cellulaire et la membrane plasmatique. La tête de flèche pointe les particules d'or.

La formation et la localisation de ces mega complexes sont cohérentes avec l'hypothèse de la formation d'une « usine biosynthétique » potentiellement couplée directement avec des mécanismes sécrétoires (Simunovic, Gherardini, et Shimkets, 2003 ; Jain et Cox, 2005), évitant

ainsi l'accumulation toxique du bacillaene dans le cytoplasme. De plus, cela faciliterait le recrutement d'un *pool* de blocs acyl-CoA ainsi que les interactions entre sous-unités.

Il reste à déterminer plus précisément comment les sous-unités PKS et NRPS s'organisent en un tel complexe. Une possibilité serait qu'une portion de chaque sous-unité s'associe à la membrane soit directement soit *via* l'intermédiaire d'une protéine membranaire. Une autre possibilité serait que les sous-unités s'associent directement ensemble pour former le mega complexe par l'intermédiaire de certains domaines et que l'interaction avec la membrane ne se fasse qu'avec quelques éléments des chaînes d'assemblage (Straight et al., 2007). Dans ce contexte, il a été proposé que des motifs *Laterally-Interacting Ketosynthase Sequences* (LINKS) disposés de part et d'autre des homodimères de KS de PKS *trans*-AT permettent la formation de tels mega complexes, en favorisant une interaction homotypique entre domaines KS. Les auteurs se sont basés sur des structures cristallographiques de ces domaines : dans neuf des dix structures de KS de *trans*-AT, les motifs LINKS s'auto-associent et dans huit de ces structures, ils assurent un positionnement parallèle des dimères de KS (Gay et al., 2016) (Figure 42 page 70).

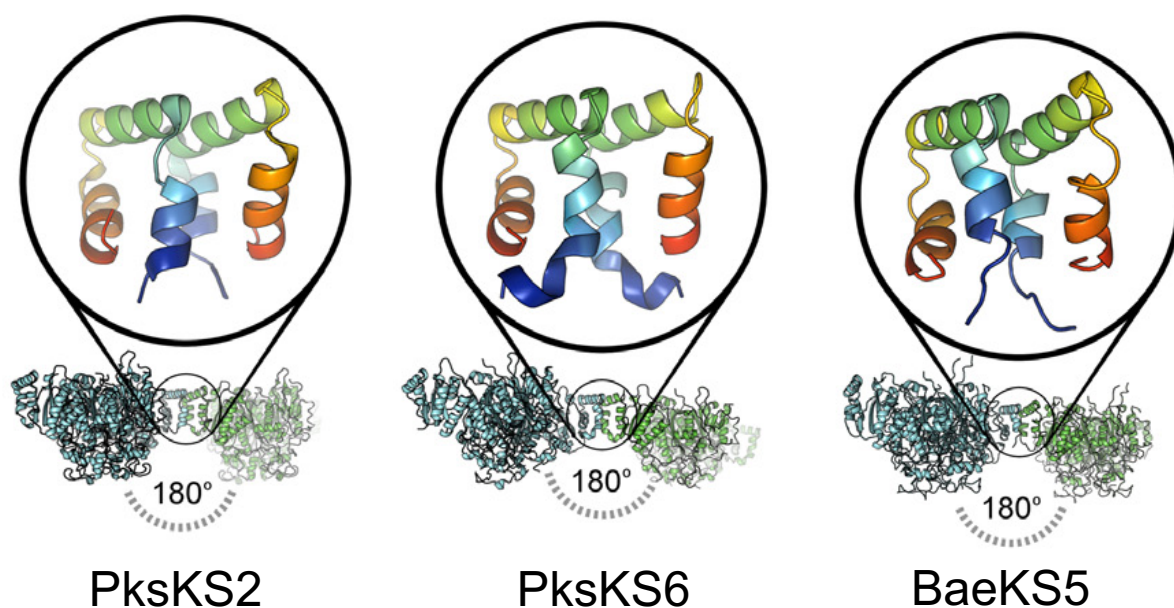


Figure 42 : Structures cristallographiques de KS de PKS *trans*-AT révélant la présence de motifs LINKS (tirée de Gay et al., 2016).

Ces KS sont issues de deux PKS *trans*-AT synthétisant la bacillaene ; les deux premières sont issues de *Bacillus subtilis* (PksKS2, PDB : 4NA1 ; PksKS6, PDB : 5ERF) et la troisième de *Bacillus amyloliquefaciens* (BaeKS5, PDB : 5ERB). Les modèles d'interaction entre motifs LINKS sont grossis : un monomère est composé de trois hélices α , les homodimères interagissent entre eux via des surfaces hydrophobes et des résidus électrostatiques. L'angle relatif entre les deux dimères est présenté en dessous des structures.

Des analyses bio-informatiques suggèrent que les séquences des LINKS sont très peu conservées et que les interactions entre ces motifs sont uniques pour chaque domaine KS,

favorisant des interactions « latérales » et stabilisant ainsi les interactions « verticales » (Figure 43 page 71).

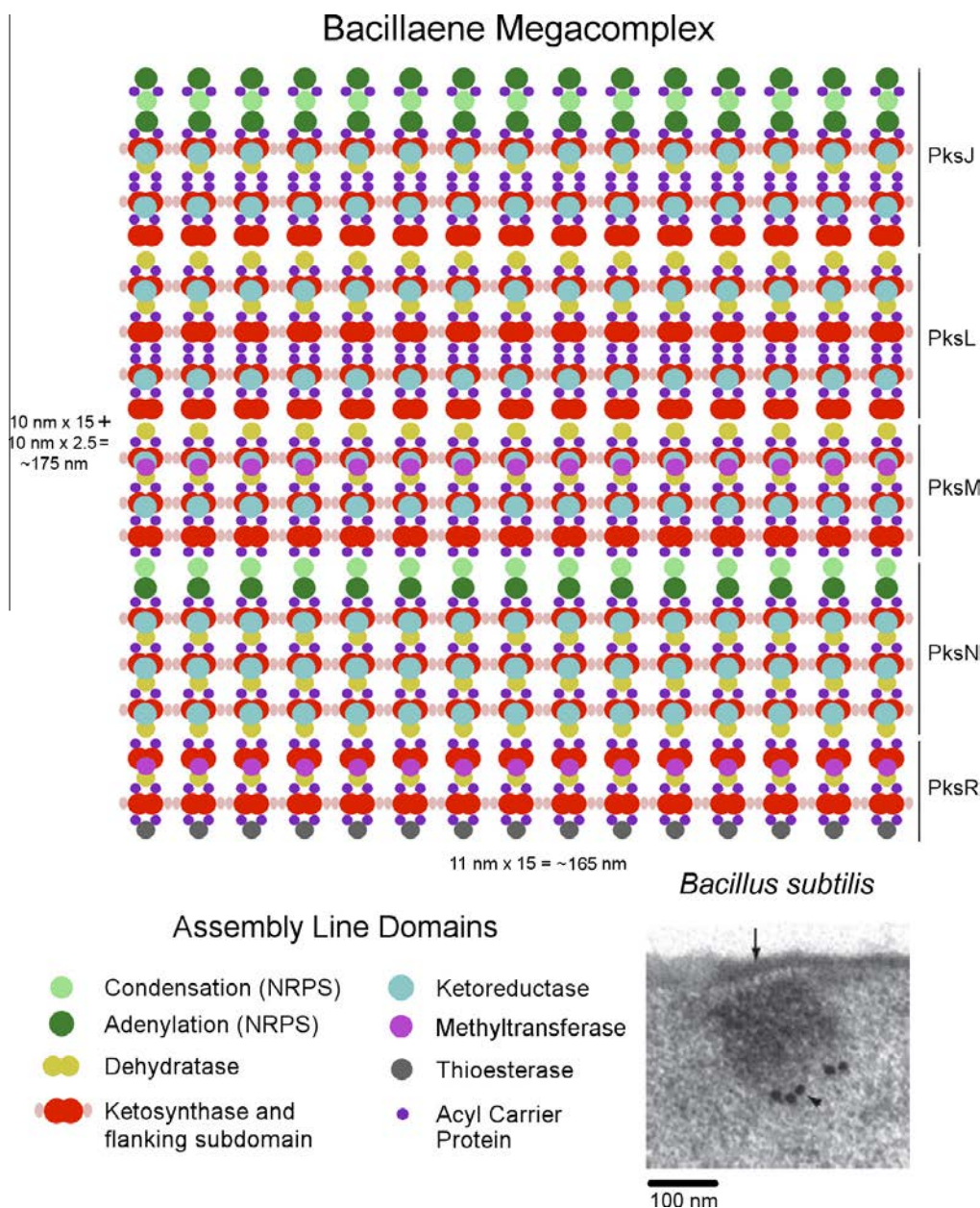


Figure 43 : Modèle proposé de formation du mega complexe de PKS trans-AT stabilisé par des interactions verticales entre sous-unités et latérales entre KS d'un même module (tirée de Gay et al., 2016).

Cependant les auteurs de cette étude n'ont confirmé ces résultats par aucune méthode plus adaptée pour déterminer l'état oligomérique de ces domaines. De plus, aucun autre domaine ou didomaine n'a été étudié, or les travaux de cryo-EM portant sur un module entier ont démontré que les structures isolées des didomaines KS-AT n'adoptent pas la même conformation lorsqu'ils sont replacés dans leur contexte modulaire. Il se peut donc que la

structuration de ces motifs LINKS ne soit pas représentative de la structuration native de cette région au sein du module.

V. Présentation de nos systèmes d'études

V.A. La virginiamycine synthase, une PKS *trans*-AT

La virginiamycine M est une molécule présentant une activité bactériostatique et produite par *Streptomyces virginiae*. Cette souche présente la particularité de produire une autre molécule bactériostatique, la virginamycine S. Un effet synergique est observé lors d'une utilisation combinée des molécules. D'ailleurs une combinaison des dérivés hémi-synthétiques de la virginiamycine M et de virginamycine S, la dalfopristine et la quinupristine respectivement, commercialisée sous le nom de Synercid®, est utilisée pour traiter des infections par des bactéries Gram-positifs résistantes à la vancomycine (Niccolai, Tarsi, et Thomas, 1997). Leur activité est alors accrue d'un facteur cent (Cocito, 1979).

La virginiamycine M est un inhibiteur de la synthèse protéique. Plus particulièrement, elle se fixe à la grande sous-unité 50s des ribosomes procaryotes, à cheval sur les sites P et A, empêchant alors la fixation de l'ARNt au niveau du site P (Hansen, Moore, et Steitz, 2003). En se fixant, un changement conformationnel de la sous-unité ribosomique est induit (Parfait et Cocito, 1980), ce qui permettrait à la virginiamycine S de se fixer, d'où l'effet synergique observé entre les deux molécules. Cette dernière se fixe également au niveau des sites A et P du ribosome mais en empêchant cette fois-ci la formation de la liaison peptidique. Alors que la virginiamycine S est synthétisée par une NRPS (Namwat et al., 2002), la virginiamycine M est quant à elle synthétisée par un hybride PKS/NRPS (Pulsawat, Kitani, et Nihira, 2007).

Le *cluster* de gènes de la virginiamycine M synthase comporte entre autres, trois gènes codant pour trois sous-unités de PKS (Figure 44 page 73).

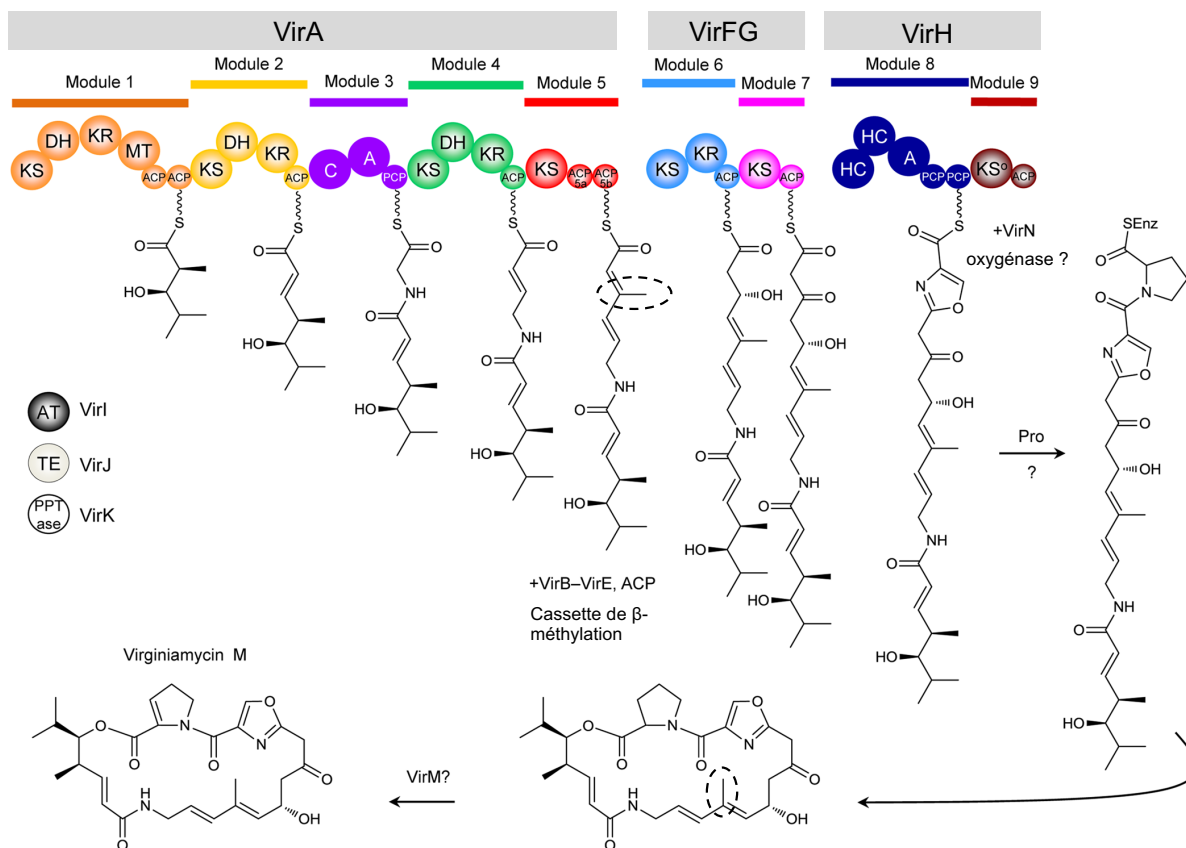


Figure 44 : Voie de synthèse proposée de la virginamycine M (adaptée de Dorival et al., 2016).

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 (Pulsawat, Kitani, et Nihira, 2007). 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 pointillés. VirI est l'AT discrète et VirJ présente l'activité TE. Le dernier module responsable de l'incorporation de la proline est manquant dans le séquençage de la PKS, et n'est donc toujours pas identifié. Le module 9 présente une KS ne possédant pas d'histidine catalytique, nommée KS°.

VirA est composée de cinq modules, tandis que VirFG et VirH sont composées de deux modules chacune. Le gène du module de chargement ainsi que celui du module 10, responsable de l'incorporation de la proline terminale, ne sont pas identifiés dans la séquence génomique de *S. virginiae* disponible dans les bases de données. Ils ont cependant mis en évidence dans un *cluster* homologue chez *S. pristinaspiraelis* qui synthétise la même molécule (Mast et al., 2011). Le *cluster* ne comporte qu'un seul gène codant pour un domaine AT : VirI, responsable de la sélection spécifique du malonyl-CoA. Deux modules NRPS sont présents au sein des sous-unités : le module 3 qui incorpore une glycine et le module 8 une sérine. Les modules 1 et 5 sont caractéristiques des PKS *trans*-AT et comportent deux domaines ACP en tandem.

Deux hypothèses sont émises quant au rôle d'un tel tandem. La première est que ces ACP peuvent travailler en parallèle et ainsi augmenter la production du polycétide (Rahman et al., 2005). La deuxième hypothèse propose au contraire, que la duplication du domaine permet

d'augmenter le temps de présence de l'intermédiaire réactionnel au sein du module et favoriserait donc des modifications supplémentaires avant qu'il ne soit transféré au module suivant (Calderone et al., 2008). D'ailleurs, une cassette de β -méthylation, composée de cinq protéines (VirAB, VirB–E)(Davison et al., 2014), intervient au niveau du tandem d'ACP (ACP_{5a} et ACP_{5b}) du module 5 et assure la méthylation du carbone β du polycétide en cours de croissance. Il est supposé que le domaine TE du module manquant assure la libération et la cyclisation de la molécule en fin de synthèse. Enfin une déshydrogénase, proposée être VirM, oxyde la proline permettant l'obtention de la molécule finale active.

A mon arrivée au laboratoire, aucune information n'était disponible quant au mécanisme d'interaction entre sous-unités de PKS *trans*-AT. La première interface étudiée au cours de ma thèse se situe donc entre les sous-unités VirA et VirFG, et plus précisément entre le module 5, qui présente le tandem d'ACP, et le module 6 qui présente une composition plus habituelle : une KS, une KR et un ACP.

V.B. L'enacyloxine synthase, une PKS hybride *cis*-/trans-AT

Un autre objectif de ma thèse consistait à caractériser des DD entre chaque interface de l'enacyloxine synthase évoquée en II.C.3.b.ii. A ce jour, aucune étude exhaustive n'est menée sur tous les DD d'un même système PKS. Ceci permettrait de mieux comprendre les déterminants clés responsables de la spécificité d'interaction entre sous-unités, et ainsi mieux appréhender et maîtriser la construction de systèmes chimériques.

L'enacyloxine est synthétisée par *Burkholderia ambifaria*, une bactérie Gram-négative appartenant au *Burkholderia cepacia complex* (Bcc), un sous-groupe d'espèces de *Burkholderia* très proches et connu pour leur production de métabolites secondaires à forts intérêts, présentant des activités antibactériennes et antifongiques (Parke et Gurian-Sherman, 2001 ; Vial et al., 2008 ; Seyedsayamdost et al., 2010). Cette molécule présente une activité antibiotique face à des souches pathogènes multi-résistantes telles que *P. aeruginosa* et *Acinetobacter baumannii* (Mahenthiralingam et al., 2011). Pour l'Organisation Mondiale de la Santé, ces deux souches sont récemment passées dans la catégorie de priorité la plus élevée de pathogènes multi-résistants pour lesquels il existe un besoin critique de recherche et de développement de nouveaux antibiotiques (Tacconelli et Magrini, 2017). Il serait donc intéressant de modifier la structure cœur de la molécule qui n'est pas stable due à l'isomérisation de doubles liaisons

(Figure 45 page 76). Afin de faciliter les échanges de modules voire de sous-unités de ce système, nous nous sommes intéressés à toutes les interfaces entre sous-unités PKS de ce système.

L'enacyloxine synthase (Figure 45 page 76) en présente six puisqu'elle est composée de sept sous-unités (5925–5919), ce qui est relativement élevé par rapport à d'autres systèmes (Helfrich et Piel, 2016 ; Medema et al., 2015). Sur ces six interfaces, cinq d'entre elles impliquent un ACP et une KS. Cette PKS est donc un excellent modèle pour mieux comprendre comment la spécificité d'interaction est médiée dans une telle configuration. De plus, l'une des interfaces, celle située entre les sous-unités 5924 et 5923, consiste en un module « splité » : la KS du module 3 est localisée à l'extrémité C_{ter} de la sous-unité 5924, alors que les domaines KR et ACP de ce même module 3 sont portés par la sous-unité suivante.

Comme évoqué précédemment, des analyses bio-informatiques démontrent que, de manière surprenante, les six premières sous-unités (5925–5920) appartiennent à la classe des PKS *cis*-AT et seule la dernière sous-unité (5919) appartient à la classe des PKS *trans*-AT. Chaque sous-unité comporte un à deux modules, mais tous les modules ne présentent pas systématiquement de domaine AT, même ceux appartenant à la classe des PKS *cis*-AT. Seuls les modules 1 et 6 en présentent un. Cependant aucun gène du *cluster* ne code pour un tel domaine ; il a donc été proposé que le domaine AT du module 1 ou 6 charge en *trans* les ACP des autres modules avec le malonyl-CoA, ou que la réaction peut être catalysée par une protéine codée par un gène situé en dehors du *cluster* de la PKS. En effet, quatre gènes ont été identifiés dans le génome de *B. ambifaria* qui codent pour des domaines AT et présentent une forte homologie avec les AT présents dans les sous-unités 5925 et 5921 (Mahenthiralingam et al., 2011).

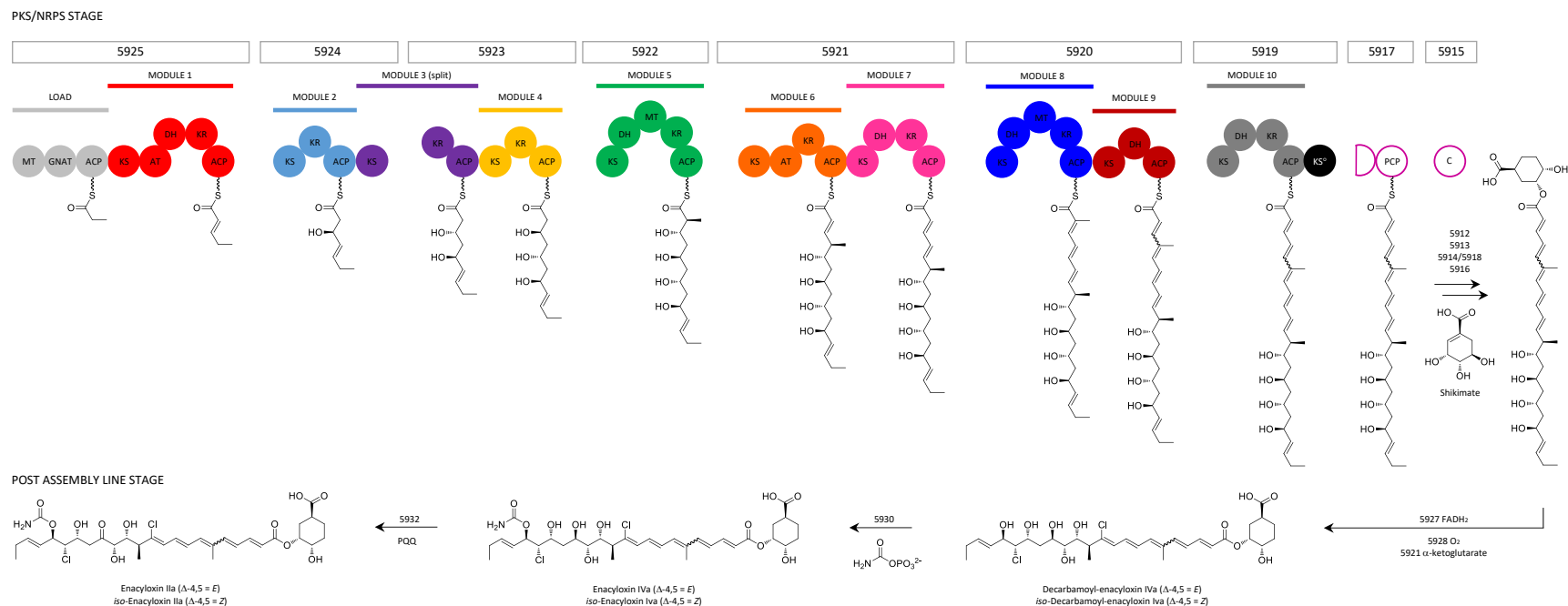


Figure 45 : Organisation en domaines et modules de la PKS synthétisant l'enacyloxine (adaptée de KJ Weissman).
 Le mécanisme de condensation de la chaîne polycétidique avec l'acide dihydroxycyclohexane carboxylique (DHCCA) et les modifications post-PKS sont présentés. L'ordre des réactions de modification du polycétide n'est que supposé et ces réactions pourraient aussi avoir lieu durant la synthèse de la molécule par la PKS.

Il existe également quelques différences entre la structure déterminée de l'enacyloxine IIa et la structure prédite de l'intermédiaire porté par l'ACP du module 9. Le module 9 est composé d'une KS, d'une DH et d'un ACP, ce qui est étonnant puisque la molécule finale présente bien une double liaison entre les carbones 4 et 5 alors qu'aucun domaine KR n'est présent au sein du module pour réduire la fonction céto en hydroxyle. Les auteurs proposent que cette réduction et l'élimination de l'eau sont catalysées par les domaines KR et DH du module 10. Il est également proposé que la double liaison entre les carbones 10 et 11 est introduite par le domaine DH du module 7, puisque le module 6 n'en comporte pas. La présence des groupements chlore sur les carbones 11 et 18, ainsi que le groupement hydroxyle du C14 et le carbamoyl du C19 seraient, quant à eux, introduits par des enzymes post-PKS présentes dans le *cluster* de gènes de la PKS (Figure 45 page 76).

Des travaux récents (Kosol et al., 2019 ; Masschelein et al., 2019) ont mis en lumière le mécanisme inhabituel de libération de la chaîne polycétidique par la PKS. Le dernier module, de nature *trans*-AT, présente à son extrémité C_{ter} un domaine KS⁰, inactif, assurant la *trans*-acylation de l'intermédiaire vers un domaine PCP (5917). Par la suite, l'intermédiaire est présenté à la protéine 5915 dont la séquence est similaire à un domaine de condensation de NRPS. Ce domaine catalyse une réaction inhabituelle d'estérification de l'intermédiaire avec l'acide dihydroxycyclohexane carboxylique (DHCCA) et la libération concomitante de la molécule. C'est la première fois qu'un tel mécanisme de libération a été mis en évidence (Masschelein et al., 2019). La résolution de la structure cristallographique du domaine de condensation a mis en évidence à son extrémité N_{ter} un DD de type 3 monomérique, adoptant le *fold* $\alpha\beta\beta\alpha\alpha$ (Kosol et al., 2019). De manière attendue, la structure RMN du domaine PCP présente une extrémité C_{ter} extrêmement flexible qui entre en interaction avec le NDD du domaine C, probablement *via* le même mécanisme d'insertion et de formation d'un brin β déjà mis en évidence (Hacker et al., 2018). Il semble aussi que les deux domaines, PCP et C, subissent des changements de conformation importants lors de l'interaction, ce qui jouerait un rôle lors de la catalyse réalisée par le domaine C.

V.C. LkcE, une enzyme post-PKS

La lankacidine est un antibiotique couramment utilisé en médecine vétérinaire pour traiter le bétail infecté par des bactéries Gram-positifs pathogènes (Hayashi et al., 1988). Elle est commercialisée sous le terme de sédécamycine ou terdécamycine, pour son variant hémisynthétique.

Tout comme la virginiamycine M, la lankacidine agit en synergie avec un autre polycétide, 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 polycétides ont co-évolué pour agir simultanément (Arakawa et al., 2007 ; Yamamoto et al., 2008). Ces deux molécules se fixent sur la sous-unité 50S du ribosome. La structure cristallographique du complexe a été résolue avec les deux molécules, permettant d'identifier le site de fixation de chaque antibiotique (Auerbach et al., 2010 ; Belousoff et al., 2011). La lankamycine se fixe au niveau du centre peptidyl-transférase alors que la lankacidine se lie au niveau du site accepteur, empêchant ainsi le positionnement adéquat de l'acide aminé porté par l'ARNt. 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 chez *S. rochei* (Kinashi et al., 1994) par un plasmide portant également les gènes de régulation : le plasmide PSLA2-L. Ce plasmide de 200 kb a été séquencé par la suite, révélant la présence de quatre *clusters* de gènes responsables de la synthèse de quatre métabolites secondaires (Suwa et al., 2000).

La lankacidine synthase est une PKS *trans*-AT qui assure la condensation de huit unités malonate 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 sont proposées pour la lankacidine (Piel, 2010). La plus probable (Figure 46 page 79) 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 (Dickschat et al., 2011). Ensuite, huit unités malonates supplémentaires sont ajoutées. Pour cela, LkcC et LkcF agiraient 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 polycétide. Il est à noter qu'une DH discrète, la LkcB interviendrait lors de plusieurs cycles d'extension (Dickschat et al., 2011 ; Tatsuno, Arakawa, et Kinashi, 2007 ; Tatsuno, Arakawa, et Kinashi, 2009).

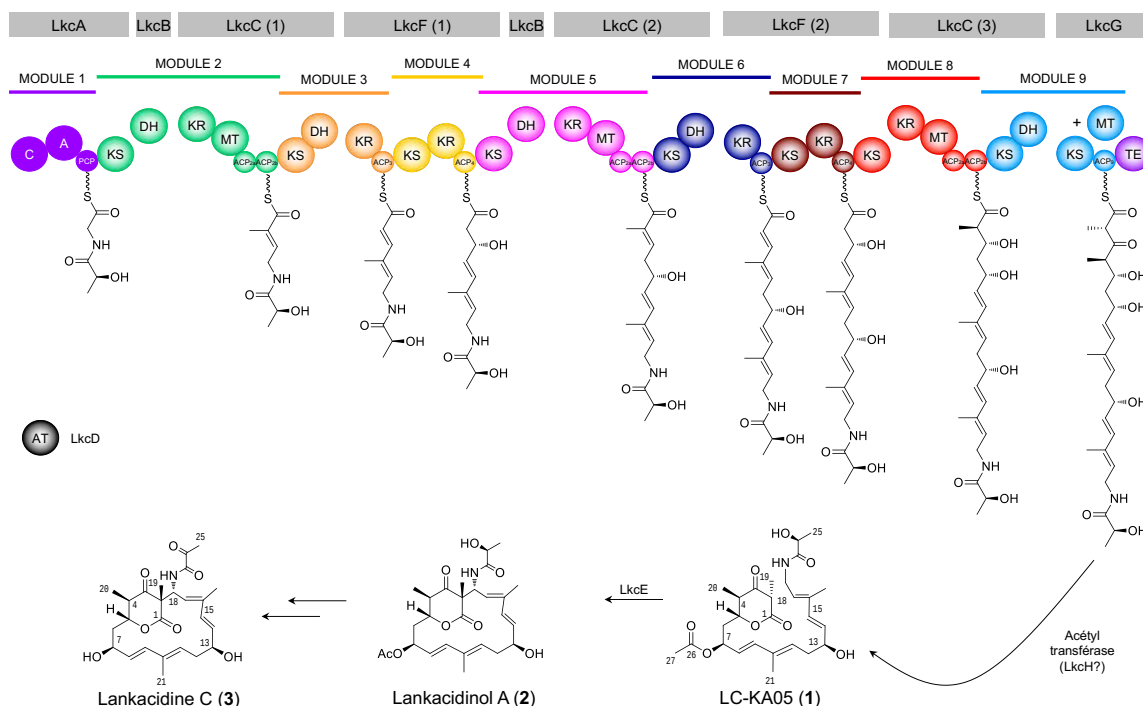


Figure 46 : Schéma de la voie de synthèse proposée pour la lankacidine C (fournie par KJ Weissman).

La libération de la chaîne par le domaine TE de LkcG n'entraîne pas la cyclisation complète du polycétide. L'intervention d'une enzyme post-PKS est donc nécessaire pour catalyser la macrocyclisation 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 ion iminium, permettant ainsi la cyclisation de la lankacidine par une réaction de type Mannich (Figure 47 page 79).

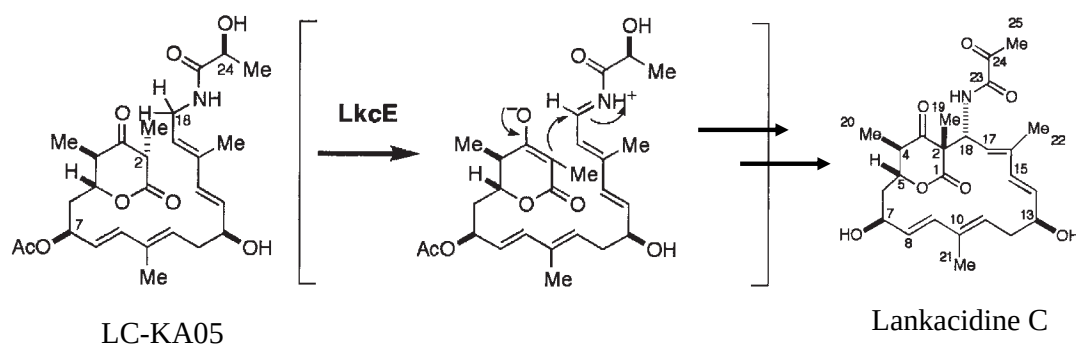


Figure 47 : Mécanisme proposé pour la macrocyclisation de la lankacidine C (tirée de Arakawa et al., 2005). La LkcE catalyserait l'oxydation de la fonction amine du LC-KA05 en ion iminium par transfert d'un électron et d'un proton sur le FAD, et utiliserait une catalyse base générale pour subtiliser un proton au niveau du carbone 2. Ce carbone ainsi activé pourrait réaliser une attaque nucléophile au niveau de l'ion iminium, permettant la cyclisation du polycétide pour former le lankacidinol A. Ce dernier subirait ensuite une oxydation au niveau du carbone 24 et une élimination du groupement acétate pour former la lankacidine C.

Une délétion du gène *lkcE* chez *S. rochei* a généré 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).

Sur la base d'études portant sur les monoamines oxydases humaines, plusieurs mécanismes sont proposés pour le transfert d'électron au groupement flavine adénine dinucléotide (FAD) permettant l'oxydation de l'amine (Fitzpatrick, 2010). La plus probable (Figure 48 page 80), soutenue par les structures cristallographiques des amines oxydases, implique un transfert de l'hydrogène en tant qu'hydrure directement sur l'azote du FAD, puisqu'aucun résidu susceptible d'agir comme catalyseur basique n'est présent au sein des sites actifs.

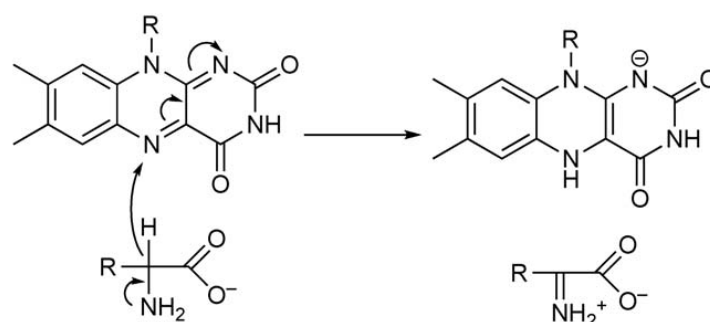


Figure 48 : Mécanisme proposé pour l'oxydation d'une amine avec transfert direct de l'hydrure depuis l'amine sur le FAD (tirée de Fagan et Palfey, 2010).

La demi-réaction d'oxydation du FAD peut impliquer différents partenaires. Dans le cas des amines-oxydases, celle-ci se fait grâce à l'oxygène moléculaire O_2 (Fitzpatrick, 2010). L' O_2 cède 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 oxydée du cycle isoalloxazine (Figure 49 page 81) (Walsh et Wencewicz, 2014).

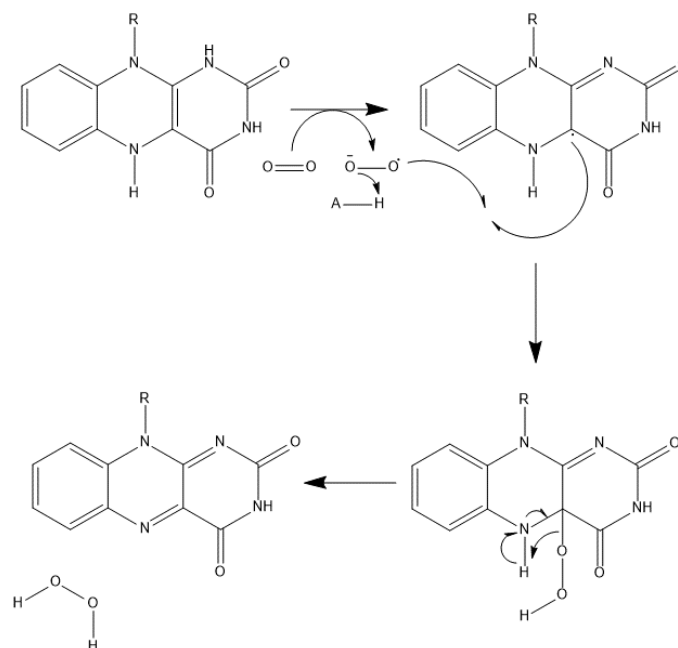


Figure 49 : Mécanisme d'oxydation du FAD par l'oxygène moléculaire O_2 (tirée de Dorival, 2016).

HYPOTHÈSES ET OBJECTIFS DE TRAVAIL

Dans le but d'améliorer les expériences d'ingénierie des PKS, plusieurs points clés de la synthèse des polycétides restent à être éclaircis. L'un d'entre eux est la base moléculaire de la communication spécifique entre sous-unités d'un même système PKS. Une meilleure compréhension de ces interactions est un point critique, puisqu'elle permettrait l'échange de modules voire de sous-unités entières entre différents systèmes, stratégie qui semble la plus prometteuse à ce jour (Whicher et al., 2014 ; Menzella et al., 2007 ; Kim et al., 2017).

A mon arrivée au laboratoire, il existait un manque important d'informations quant aux DD de PKS *trans*-AT, des PKS modulaires qui sont *a priori* les mieux adaptées aux remaniements/manipulations de leur chaîne d'assemblage (Tille et Race, 2014). En effet, les quelques structures de DD disponibles (Broadhurst et al., 2003 ; Buchholz et al., 2009 ; Whicher et al., 2013) provenaient exclusivement de PKS *cis*-AT. Nous nous sommes donc intéressés à la première interface de la virginiamycine (Vir) synthase, une PKS *trans*-AT. Elle se situe à l'extrémité C_{ter} de la première sous-unité, au niveau du module 5, et à l'extrémité N_{ter} de la seconde sous-unité, au niveau du module 6 de la Vir PKS. Cette *trans*-AT synthétise un antibiotique à fort intérêt thérapeutique dont l'interaction avec le ribosome est bien caractérisée (Schlünzen et al., 2001), ce qui fait de ce système une cible de choix pour l'ingénierie.

Dans ce contexte, une étude précédente de notre équipe portait sur la structure adoptée par le module 5. Celui-ci présente une composition particulière en domaines, puisqu'il comporte une KS et deux domaines ACP en tandem (ACP_{5a} et ACP_{5b}), et permettrait l'intervention de plusieurs protéines appartenant à une cassette de β -méthylation (Davison et al., 2014). En utilisant une approche combinant RMN, SAXS et modélisation par homologie de séquence, la structure de ce module a été obtenue – première fois qu'une structure de module de PKS *trans*-AT est mise en évidence (Figure 50 page 84) (Davison et al., 2014). La structure dimérique globale de ce module adopte une forme d'arche, semblable à celle adoptée par le module 5 de la pikromycine synthase (Dutta et al., 2014 ; Whicher et al., 2014). Au sommet de cette arche se trouve le dimère de domaines KS, suivi directement par les ACP_{5a}, alors que les ACP_{5b} se situent aux extrémités. Il existe à l'extrémité C_{ter} de cette construction, en aval de l'ACP_{5b}, une séquence d'acides aminés qui a été, dans un premier temps, assignée comme étant un domaine de dimérisation (Davison et al., 2014). Or le modèle structural obtenu pour le module 5 va à l'encontre de cette hypothèse, puisque l'arche reste ouverte, les extrémités C_{ter} ne se rejoignant pas (Figure 50 page 84). De plus, un domaine partenaire potentiel a été identifié à l'extrémité N_{ter} du module 6 porté par la sous-unité suivante.

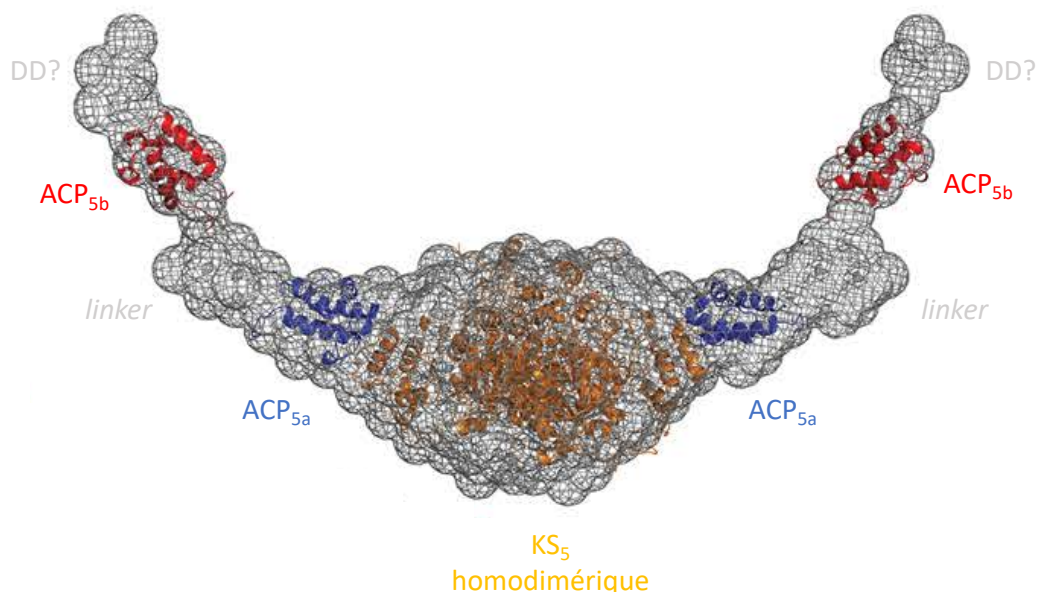


Figure 50 : Modèle structural du module 5 de la virginiamycine synthase (tirée de Davison et al., 2014).
 En gris, l'enveloppe SAXS moyenne calculée, dans laquelle est replacé un modèle par homologie de la KS₅ homodimérique (orange), les structures RMN de l'ACP_{5a} (bleu) (PDB : 2M4F), et de l'ACP_{5b} (rouge) (PDB : 4CA3).

Nous avons donc émis l'hypothèse que ces deux domaines situés aux extrémités des sous-unités sont en réalité des domaines de *docking* (DD).

Une deuxième étude a donc été menée par notre équipe sur ces régions à l'interface des deux sous-unités. Le but était de mettre en évidence l'interaction entre ces deux DD potentiels, puis de la caractériser plus en détails notamment en résolvant la structure du complexe par RMN. Les résidus impliqués seraient ainsi identifiés et leur mutation rationnelle pourrait être testée pour la biologie synthétique. Enfin, pour replacer cette interface dans le contexte modulaire de la PKS, la disposition des DD et de leur domaine flanquant a été étudiée par SAXS.

Toujours concernant les DD, toutes les études structurales précédentes se sont portées sur une à deux interfaces maxima d'un même système PKS *cis*-AT (Broadhurst et al., 2003 ; Buchholz et al., 2009 ; Whicher et al., 2013). Aucune étude complète de toutes les interfaces d'un même système n'a été menée à ce jour. Cependant, ceci permettrait de mieux appréhender les éléments clés responsables de la spécificité d'interaction et du bon ordonnancement des sous-unités de la chaîne d'assemblage. Ces informations sont essentielles en vue de construire des PKS chimères résultantes de l'échange de modules et de sous-unités. Afin de remédier à ce manque, nous nous sommes penchés sur les six interfaces de l'enacyloxine synthase. Cette PKS est le fruit d'un brassage génétique inhabituel, puisqu'elle est seulement le deuxième système

PKS hybride *cis-/trans*-AT caractérisé à ce jour (Mahenthiralingam et al., 2011). Les six premières sous-unités sont de nature *cis*-AT et seule la dernière est de nature *trans*-AT. De plus, l'enacyloxine est une molécule à activité antibiotique prometteuse puisqu'elle est efficace contre *Acinetobacter baumannii*, une bactérie multi-résistante, responsable d'infections nosocomiales sévères (Howard et al., 2012).

Les objectifs de ce projet étaient nombreux : i) déterminer la limite des DD potentiels aux différentes interfaces, ii) mettre en évidence ou non leur interaction par ITC, iii) caractériser structuralement ces domaines en combinant CD, SLS, SAXS et RMN et éventuellement déterminer la structure adoptée par certains de ces complexes par RMN et iv) ainsi déterminer de manière plus précise la classe à laquelle ces DD appartiennent ainsi que les résidus impliqués dans cette interaction, et enfin, v) déterminer le rôle possible joué par les domaines adjacents aux DD dans la formation d'un complexe stable.

Une autre approche d'ingénierie peu exploitée et pourtant d'intérêt certain, consiste en l'utilisation d'enzymes post-PKS pour modifier et complexifier des molécules existantes. Les activités de ses enzymes sont extrêmement diverses et les modifications apportées confèrent souvent leur activité biologique aux polycétides (Rix et al., 2002). Elles constituent des cibles privilégiées puisqu'elles agissent en fin de synthèse, et donc théoriquement leur utilisation et leur modification entraînent moins de problèmes de spécificité de substrat. La PKS synthétisant la lankacidine C, un antibiotique utilisé en médecine humaine et vétérinaire (Hayashi et al., 1988), produit un intermédiaire polycétidique linéaire non fonctionnel. C'est une enzyme post-PKS, LkcE, qui permet la macrocyclisation du polycétide (Arakawa et al., 2005), une activité intéressante puisque cette réaction est très difficile à réaliser en chimie traditionnelle. De plus, cette enzyme serait une monoamide oxydase, seulement la deuxième mise en évidence (Walsh & Wencewicz, 2014), et présentant une deuxième activité de macrocyclase par une réaction de Mannich intramoléculaire.

Jonathan DORIVAL, un doctorant ayant soutenu sa thèse au sein de notre équipe, avait pour objectif de réaliser une étude structure/fonction de cette enzyme. Pour cela, il a mis au point le protocole de production et de purification de l'enzyme et obtenu des structures par cristallographie et diffraction des rayons X de l'enzyme dans trois états : sans substrat, en présence de deux analogues mimant chacun une partie du substrat naturel, et enfin en présence du substrat natif. Ce substrat était extrait par nos soins à partir d'une souche de *S. rochei* dont

le gène *lkce* a été supprimé, afin d'accumuler le substrat linéaire. La molécule était ensuite purifiée par une équipe de collaborateurs chimistes située à l'université d'Hanovre (Pr. Andreas KIRSCHNING). Jonathan a ainsi pu mettre en évidence plusieurs résidus potentiellement impliqués dans la catalyse de la macrocyclisation et il a pu débiter le clonage et la production de certains de ces mutants avant de finir son doctorat.

Ainsi mon rôle dans cette étude était de finaliser le projet : poursuivre le clonage et la production des mutants, et mettre au point un test enzymatique permettant de déterminer les paramètres cinétiques de l'enzyme et ainsi prouver l'implication de certains résidus dans la catalyse. L'ensemble de ces informations permettraient de proposer un mécanisme réactionnel de l'enzyme et faciliterait par la suite son utilisation dans le cadre d'expériences de biologie synthétique.

RÉSULTATS

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

Dorival, J., Annaïval, T., **Risser, F.**, Collin, S., Roblin, P., Jacob, C., Gruez, A., Chagot, B. & 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 détaillée de la structuration et de la dynamique des domaines du module 5 de la virginiamycine synthase (Davison et al., 2014), notre équipe s'est intéressée de plus près au rôle joué par le petit domaine situé à l'extrémité C_{ter} de ce module, se structurant possiblement en hélices α . Nous sommes partis de l'hypothèse qu'il pouvait s'agir d'un DD (Broadhurts et al., 2003 ; Buchholz et al., 2009 ; Whicher et al., 2013). La sous-unité suivante, VirFG, présente un petit domaine analogue non assigné en N_{ter}, en amont de la KS. Dans un premier temps, les deux DD potentiels d'environ trente-cinq résidus chacun, ont été clonés et produits afin de caractériser leur interaction par gel filtration analytique (S75 10/300) puis par RMN (HSQC ^1H , ^{15}N). L'interaction a ensuite été confirmée et mesurée par ITC, une constante de dissociation de $5,8 \pm 0,2 \mu\text{M}$ a été mesurée ainsi qu'une stœchiométrie de 1:1. Des affinités et stœchiométries similaires ont également été mesurées entre les constructions ACP-VirA $^{\text{C}}$ DD et VirFG $^{\text{N}}$ DD, de même qu'entre VirA $^{\text{C}}$ DD et VirFG $^{\text{N}}$ DD-linker. Cette dernière construction comprend le DD et le *linker* le reliant à la KS₆, la construction comportant aussi la KS₆ n'était pas soluble. Ces expériences démontrent que les domaines ou les *linkers* flanquant les DD ne semblent pas intervenir dans l'interaction entre ces deux sous-unités.

Par la suite, la structure d'une protéine de fusion entre les deux DD a été résolue par RMN, les deux partenaires ont été liés par un *linker* de huit résidus (Gly₃Ser)₂ (Whicher et al., 2013). Cette structure révèle que les DD sont composés chacun de deux hélices α qui interagissent entre elles avec une topologie unique pour une interaction protéine-protéine, révélant une différence considérable entre cette interface de PKS *trans*-AT et celles déjà caractérisées chez les PKS *cis*-AT. Une autre caractéristique notable de cette interface réside dans le caractère intrinsèquement désordonné du VirFG $^{\text{N}}$ DD. En effet le spectre CD de cette protéine est caractéristique d'une protéine peu structurée. D'autres expériences de CD ont

permis de mettre en évidence une capacité de ce partenaire à subir un repliement induit en présence du VirA^cDD. Ce dernier, au contraire, présentait un spectre CD caractéristique d'une protéine majoritairement structurée en hélices α . Ces résultats ont été confortés par des expériences SAXS menées sur le VirFG^NDD : le *Kratky plot* est caractéristique d'une protéine plutôt allongée, donc probablement désordonnée, présentant toutefois des éléments de structures secondaires.

Toutes les données SAXS de cette étude ont été enregistrées au synchrotron SOLEIL sur la ligne SWING qui dispose d'un système d'injection couplé à un système de filtration sur gel 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. Par ailleurs, les données sont enregistrées directement à plusieurs concentrations dans un pic de sortie de chromatographie, 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).

Grâce à la structure RMN de la protéine de fusion, trois couples de résidus polaires semblant importants pour l'interaction ont été mis en évidence : ViFG^NDD Lys4/VirA^cDD Glu6954, ViFG^NDD Lys11/VirA^cDD Asn6949, ViFG^NDD Arg17/VirA^cDD Asp6940. Par alignements de séquences, il est démontré que la nature polaire de ces résidus est conservée parmi les DD pouvant appartenir à la même classe que les DD de VirA et VirFG, bien que le code électrostatique puisse varier au sein de ces couples potentiels. Par la suite nous avons muté chacun des six résidus polaires impliqués, soit en petit acide aminé hydrophobe soit en un résidu de charge opposé. Nous avons constaté que les mutants du VirFG^NDD ne présentaient pas les mêmes spectres CD que la protéine sauvage, ils perdaient en partie leur caractère désordonné au profit d'une pré-structuration en hélice α . Il est donc devenu impossible de déterminer si ces mutants entraînaient une perturbation de l'interaction ou bien s'ils perdaient leur capacité à se replier en présence de leur partenaire. Les cinq mutants du VirFG^NDD n'ont donc pas été utilisés pour la suite de l'étude. Au contraire, les mutants de VirA^cDD présentaient le même spectre CD que la protéine sauvage et ont donc pu être utilisés : nous avons ainsi pu constater une diminution de l'affinité pour chaque mutant du VirA^cDD avec son partenaire sauvage VirFG^NDD. Nous en avons conclu que l'interaction entre les deux DD est bien médiée par ces interactions polaires mais elles sont aussi possiblement stabilisées par un *cluster* d'acides aminés hydrophobes : Leu6939, Leu6943, Val6946, Ile6953, Ala6956, Val6957, et Leu6960 pour le VirA^cDD et Ala3, Leu7, Phe10, Ala20, et Leu24 pour le VirFG^NDD. La surface d'interaction entre les deux partenaires est

d'environ 730 Å², ce qui est comparable avec les autres classes de DD : 1000 à 1100 Å² pour les DD de classe 1a et 1b respectivement, et environ 1800 Å² pour les DD de classe 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 et la surface d'interaction serait alors de 1460 Å² 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, les constantes de dissociations sont comprises entre 2 et 24 µM.

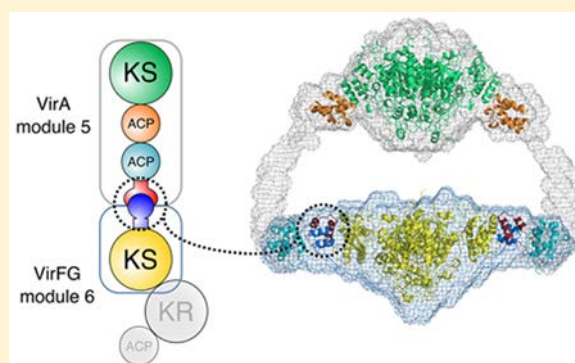
Des expériences par SAXS ont également été menées sur différentes constructions afin d'identifier la position des domaines flanquant. Ainsi, nous avons résolu la structure à faible résolution de la protéine de fusion ACP_{5b}-VirA^{CDD}/VirFG^{NDD}-*linker*-KS₆. Les domaines de docking adoptent une position centrale et sont entourés de part et d'autre par l'ACP_{5b} et la 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 obtenues précédemment sur le tridomaine KS₅-ACP_{5a}-ACP_{5b} et celles de la fusion ACP_{5b}-VirA^{CDD}/VirFG^{NDD}-*linker*-KS₆. En effet, les deux enveloppes se superposent parfaitement au niveau de l'ACP_{5b}, commun aux deux constructions. C'est la première fois qu'une interface de communication entre sous-unité de *trans*-AT PKS est reconstituée.

Dans ce travail collaboratif au sein de l'équipe, j'ai participé à la caractérisation des DD sauvages, à la production et à la purification des mutants, et réalisé les expériences de CD et d'ITC avec ces derniers. Les expériences et structures RMN ont été réalisées par les Dr. Jonathan DORIVAL, Thibault ANNAVAL et B. CHAGOT. La partie SAXS a été réalisée par le Dr. Jonathan DORIVAL avec le Dr. Arnaud GRUEZ. Sabrina COLLIN nous a aidés pour la purification de la fusion ACP_{5b}-VirA^{CDD}/VirFG^{NDD}-*linker*-KS₆.

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

S 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

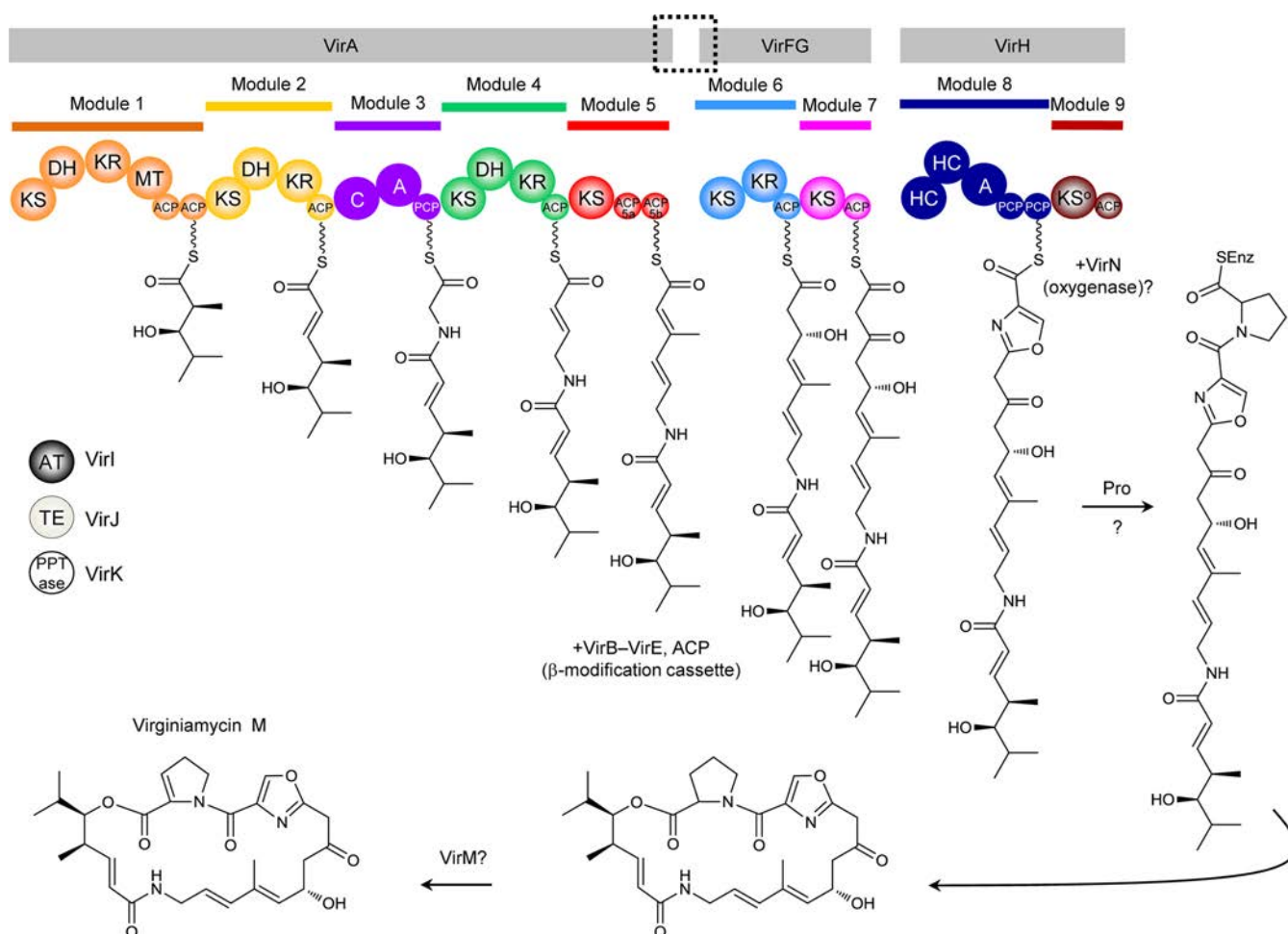


Figure 1. Virginiamycin (Vir) *trans*-AT polyketide synthase. The PKS comprises three subunits, VirA, VirFG (whose N-terminal sequence was recently revised by us),²¹ and VirH as well as additional modules that were not identified in the initial cluster description⁴ for starter unit selection and introduction 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

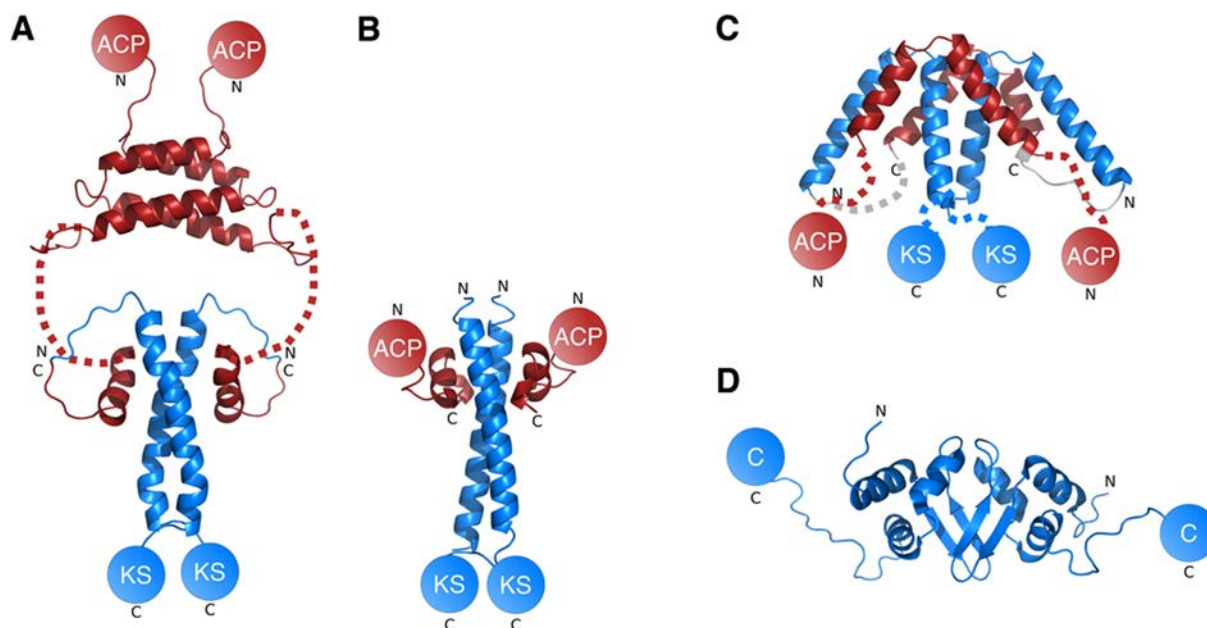


Figure 2. Three classes of docking domains from *cis*-AT PKSs. (A) NMR structure of a complex of covalently fused Class 1a docking domains (Protein Databank (PDB): 1PZQ, 1PZR).¹⁵ The C-terminal DDs (red) comprise three α helices, of which the first two form a four α -helical bundle that acts as a dimerization 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: 4MYJ).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 (Δ*H*), 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 μ L of protein sample at 100 μ M was used for all the measurements except that of 10 μ M VirA^{CDD} (100 μ L of sample were used in a 0.1 cm cuvette). Measurements were made at least in triplicate, and sample spectra were corrected for buffer background by subtracting the average spectrum of buffer alone. 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 μ L each at 100 μ M) 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 μ L at 100 μ M), 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 $\text{\AA}^{-1}$. 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 μ L 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 (Schrödinger, 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 $0.3 \text{ \AA}^{-1}$ ($\chi^2 = 6.61\text{--}10.39$). 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 μM). 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 $0.3 \text{ \AA}^{-1}$) 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 $66 \text{ \AA}$, and M_w s of 5 kDa (SAXS MoW) and $4.2 \pm 0.4 \text{ kDa}$ (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 = $160 \text{ \AA}$, 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

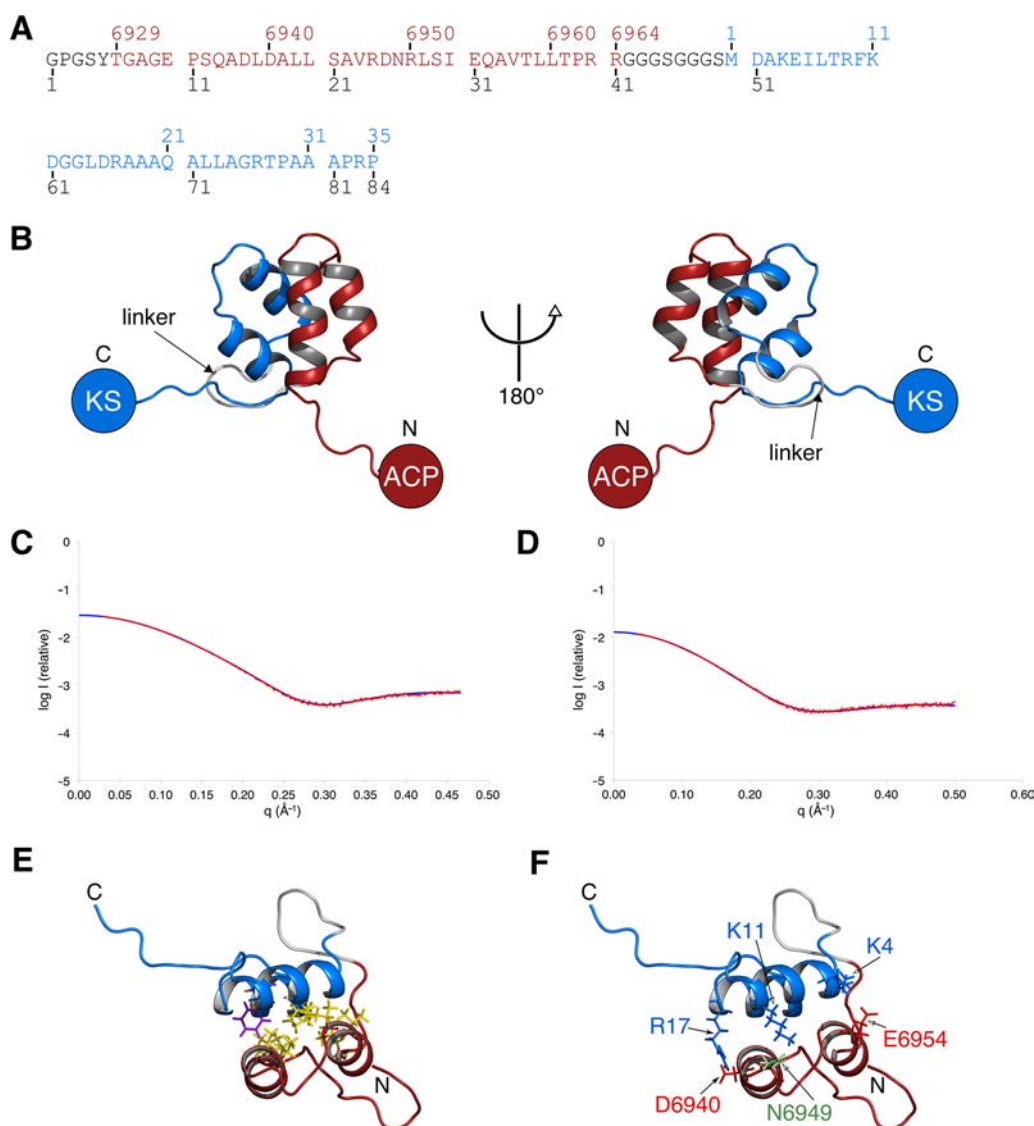


Figure 3. Analysis of docking between VirA^{CDD} and VirFG^{NDD}. (A) Sequence of the VirA^{CDD}–VirFG^{NDD} fusion construct, with VirA^{CDD} indicated in red, and VirFG^{NDD} indicated in blue. The N-terminal vector-derived residues and the added 8-residue linker are shown in black. The numbering above 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 (CRYSOLOG⁵⁰) (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 ~ 730 Å², an extensive surface area given the limited number of participating residues (17 amino acids within a 4 Å 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 ~ 1460 Å², which is in between Class 1 (~ 1100 Å²) and Class 2 (~ 1800 Å²) 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 ($K_d = 5.8$ μ M) by factors of ca. 15 (85 ± 23 μ M) and 4-fold (20 ± 2 μ 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 ± 16 μ 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 μ 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

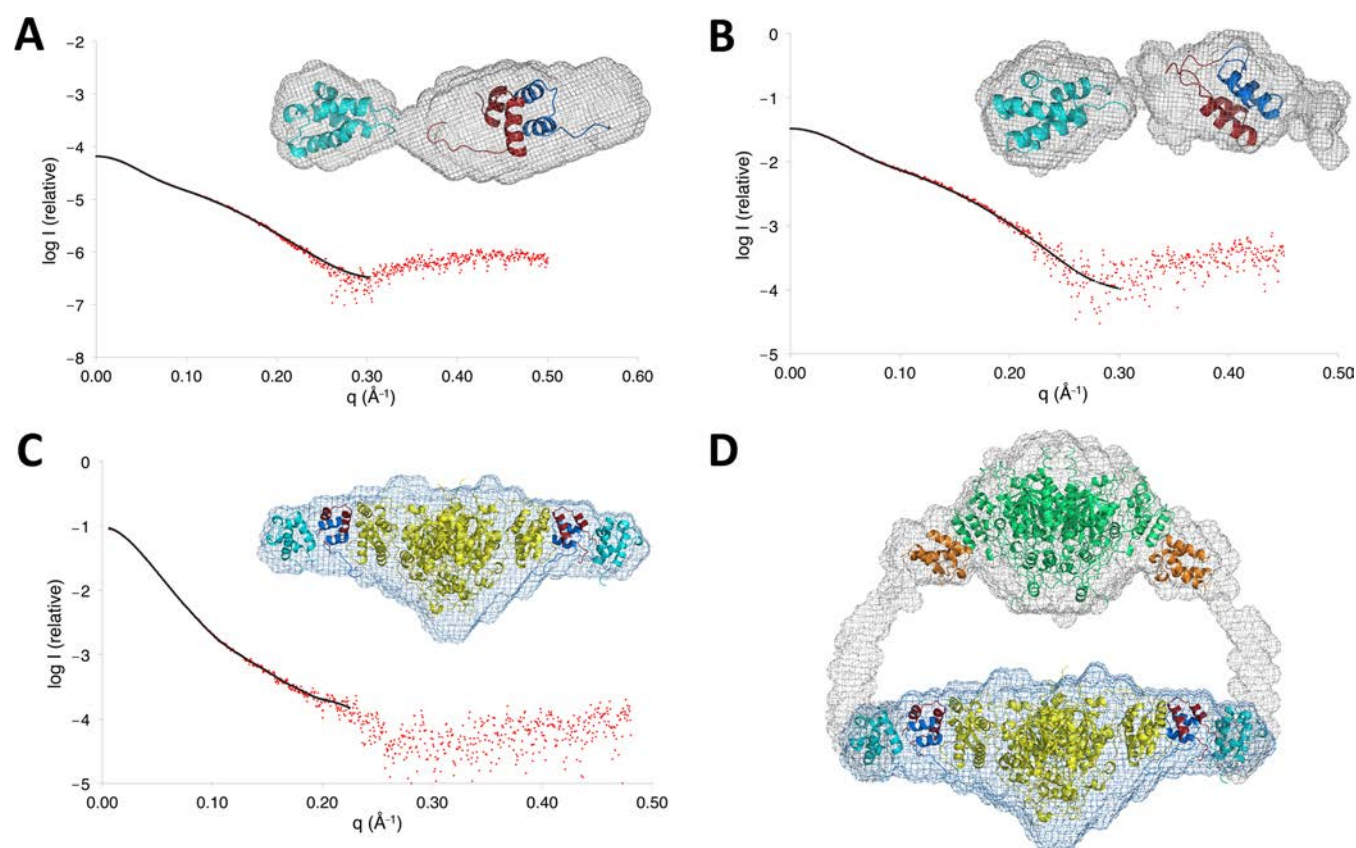


Figure 4. Analysis of the interface between subunits VirA and VirFG by SAXS. (A) Analysis by SAXS of the VirA ACP_{sb}-CDD/VirFG^{NDD} complex. Fit between the *ab initio* model computed with DAMMIN⁴² (solid black line) and the experimental SAXS data (red dots). This analysis yielded a molecular envelope 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 (data from ref 21), 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_{sb}-^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_{sb}-^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_{sb} does not contact it directly. Analysis of VirA ACP_{sb}-^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_{sb} 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_{sb} 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_{sa}, and ACP_{sb} domains (KS₅–ACP_{sa}–ACP_{sb})²¹ (Figure 4D). This analysis shows that the two sets of ACP_{sb} domains superimpose nearly perfectly and therefore that the global form observed for VirA ACP_{sb}-^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_{sb} to ACP_{sb} in module 5 alone = 279 Å²¹ vs 200 Å in KS₅–ACP_{sa}–ACP_{sb}), 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 PKSs.

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_{sb} 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_{sb} 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.; Götz, M.; Ort-Winkelbauer, 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-Lazetić, G. M.; Mitić, N. S.; Kovačević, J. J.; Obradović, Z.; Malkov, S. N.; Beljanski, M. V. *BMC Bioinf.* **2011**, *12*, 66.

SUPPLEMENTARY INFORMATION

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 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

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

CONTENTS

SUPPLEMENTARY METHODS

Table S1. Primers used in this study.

Representative analysis of SAXS data obtained on construct VirA^{CDD}–VirFG^{NDD}.

SUPPLEMENTARY RESULTS

Figure S1. Comparative sequence analysis of VirFG^{NDD} and Class 2 docking domains.

Table S2. Analysis of intersubunit interfaces in *trans*-AT PKS.

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

Table S3. Putative docking domain families in *trans*-AT PKS.

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

Table S4. Sequences of constructs investigated in this study.

Figure S4. Analysis by gel filtration of the interaction between VirA^{CDD} and VirFG^{NDD}.

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

Figure S6. NMR studies of VirA^{CDD} and VirFG^{NDD}.

Figure S7. Analysis by SAXS of the various constructs.

Table S5. Summary of SAXS data.

Figure S8. Analysis by circular dichroism of wild type and mutant docking domains.

Table S6. NMR and refinement statistics for VirA^{CDD}–VirFG^{NDD}.

Table S7. Putative interface residues in the VirA^{CDD}/VirFG^{NDD} docking domain families.

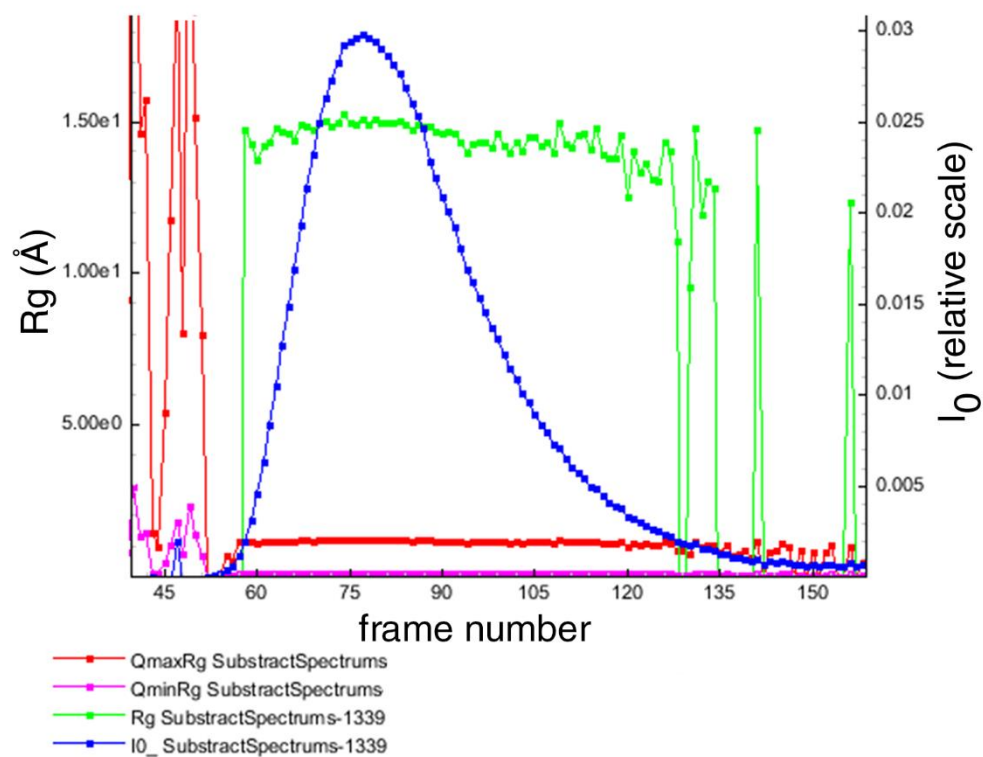
SUPPLEMENTARY REFERENCES

SUPPLEMENTARY METHODS

Table S1. Primers used in this study.

Constructs	Primer type	Oligonucleotides (5'-3') ^a
VirA ^C DD	f	TTTCCGGGATCCTATACCGGCGCCGGGGAGCCCTC
	r	TTTCGAAGCTTTTCAGCGCCGCGGGGTGAGCAG
VirFG ^N DD	f	TTTCCGGGATCCATGGACGCGAAAGAGATCCTGACCCGATTCAAG
	r	TTTCGAAGCTTTCAATACGGGCGCGGTGCGGCCG
VirFG ^N DD-linker	f	TTTCCGGGATCCATGGACGCGAAAGAGATCCTGACCCGATTCAAG
	r	TTTCGAAGCTTTTCAGGGCCGCGCCGCGCCGT
VirA ACP _{5b} - ^C DD	f	TTTCGCGGATCCGCCGTGGCCGTCGACCCCG
	r	TTTCGAAGCTTTTCAGCGCCGCGGGGTGAGCAG
VirFG ^N DD-KS ₆	f	TTTCCGGGATCCATGGACGCGAAAGAGATCCTGACCCGATTCAAG
	r	TTTCGAAGCTTTTCAGCGTTTCGACGTAGTGGCGGTGCGG
	BamHI f	GACACGTTCTGGCAGCGCATCCTCGACGGCGAC
	BamHI r	GTCGCCGTCGAGGATGCGCTGCCAGAACGTGTC
VirA ^C DD-VirFG ^N DD (covalent fusion)	f	TTTCCGGGATCCTATACCGGCGCCGGGGAGCCCTC
	r	TTTCGAAGCTTTTCAGGGGCGCGGTGCGGCCG
	Overlap1	GGAGGTAGTGGTGGAGGCAGTATGGACGCGAAAGAGATCCTGACCCGATTCAAG
	Overlap2	GCCTCCACCACTACCTCCCCCGCGCCGCGGGGTGAGCAG
VirA ACP _{5b} - ^C DD-VirFG ^N DD (covalent fusion)	f	TTTCGCGGATCCGCCGTGGCCGTCGACCCCG
	r	TTTCGAAGCTTTTCAGGGGCGCGGTGCGGCCG
	Overlap1	GGAGGTAGTGGTGGAGGCAGTATGGACGCGAAAGAGATCCTGACCCGATTCAAG
	Overlap2	GCCTCCACCACTACCTCCCCCGCGCCGCGGGGTGAGCAG
VirA ACP _{5b} - ^C DD-VirFG ^N DD-KS ₆ (covalent fusion)	f	TTTCGCGGATCCGCCGTGGCCGTCGACCCCG
	r	TTTCGAAGCTTTTCAGCGTTTCGACGTAGTGGCGGTGCGG
	Overlap1	GGAGGTAGTGGTGGAGGCAGTATGGACGCGAAAGAGATCCTGACCCGATTCAAG
	Overlap2	GCCTCCACCACTACCTCCCCCGCGCCGCGGGGTGAGCAG
VirA ^C DD D6940A	f	TCCCAGGCGGACCTGGCCGCGCTGCTCAGCGCC
	r	GGCGCTGAGCAGCGCGGCCAGGTCCGCCTGGGA
VirA ^C DD D6940R	f	TCCCAGGCGGACCTGCGCGCGCTGCTCAGCGCC
	r	GGCGCTGAGCAGCGCGCGCAGGTCCGCCTGGGA
VirA ^C DD L6943A	f	GACCTGGACGCGCTGGCCAGCGCCGTCCGCGAC
	r	GTCGCGGACGCGCTGGCCAGCGCGTCCAGGTC
VirA ^C DD N6949A	f	AGCGCCGTCCGCGACGCCCGGCTGTCCATCGAG
	r	CTCGATGGACAGCGGGCGTCTGCGGACGGCGCT
VirA ^C DD N6949K	f	AGCGCCGTCCGCGACAAACGGCTGTCCATCGAG
	r	CTCGATGGACAGCGTTTGTCTGCGGACGGCGCT
VirA ^C DD E6954A	f	AACCGGCTGTCCATCGCGCAGGCGGTACCCCTG
	r	CAGGGTGACCGCCTGCGCGATGGACAGCCGGTT
VirA ^C DD E6954K	f	AACCGGCTGTCCATCAAGCAGGCGGTACCCCTG
	r	CAGGGTGACCGCCTGCTTGATGGACAGCCGGTT
VirFG ^N DD K4A	f	GGATCCATGGACGCGGAGAGATCCTGACCCGA
	r	TCGGGTCAAGATCTCTGCCGCGTCCATGGATCC
VirFG ^N DD K4E	f	GGATCCATGGACGCGGAAGAGATCCTGACCCGA
	r	TCGGGTCAAGATCTCTTCCGCGTCCATGGATCC
VirFG ^N DD K11A	f	ATCCTGACCCGATTTCGCGGACGGGGGCTGGAC
	r	GTCCAGGCCCCCGTCCGCGAATCGGGTCAGGAT
VirFG ^N DD R17A	f	GACGGGGGCTGGACGCGCGCCGCCAGGCGG
	r	CGCTGGGCGGCGCGCGTCCAGGCCCCGTC
VirFG ^N DD R17D	f	GACGGGGGCTGGACGACGCCGCCAGGCGG
	r	CGCTGGGCGGCGCGCTCTCCAGGCCCCGTC

^aIntroduced restriction sites are underlined, while the added Tyr is indicated by bold letters.



Representative analysis of SAXS data obtained on VirA^CDD–VirFG^NDD. Shown are the variation in R_g and $I(0)$ after gel filtration chromatography, as a function of the frame number.

A

B

[illegible]

C

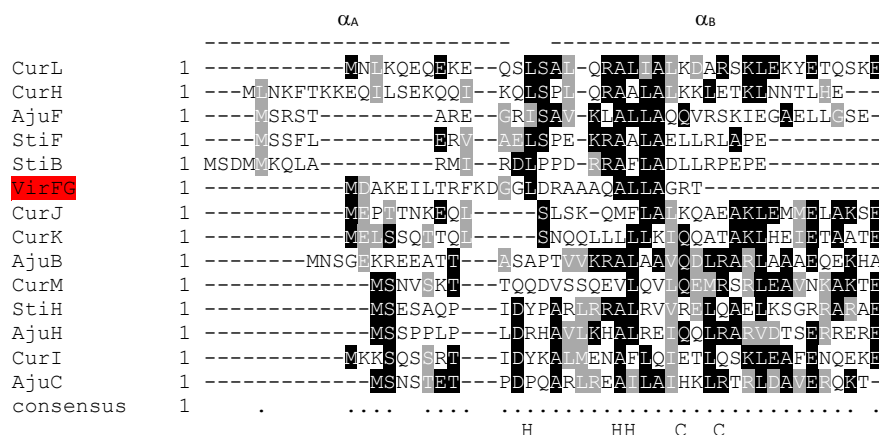


Figure S1. Comparative sequence analysis of VirFG^{NDD} and Class 2 docking domains. (A) Sequence of subunit VirFG upstream of the conserved KS domain. The first 28 residues (in red) have been designated as VirFG^{NDD}. (B) Multiple alignment using ClustalOmega¹ of the C-terminal regions of PKS subunits from *trans*-AT PKSs (in red), with those of *cis*-AT PKS² (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^{NDD}) from the *trans*-AT PKSs 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 *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.² (C) Multiple alignment using ClustalOmega¹ of the N-terminal regions from the partner *cis*-AT PKS subunits to those shown in (B), alongside that of VirFG (in red). The positions of the two 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, ajudazol (*cis*-AT);¹¹ Ela, elansolid (*trans*-AT);¹² Vir, virginiamycin (*trans*-AT);¹³ Sna, pristinemycin (*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 ^a
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
C	N-MT	✓	Ttt
PCP	KS		Ozm, Nsp, Sna, Tdp(2)
O-MT	ACP	✓	Ozm
C-MT	ACP	✓	FR9, Riz
Ox	KR	✓	FR9, Tst
E	C	✓	Tdp

^aThe 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, FkbH-like domain; ACP, acyl carrier protein; C, condensation; HC, heterocyclization; ER, enoyl reductase; PS, pyrane synthase; A, adenylation; GNAT, Gcn5-related N-acetyltransferase; PCP, peptidyl carrier protein; O-MT, O-methyltransferase; Ox, oxidase; E, epimerase; °, inactive.

Key to *trans*-AT PKS systems:

Atc, anthracimycin;^{15,16} Chi, chivosazol;¹⁷ Cor, corallopyronin;¹⁸ Ela, elansolid;¹² FR9, FR901464;¹⁹ Kir, kirromycin;²⁰ Mmp, mupirocin;⁶ Ooc, oocydin;²¹ Ozm, oxazolomycin;²² Rhi, rhizoxin;⁵ Sna, pristinamycin;¹⁴ Tai, thailandamide;⁹ Tdp, thailandepsin;²³ Tmp, thiomarinol;²⁴ Vir, virginiamycin;¹³ 592, enacyloxin;²⁵ Cal, calyculin;²⁶ Nsp, nosperin;²⁷ Onn, onnamide;²⁸ Ped, pederin;²⁹ Psy, psymberin;³⁰ Bry, bryostatin;³¹ Etn, etnangien;³² Ta, myxovirescin (antibiotics TA);¹⁰ Bae, bacillaene;³³ Bat, kalimantacin/batumin;³⁴ Dis, disorazol;^{35,36} Riz, rhizopodin;³⁷ Dif, diffidin;³⁸ Lnm, leinamycin;³⁹ Lum, luminaolide;⁴⁰ Mis, miskinolide;⁴⁰ Mln, macrolactin;⁴¹ Sor, sorangicin;⁴² Tst, thailanstatin;⁴³ Bon, bongkrekeic acid;⁴⁴ Tto, tolytoxin;⁴⁰ Lac, lactimidomycin;⁴⁵ Lkc, lankacidin;⁴⁶ Mgs, migrastatin;⁴⁷ Cto, chlorotonil;¹⁶ Alb, albidin;⁴⁸ Ttt, thalassospiramide.⁴⁹

ChiB_ACP	1	-----LSSFAADVRLPGAAKIDTAALFG-----	23
TaiL_ACP	1	--ASELAALARRRAGGAPLPDAEAAASATTATTASASGTAHAPNAA	42
OzmM_ACP	1	-----AATGAAGTAAAGTAAAGPARA	19
OzmH_ACP	1	-----DALLGPAAQEP-	11
ChiE_ACP	1	-----VDHHE-AALRASEPPEGVGGGAEFVRPRANRPERPERADPP--	39
ChiC_ACP	1	-----ERDPALAEIKAVQLRAQGA-	19
KirAV_ACP	1	---AGRFDGGTGG--ADDLPALFADFDRLA---	25
5925_ACP	1	-----LAANAAPAG-	8
ViaG_ACP	1	-----AGR-AAVPVA-	9
KirAiif_ACP	1	HDKYGPLPATATPEPPPAPVTSSATAPEPTSVPVASSAAAPIPA-	46
SnaE2_ACP	1	-----TSR-TAPQFA-	9
TaiE_ACP	1	-----GTQL-GARLFAAQARAGAF--AGV--	22
AtcE_ACP	1	---HAG-HAAAPPGA-DGA-	14
DisB_ACP	1	-----SRR-GARAPESRAKARSGA-AGA-	21
Ta-1_ACP	1	-----GN-	2
RhiC_ACP	1	-----ADNQRTAMAVSPN-	13
RizC_ACP	1	---RALV-TEAPRGA-	12
TaiD_ACP	1	-----AERIARDG--GRRAAEARAPDAAMAS--DVA-	28
5923_ACP	1	-----AR--GALAE-LFA-	10
EtnH_ACP	1	---CAEHDDLLARRQAG-	14
BaeM_ACP	1	-----LTEHKETIETFYRTE-	15
MlnD_ACP	1	-----SSI-RNR--SAD-	9
OzmK_ACP	1	-----	
BryA_ACP	1	-----RKQIIIS-NDE-KQH-	12
Bat1_ACP	1	-----VRIVSAQPREAS-SARTLDVVVGGERLDTQ-SSD-	31
ElaK_ACP	1	-----SSEMT-NISTY-EPV-	13
BryB_ACP	1	-----KQIEQR-	7
Consensus aa:		 h	
Consensus ss:		 h	

			α1	
ChiB_ACP	24	--DGAAPAPAPA-----PAPVD-EPRGGEPEG-----D-----E PLWR	53	
TaiL_ACP	43	GDAHAPNAAGARAATPA P AENAA GAVHAAR ANPAN PAT--LAN PAPGG-----G-----V PLDD	95	
OzmN_ACP	20	A---E PEAGIEQLVDSLSDTEVD -----	39	
OzmH_ACP	12	----PAP-----PAPPAS--SASSASSG-----T-----S APPD	34	
ChiE_ACP	40	----ERAGVAPGRRAPSP AAAVERFDGA--VPED--LAGEVARM--SD--AEVR	83	
ChiC_ACP	20	----PGP-----GAPE-----LAVG-----H----- DELWR	36	
KirAV_ACP	26	-----AGL TEAATDEVFDRDRVVRVRLGLLTAL -----SPATP-----EA PPD	62	
5925_ACP	9	-----SAVEAAs-----VAPAVEAT-----EPAA-----APLDT	33	
VLA _ACP	10	-----AAAEPA-----LPLPVTGAG-----EPSQ-----AD LA	74	
KirAIi_ACP	47	-----PVAPSAA-----APAFVASSAA AEPPPAPVA--ADAVG	38	
SnaE2_ACP	10	-----GAVPAG-----SVSA-----AD LDA	24	
TaiE_ACP	23	-----EPGEPDARAL PAARAA--APPAH--TD--AAAH--TDTA	55	
AtcE_ACP	15	-----G-----PERA-----TD VED	24	
DisB_ACP	22	-----DLST	25	
Ta-1_ACP	3	-----EIQTTSVAG-----APAQGE-----P-----PS LDE	23	
RhiC_ACP	14	-----DLDHVASG-----NNVRG-----QSRPI EIHPPFADQ	41	
RicZ_ACP	13	-----TQSANGH-----ALE-----CMSD-----GN LEE	31	
TaiD_ACP	29	-----EASVVSEATEASDASEASDASEAS--EASEASEA-----SKAP-----AD LAA	69	
5923_ACP	11	-----FAIVAAPAQ--LAN--AAADASLG--DHSa-----AELAR	39	
EtnH_ACP	15	-----TQANG-----GA-----VLER-----RAVEA	30	
BaeM_ACP	16	-----ETETEAA--APESK-----EYTD-----QE IIA	56	
MlnD_ACP	10	-----KPYNGNFREKS--GFGLEEIAYTLTG R--EPFECLRALITCSNI GD	31	
OzmK_ACP	1	-----EAE	3	
BryA_ACP	1	-----QPSISTIFP-----TS LDE	26	
BatI_ACP	32	-----DRT--MKMLDFRDG--NVKIEEV EGILDRW--TKTQS--SDDR TMK	70	
ElaK_ACP	4	-----QSLQQEF-----SLDD	24	
BryB_ACP	18	-----NDIQNLSSSSSYFSEDN-----QEVA-----LS LKE	32	
Consensus aa:		 s . h . .		
Consensus ss:		hhhhhh		hhh

S7

ChiB_ACP		-----	
TaiL_ACP		-----	
OzmN_ACP		-----	
OzmH_ACP		-----	
ChiE_ACP		-----	
ChiC_ACP	70	MV-----	71
KirAV_ACP	88	EDR-----	90
5925_ACP		-----	
VirA_ACP		-----	
KirAII_ACP		-----	
SnaE2_ACP		-----	
TaiE_ACP	86	PLAQP-----	90
AtcE_ACP		-----	
DisB_ACP		-----	
Ta-1_ACP		-----	
RhiC_ACP	72	I-----	72
RizC_ACP		-----	
TaiD_ACP		-----	
5923_ACP		-----	
EtnH_ACP		-----	
BaeM_ACP		-----	
MlnD_ACP	118	EDLYEEELKPVL SL PLY PFS KNRHHW IHL NKDEA Q KAEAFSNT ELH NSNKDTADGEIMSLLEKA Q TGEMN	187
OzmK_ACP		-----	
BryA_ACP	54	DMELHELL-----	61
Bat1_ACP		-----	
ElaK_ACP		-----	
BryB_ACP		-----	
Consensus aa:			
Consensus ss:			

ChiB_ACP		-----	
TaiL_ACP		-----	
OzmN_ACP		-----	
OzmH_ACP		-----	
ChiE_ACP		-----	
ChiC_ACP		-----	
KirAV_ACP		-----	
5925_ACP		-----	
VirA_ACP		-----	
KirAII_ACP		-----	
SnaE2_ACP		-----	
TaiE_ACP		-----	
AtcE_ACP		-----	
DisB_ACP		-----	
Ta-1_ACP		-----	
RhiC_ACP		-----	
RizC_ACP		-----	
TaiD_ACP		-----	
5923_ACP		-----	
EtnH_ACP		-----	
BaeM_ACP		-----	
MlnD_ACP	188	MKETSELIEELLFHE	202
OzmK_ACP		-----	
BryA_ACP		-----	
Bat1_ACP		-----	
ElaK_ACP		-----	
BryB_ACP		-----	
Consensus aa:			
Consensus ss:			

Tao_ER		GEGH RIGVL --QQPPRGRDCHEADDGSRRI-----PQPPVPVQVGRPGRG	42
KirAIII_C		-----	
Bat2_C		-----	
EtnI_ER	1	PEHSYEP----- RGEST -----	12
ChiD_HC	1	---QRP----- DRG ---	6
5924_KS	1	--NKPTS-----SDGWK-	10
RizD_HC	1	--PFK-----	3
OzmL_C		-----	
DisC_HC		-----	
SnaE3_KS	1	AEGPSRLPL--AEPDAV--	15
VicF _KS		-----	
TaiM_KS		-----	
TaiK_KS		-----	
ChiC_KS		-----	
AtcF_KS		-----	
OzmH_KS		-----	
RhiD_KS		-----	
ElaO_KS		-----	
TaiE_C	1	PRWRPTWLKR LSC PQPRHPHTPKHPHTPKHKPHKPHRHPKHPI ISRR CSSASARGKWISTTSSTWFDG	70
MlnE_KR	1	-----NEK-----	3
OzmJ_KS		-----	
BaeN_C		-----	
BryC_ER		-----	
5922_KS		-----	
ChiF_KS	1	-----AMT-----	3
KirAVI_KS		-----	
BryB_ER	1	-----VVVE-----E	5
Consensus aa:			
Consensus ss:			

		$\alpha 1$		$\alpha 2$			
TaO_ER	43	AVMTRLEILQALQS	----	GRMA	-----LADAKKALAAIEAQVQPED-----	79	
KirAIII_C	1	--SRALRDVLAAVA	----	GELP	-----DETAESLRGLRSQDPVPPEG----	36	
Bat2_C	1	--TRRDLLFAYKN	----	RSLT	-----TQELVAALTEINLTER----	33	
EtnI_ER	13	VKSQIEALLREYQS	----	GLLS	-----EABLRLARYAQIVASGSD-----	48	
ChiD_KS	7	--RRLRYMEKAVTD	----	AMF	-----DERSREEKLSVFRALRDLRLRTPPHG---GAADG	53	
5924_KS	11	--DDYLSRLSLRLSK	----	NQLM	-----ALALKLKQQQLGQPPAA-----	43	
RizD_KS	4	--NDSALNEFR	----	LLS	-----REEKLQRIAAALIEEKSSAKS-----AASSEA	42	
OzmL_KS	1	---REDLATSRK	----	ARLS	-----PQKRALLLEKLSGAARTS-----GTTTRI	36	
DisC_HK	1	---ESAMITQEFAN	----	LS	-----AEEKVQVLRLLDR-RASWQ-----AAPGEPAAIA	71	
SnaE3_KS	16	DRDRAALLRAVAG	QGRGPAAGATASARPA	PGVWD	RAFAAALLRAAAGERRPG	-GDGPAGGS-MGR	49
YnfP_KS	1	--D-----E-----T-----P-----	----	GC	D-----A-----Q-----AGRTPAAAPR--PSEPAAP-TRP	44	
TaiM_KS	1	ERQNLNRLVLEKLA	----	AAIG	-----PDEAKRMLSARAGDA-----ADK	37	
TaiM_KS	1	--RSSKEIFEARLD	----	GRLS	-----REEAHAAALRSARAAADALGASGAFEPRGAEA	48	
ChiC_KS	1	-KRDPRVIFEQLNK	----	QGIS	-----RAEAHRLLKARS DGR EQEATGPGEAEEGAADA	49	
AtcF_KS	1	--NSAVKRVLELVET	----	RQIS	-----PEEGRRLRLTALGTAGDRPVA-----GEAGDR	43	
OzmH_KS	1	-----	----	RRRS	-----MDDKRALLLKLLASVRRQQAAPA-----	26	
RhD_KS	1	SSQEQFSIYIAGLN	----	RQIT	-----VAAEQRLRLKKAASATL-GTEQ-----TEAADK	46	
ElaO_KS	1	--DSLAILKALST	----	GAIS	-----AGAAREALKQLAVAPVSVKPP-----	35	
TaiE_C	71	DHYNKKDILLLAYRE	----	GLLD	-----TGSAQVRLDALRLERSASAP-----	107	
MlnE_KR	4	-TMQIEQIVKKVKE	----	CSLT	-----PEEGLEI-----	28	
OzmJ_KS	1	SSANFRDILLGYVR	----	GAVG	-----AEQARGLLDAGRTPDPQDK-----EER	40	
BaeN_KS	1	--NHKELLDAYRS	----	GLTI	-----TAEVEQLQAFKPTT-----AKR	33	
BryC_ER	1	--KAKVIFQGLKS	----	NEIS	-----IEQAEKMLLSLDTESKHGHPESKLSDIKSDLIH	47	
5922_KS	1	TDMDKDLLLSQISQT	----	IREL	-----KTRLAQAEQGHHEA-----	32	
ChiF_KS	4	-LRSQDLMDMAK	----	-----	-----RESLVRLHQQKAN-----	28	
KirAIV_KS	1	--SSEDKRLDYLRR	----	VTAE	-----LQVTRRLRREVESRSK-----	42	
BryB_ER	6	FFVYSYRDLKALQ	----	EKIS	-----FEAKYKLIKRRDKDKSKQRLN-----HDREL	39	
Consensus aa:		...p...hhp.hpp	s.hs		l.....		
Consensus ss:		hhhhhhhhhhhh			hhhhhhhhhhhh		

S9

Figure S2. Identification of putative families of docking domains in *trans*-AT PKSs. (A) Secondary structure/sequence-based alignment by PROMALS3D⁵⁰ of VirA^CDD 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 α 1 and α 2 in VirA^CDD 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, pristinamycin; 592, enacyloxin; Ta-1, myxovirescin; Rhi, rhizoxin; Riz, rhizopodin; Etn, etnangien; Bae, bacillaene; Mln, macrolactin; Bry, bryostatin; Bat, kalimantacin/batumin; Ela, elansolid. (B) Secondary structure/sequence-based alignment by PROMALS3D⁵⁰ of VirFG^NDD 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 α 1 and α 2 in VirFG^CDD 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^CDD–VirFG^NDD 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/KirAVI ; OzmN/OzmH ; OzmH/OzmJ ; <i>RhiC/RhiD</i> (ACP–ACP/KS); <i>SnaE2/SnaE3</i> (ACP–ACP/KS); TaiE/TaiK (ACP–ACP/KS); TaiL/TaiM (ACP–ACP/KS); <i>VirA/VirFG</i> (ACP–ACP/KS); 5923/5922 ; 5925/5924
ACP	C	<i>BaeM/BaeN</i> ; <i>Bat1/Bat2</i> ; KirAII/KirAIII ; OzmK/OzmL ; TaiD/TaiE
ACP	HC	ChiC/ChiD ; <i>DisB/DisC</i> ; <i>RizC/RizD</i>
ACP	KR	<i>MlnD/MlnE</i>
ACP	ER	BryA/BryB ; BryB/BryC ; <i>EtnH/EtnI</i> ; <i>Ta-1/TaO</i>

^aThe host organisms of these 19 systems include six *Streptomyces* strains (*Streptomyces* strains CNH365/T676 (anthracimycin, Atc), *Streptomyces collinus* Tü 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 etnangien, Etn), *Chitinophaga pinensis* DSM2588 (Elansolid, Ela), *Sorangium cellulosum* So ce12 (disorazol, Dis), *Stigmatella aurantiaca* Sg a15 (rhizopodin, Riz) and *Myxococcus xanthus* DK1622 (myxovirescin, Ta)) 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 amyloliquefaciens* FZB42 (macrolactin, Mln)), and one *Pseudomonas* (*Pseudomonas fluorescens* BCCM_109359 (kalimanticin/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).

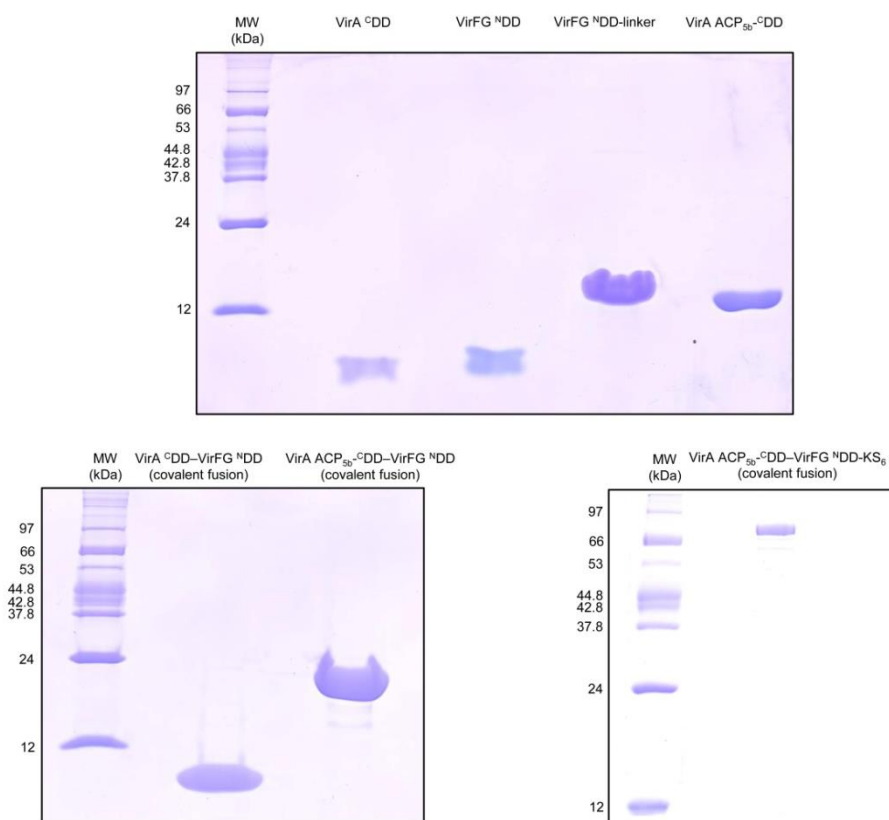


Figure S3. SDS-PAGE analysis of protein preparations used in this study. VirA^CDD (calc'd: 4.3 kDa); VirFG^NDD (calc'd: 4.1 kDa); VirFG^NDD-linker (calc'd: 7.2 kDa); VirA ACP_{5b}-^CDD (calc'd: 14.3 kDa); VirA^CDD-VirFG^NDD (covalent fusion) (calc'd: 8.4 kDa); VirA ACP_{5b}-^CDD-VirFG^NDD (covalent fusion) (calc'd: 18.4 kDa); VirA ACP_{5b}-^CDD-VirFG^NDD-KS₆ (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 ^C DD	GP GS ^Y GTGAGEPS QADLDALLSAVR DNRLS IEQAVTLLT PRR
VirFG ^N DD	GP GSMD AKEILTRFK DGGLD RAAAQALL AGRT PAAAPRP ^Y
VirFG ^N DD-linker	GP GSMDAKEILTRFKDGGLDRAAAQALLAGRT PAAAPRP SE PAAAPTRPAVP PAVP AVEPAA GTTVAAEGTAGRP
VirA ACP _{5b} - ^C DD	GP GS AVAVDPAPVARALREELARTLYCEPGDIDDEAS FNTLGLDSILGVEFVAFVNQTYGL DEKAGILYDHPSLAALSRHVAGRAAPV PAAAAEPAALPLPV
VirFG ^N DD-KS ₆	GP GSMDAKEILTRFKDGGLDRAAAQALLAGRT PAAAPRP SE PAAAPTRPAVP PAVP AVEPAA GTTVAAEGTAGRPEPVAVIGYSARFPGAADADTFWQRI LDGDDLVTEVPPERWRTEEFYDP DPAAEGRSVSRWGAYVADADRFDA DFRMTPREAELTDPQARLFLQEAWRALEHAGRDARS LAGTRCGVYAGV MLNDYQDLVERESPYKRLPQVMQNSNSILAARIAYHLDLKGPAVTVD T ACSSSLTALHLACQSLWLGETDLTVVGGVTLYLTELPHVFM SAAGMLSPNGRCRPFDAAD GIVPGE CAVVVLKPLSKALADGDPVHAVIRASGLNQDRGPQRGITAPSARSQTALVRD TL QRFPAVD PAGIDYVECHGTGTPLGDP IEVTALNEAFAGAGLAPASVP IGSVKGNIGHTSAA AGLAGLLKAAGVVRTGLVPPSLHYARANPQIPFDQGPFTVAGERRELGRPDGGRPRRATVS SFGFSGTNAYVVVEQAPEQAAR PAGDDGAPPVLVPLSGPARADAPAAHAADLARWLRGPGS DASPADVAHTLAVARTHHTYRSALLVSGRDELLES LDLLAAGSADPRRTDTAPDAAAPDAA VRRQAALLARLVERAGENPGPVELAALAKLYTQGH TVDWAAVSPPARHRRLSLPAYPFAP HRHYVER
VirA ^C DD–VirFG ^N DD	GP GS ^Y GTGAGEPS QADLDALLSAVR DNRLS IEQAVTLLT PRR GGGSGGG SM AKEILTRFKD GGLDRAAAQALLAGRT PAAAPRP
VirA ACP _{5b} - ^C DD–VirFG ^N DD	GP GS AVAVDPAPVARALREELARTLYCEPGDIDDEAS FNTLGLDSILGVEFVAFVNQTYGL DEKAGILYDHPSLAALSRHVAGRAAPV PAAAAEPAALPLPVTGAGEPSQADLDALLSAVR DNRLS IEQAVTLLT PRR GGGSGGG SM AKEILTRFKDGGLDRAAAQALLAGRT PAAAPRP
VirA ACP _{5b} - ^C DD–VirFG ^N DD-KS ₆	GP GS AVAVDPAPVARALREELARTLYCEPGDIDDEAS FNTLGLDSILGVEFVAFVNQTYGL DEKAGILYDHPSLAALSRHVAGRAAPV PAAAAEPAALPLPVTGAGEPSQADLDALLSAVR DNRLS IEQAVTLLT PRR GGGSGGG SM AKEILTRFKDGGLDRAAAQALLAGRT PAAAPRP SE PAAAPTRPAVP PAVP AVEPAA GT TTVAAEGTAGRPEPVAVIGYSARFPGAADADTFWQRI LDG DDL VTEVPPERWRTEEFYDP DPAAEGRSVSRWGAYVADADRFDA DFRMTPREAELTDPQA RLFLQEAWRALEHAGRDARS LAGTRCGVYAGVMLNDYQDLVERESPYKRLPQVMQNSNSI LAARIAYHLDLKGPAVTVD TACSSSLTALHLACQSLWLGETDLTVVGGVTLYLTELPHVFM SAAGMLSPNGRCRPFDAADGIVPGE CAVVVLKPLSKALADGDPVHAVIRASGLNQDRGP QRGITAPSARSQTALVRD TLQRFPAVD PAGIDYVECHGTGTPLGDP IEVTALNEAFAGAGLA PASVP IGSVKGNIGHTSAAAGLAGLLKAAGVVRTGLVPPSLHYARANPQIPFDQGPFTVAG ERRELGRPDGGRPRRATVSSFGFSGTNAYVVVEQAPEQAAR PAGDDGAPPVLVPLSGRRAD APAAHAADLARWLRGPGSDASPADVAHTLAVARTHHTYRSALLVSGRDELLES LDLLAAGS ADPRRTDTAPDAAAPDAAVRRQAALLARLVERAGENPGPVELAALAKLYTQGH TVDWAAV SPPARHRRLSLPAYPFAPHRHYVER

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

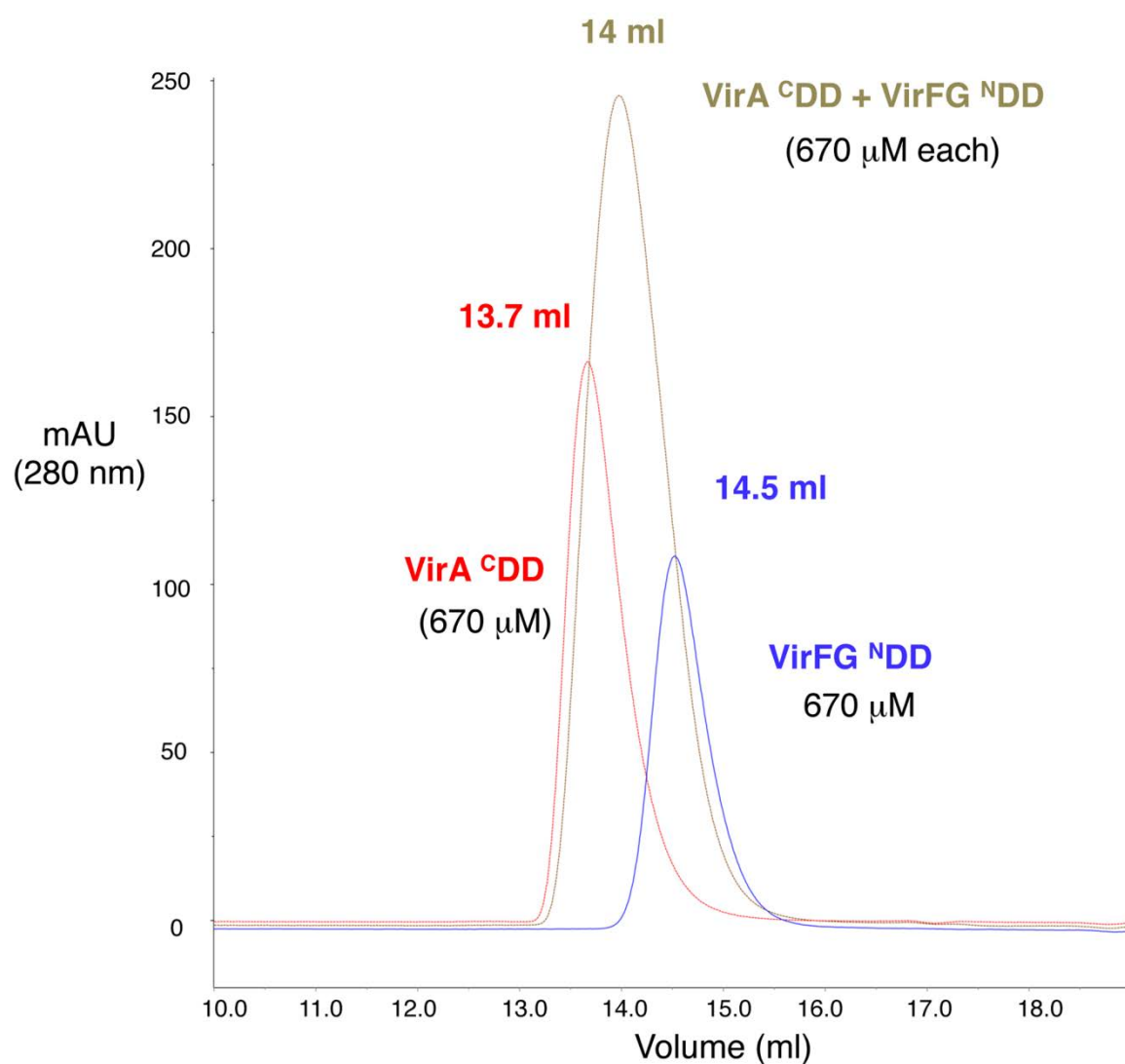
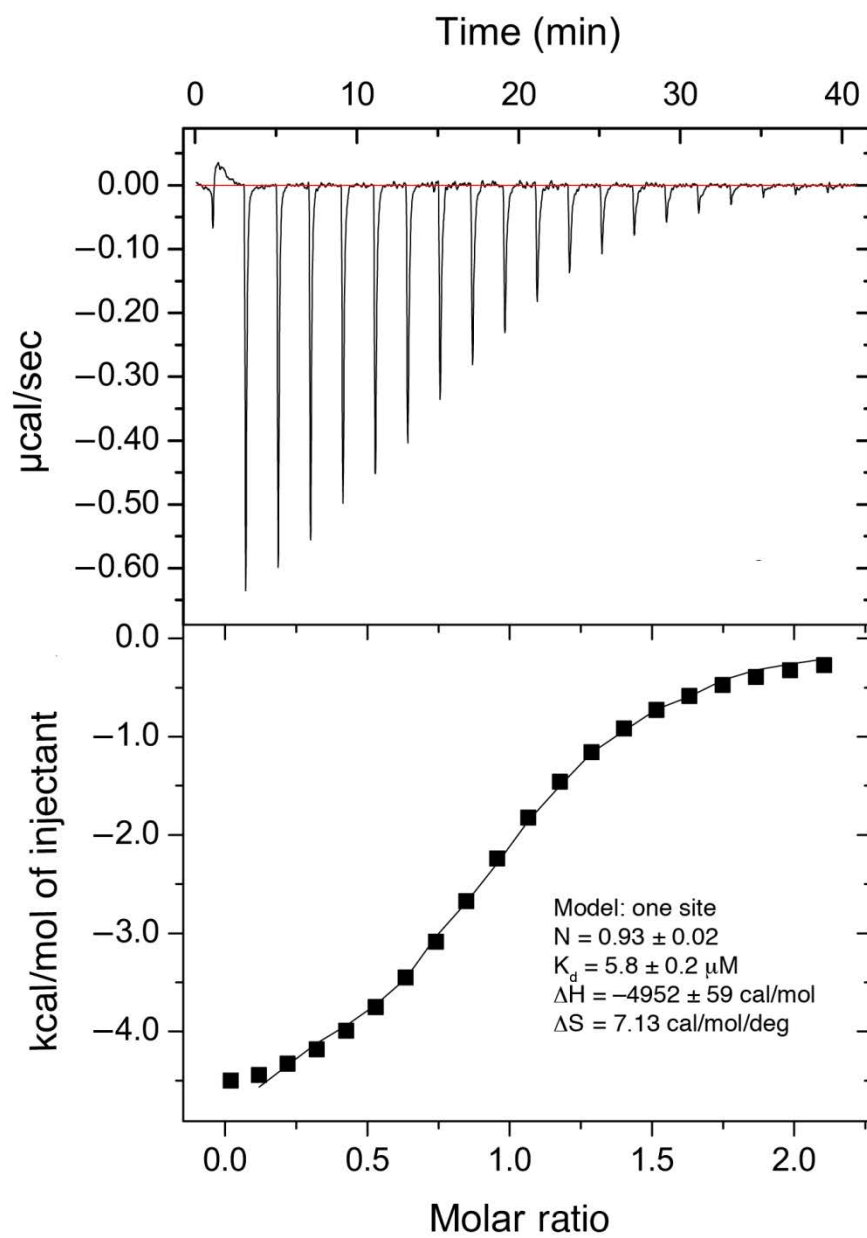
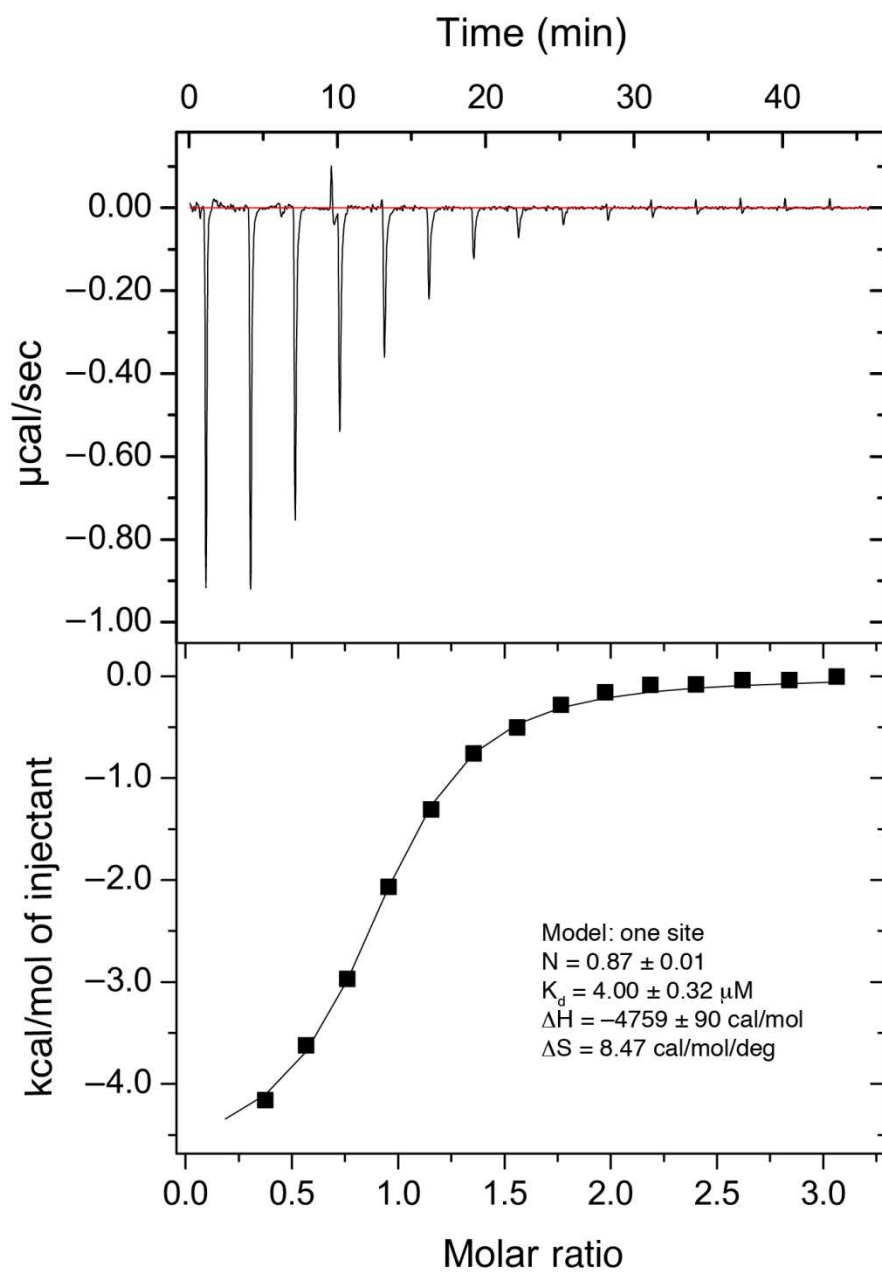


Figure S4. Analysis by gel filtration of the interaction between VirA^{CDD} and VirFG^{NDD}. The individual docking domains (both at an initial concentration of 670 μM) exhibit elution volumes of 13.7 ml and 14.5 ml, respectively. When combined at the same concentrations, a single new peak is observed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

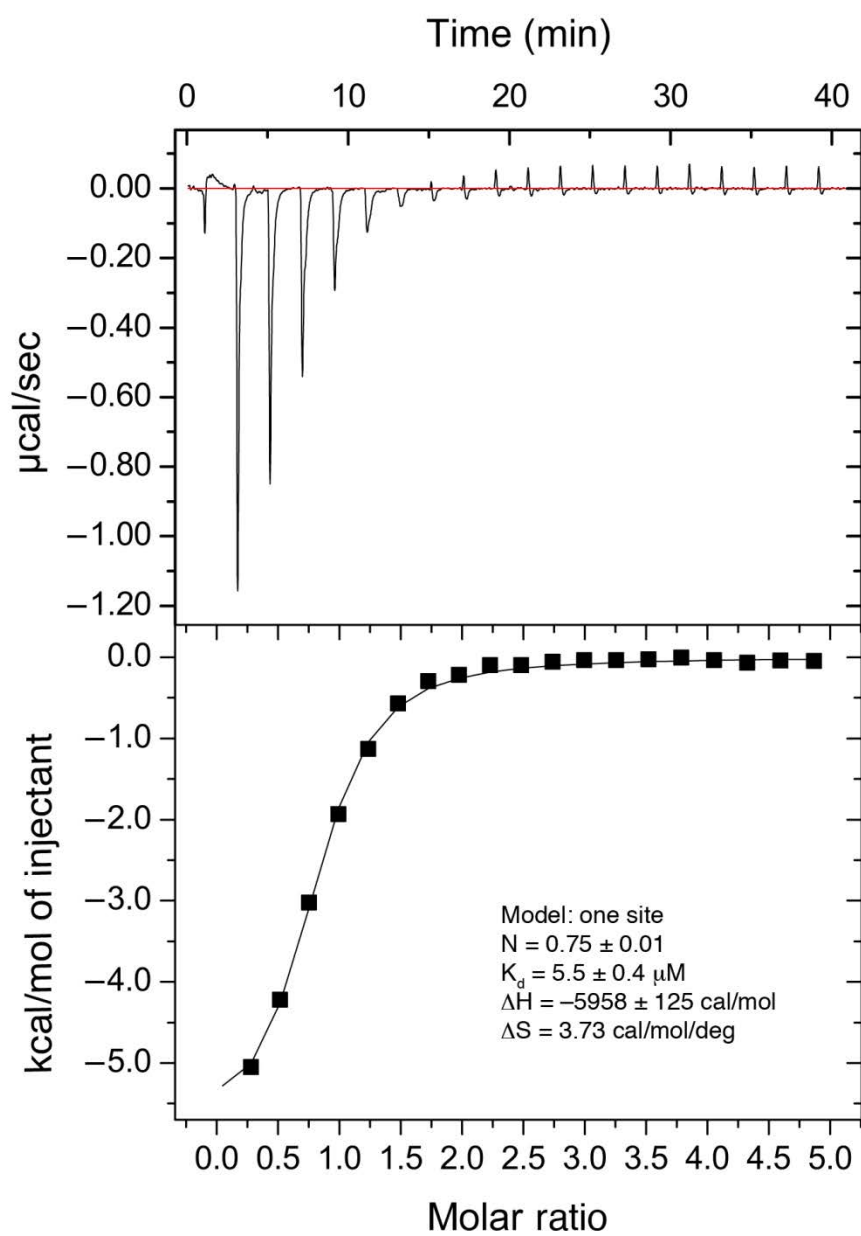
A



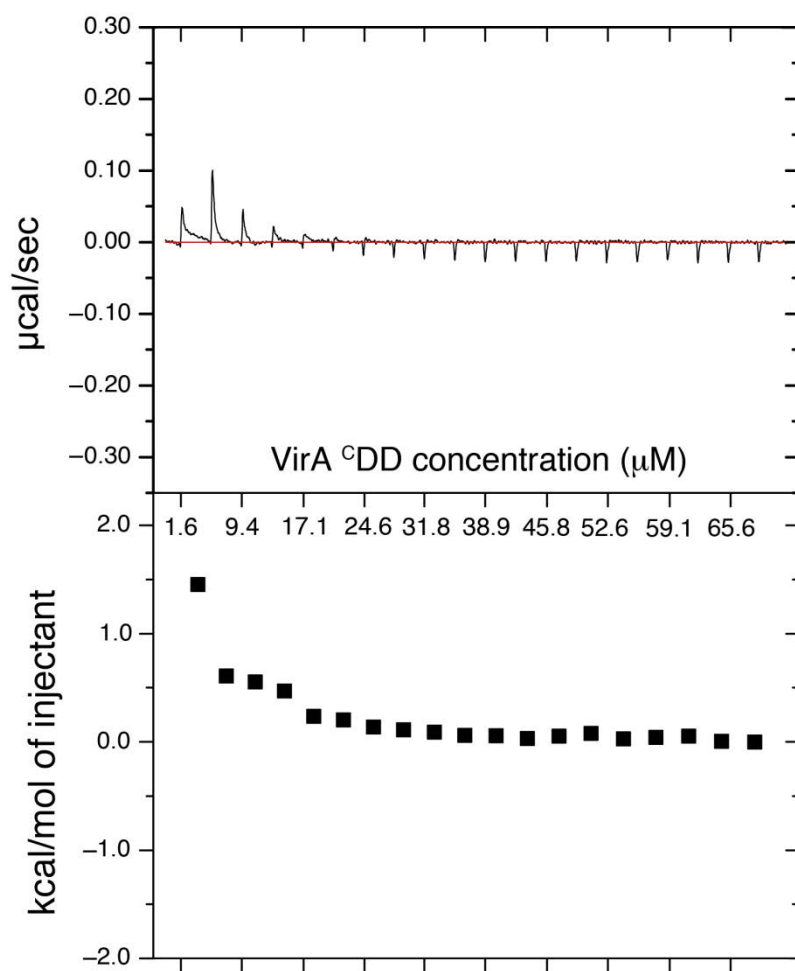
B



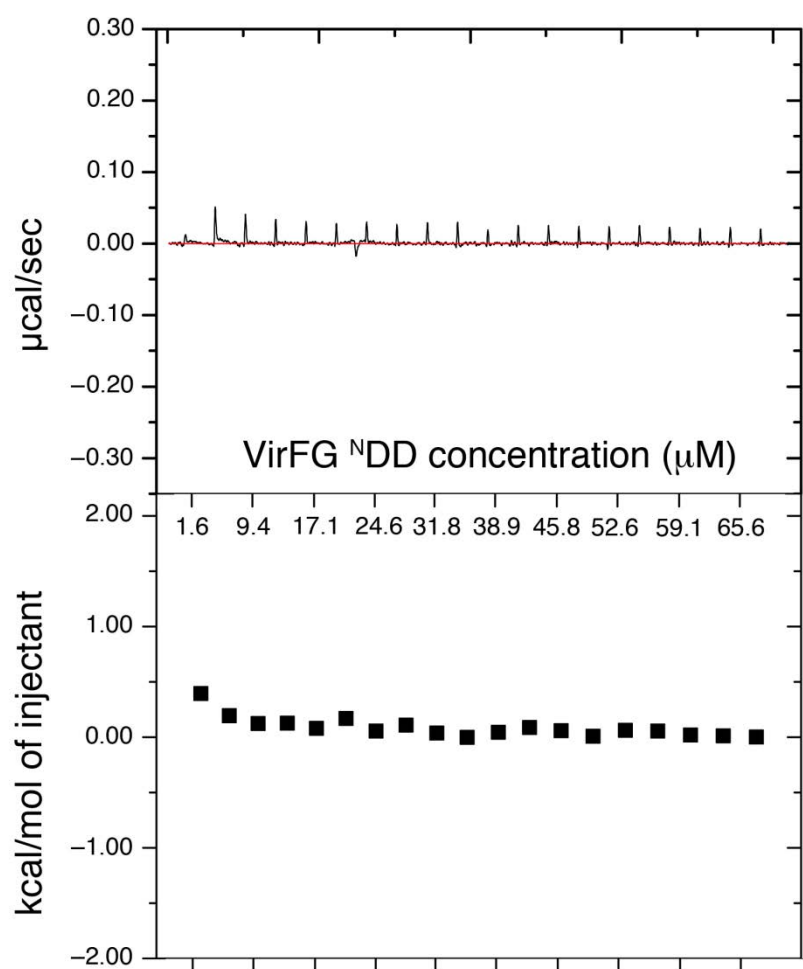
c



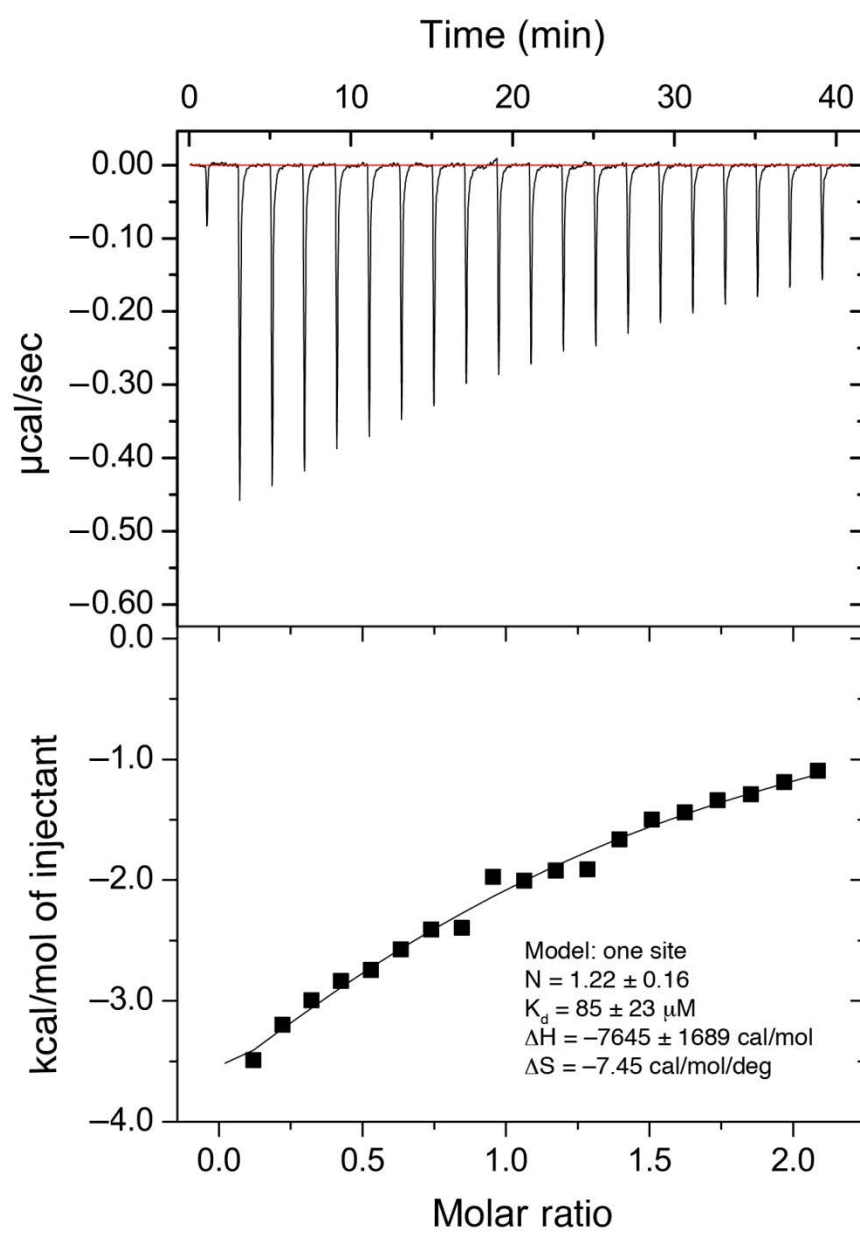
D



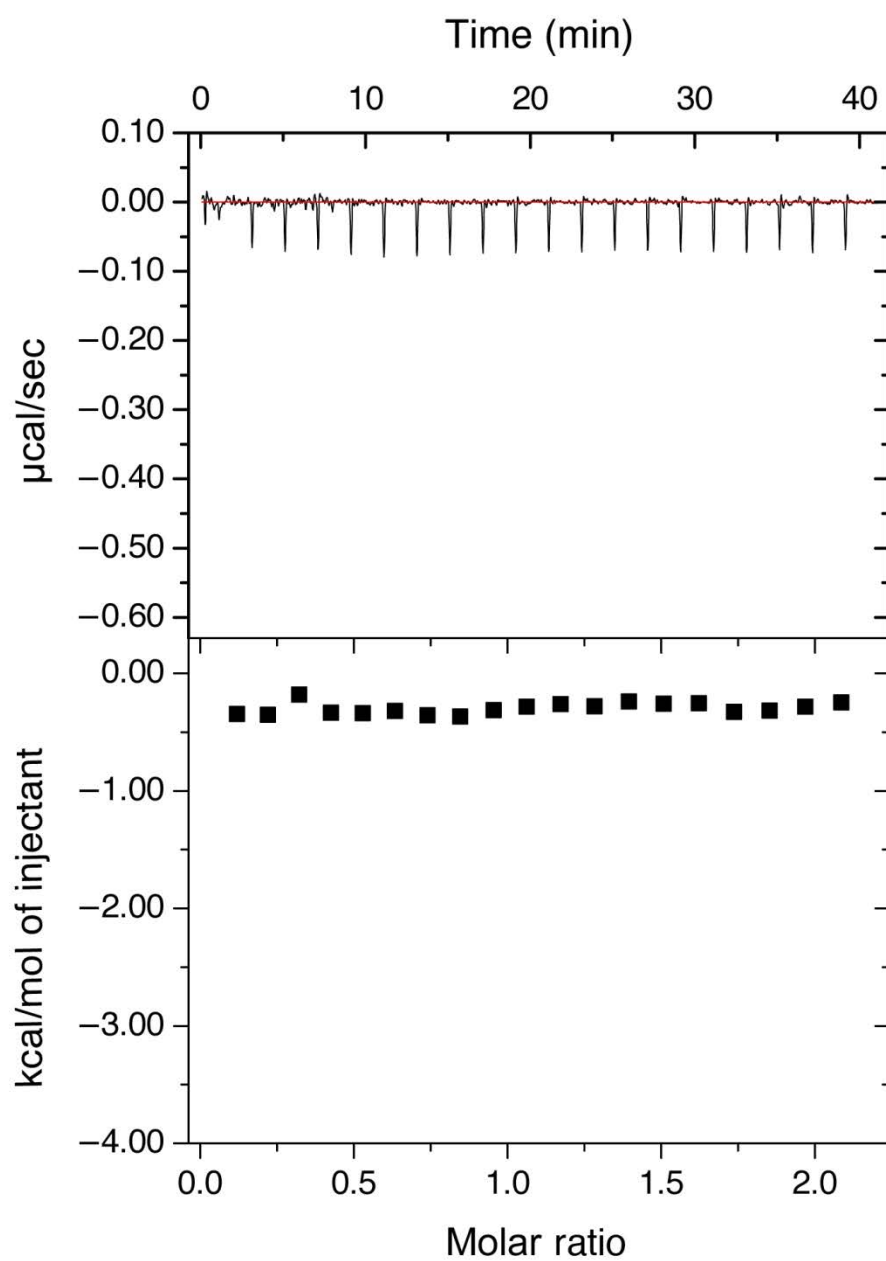
E



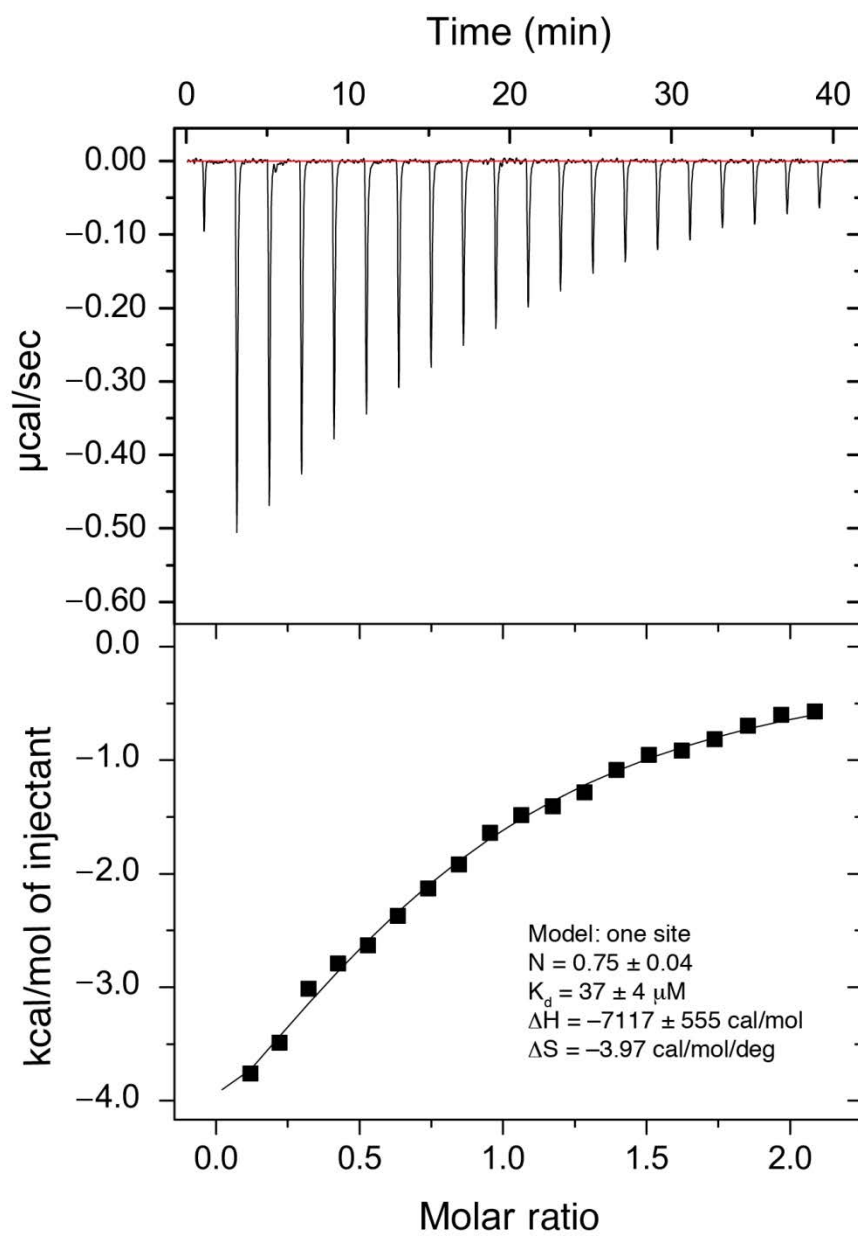
F



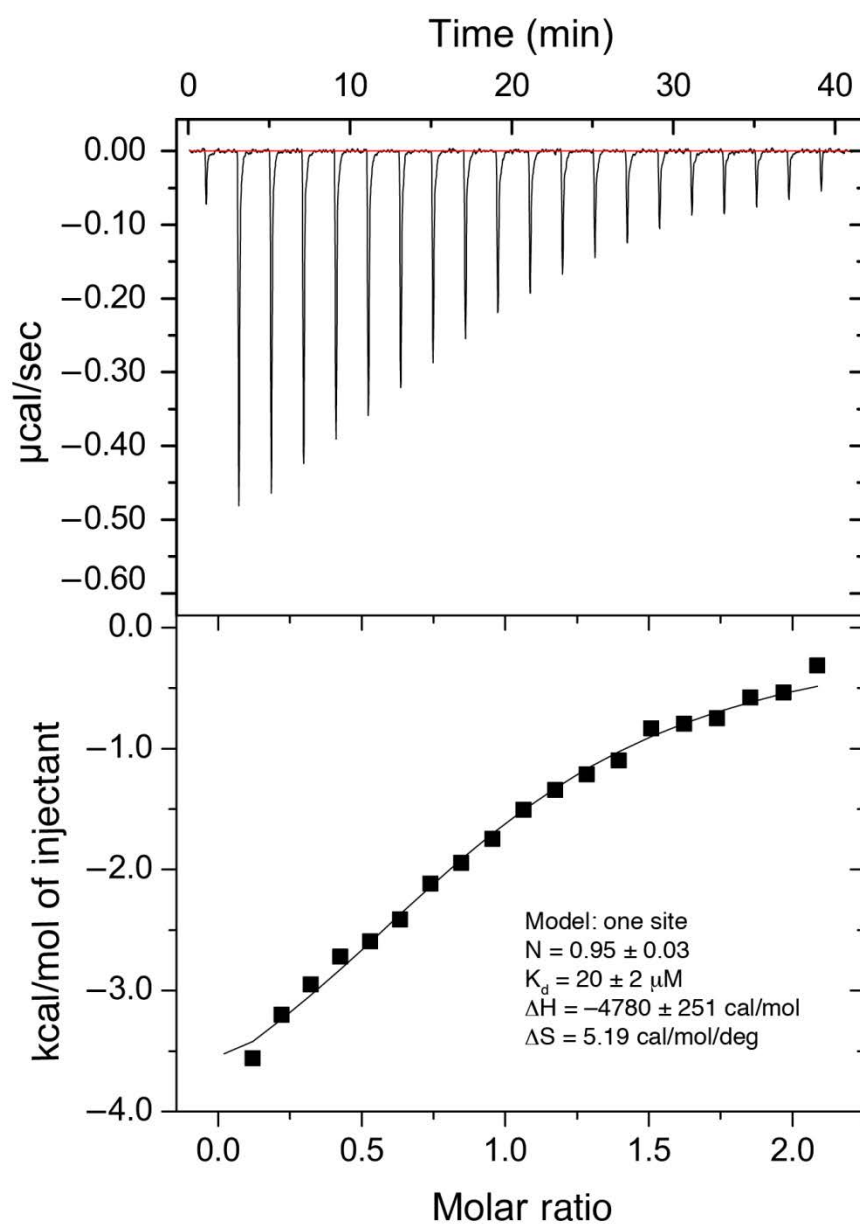
G



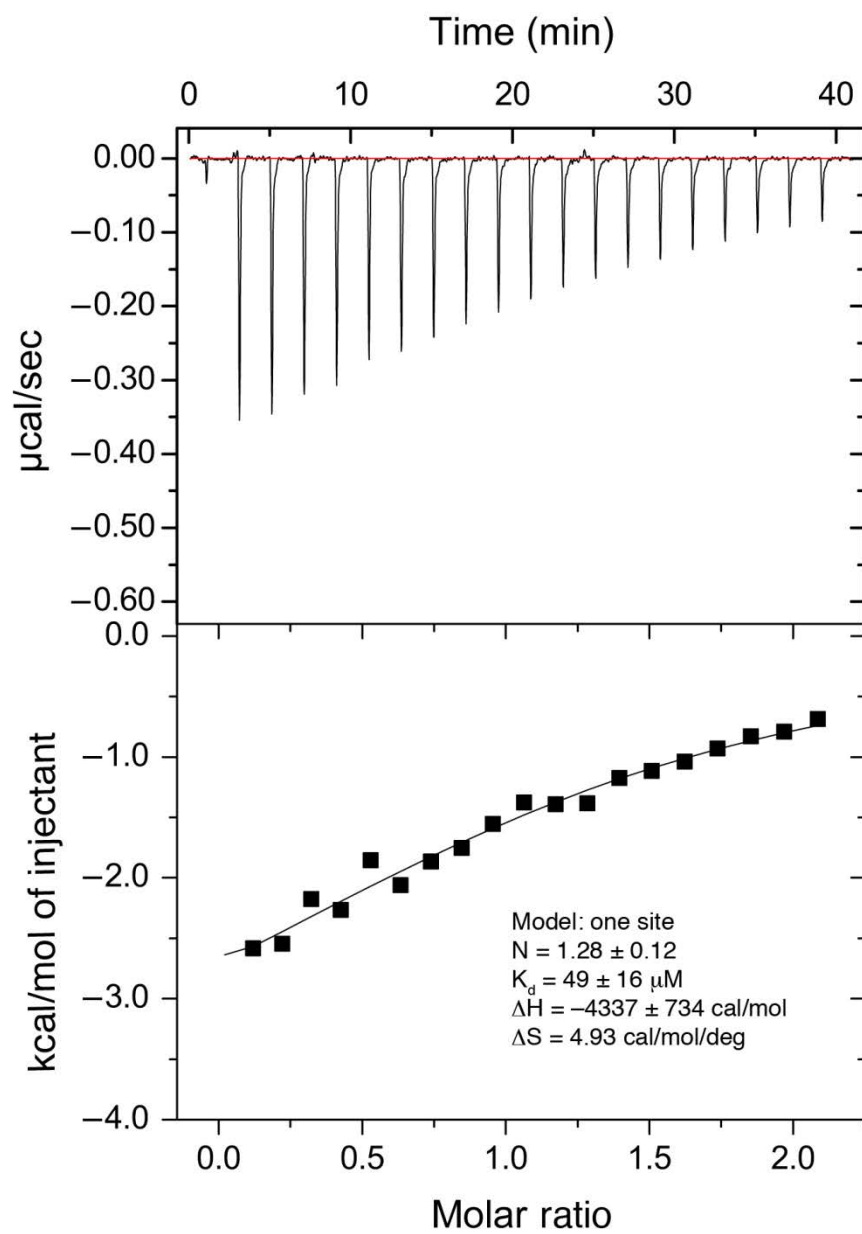
H



I



J



K

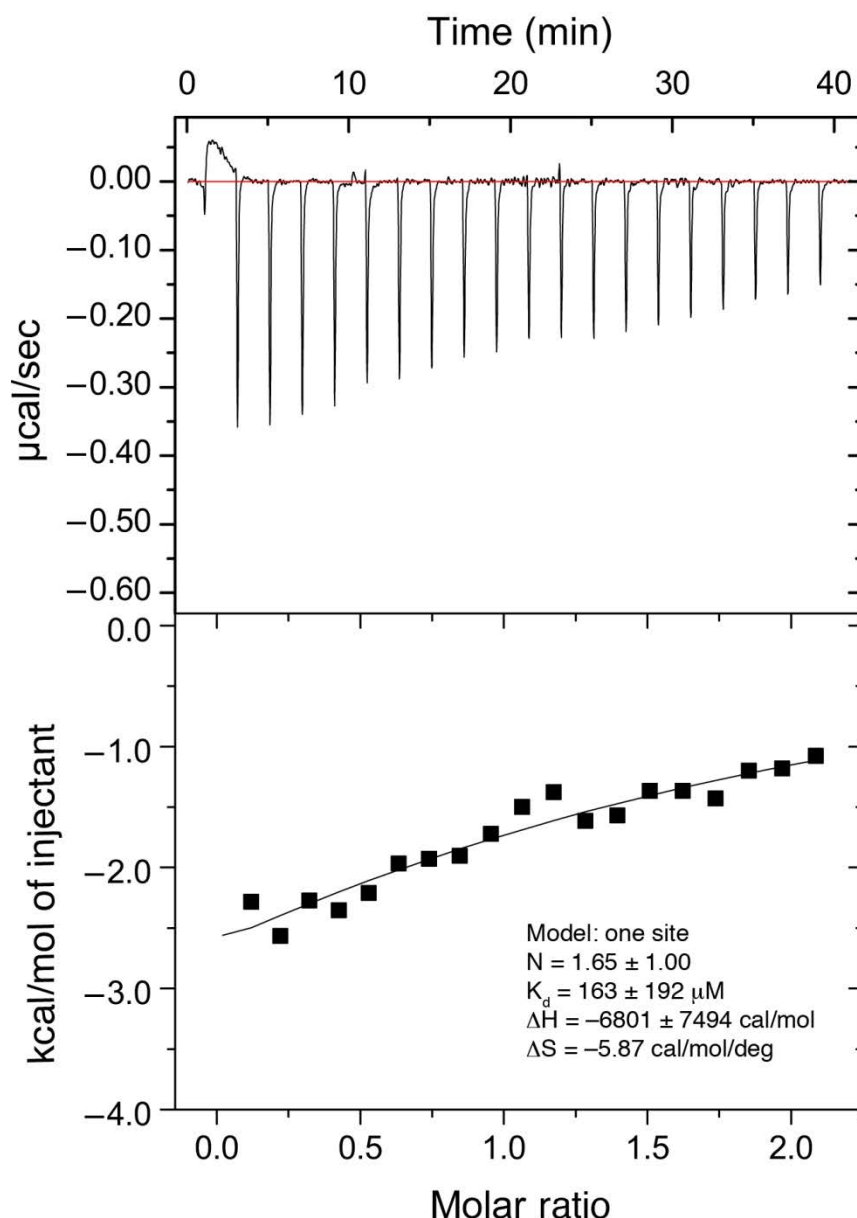
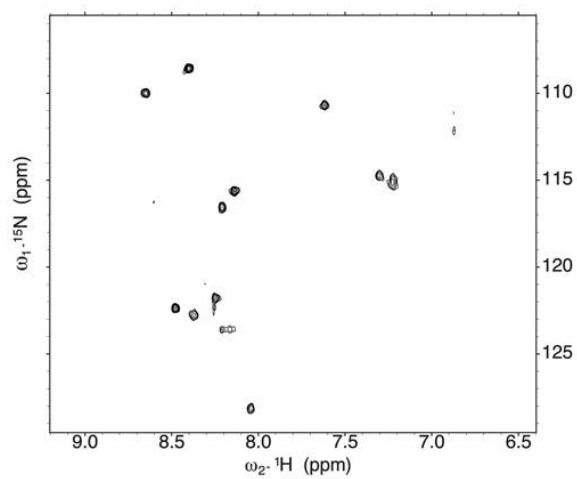
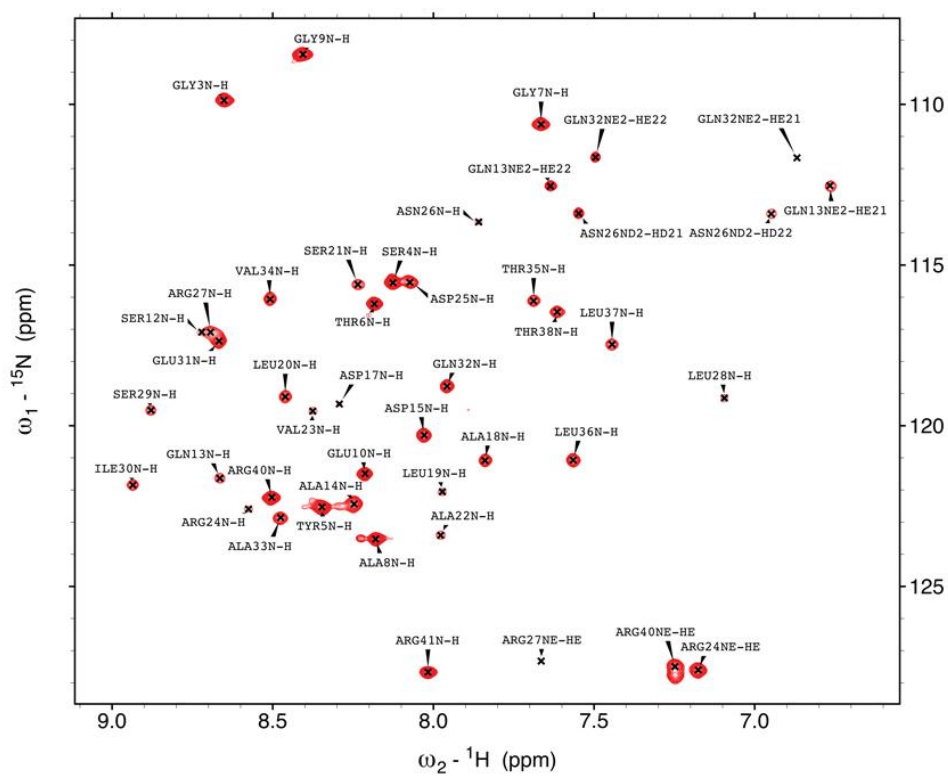


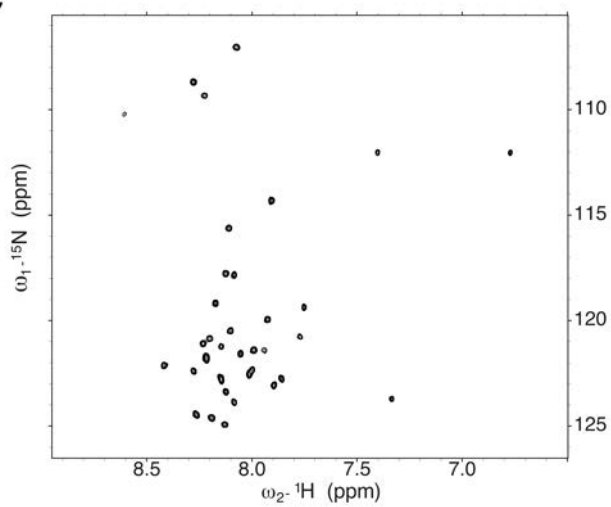
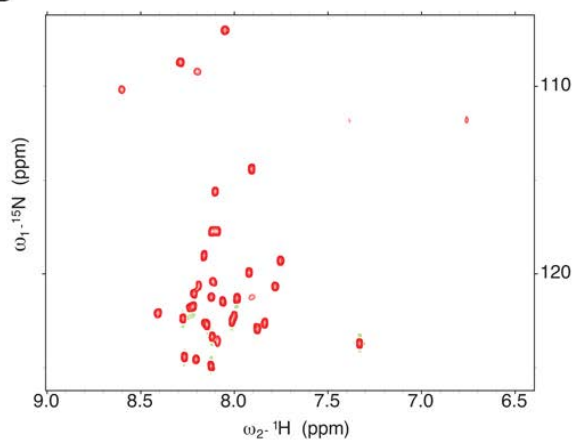
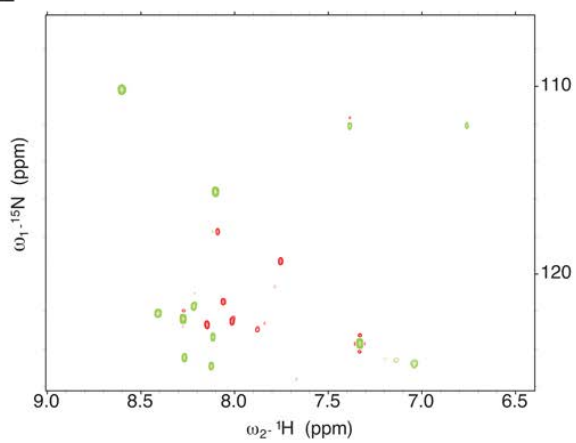
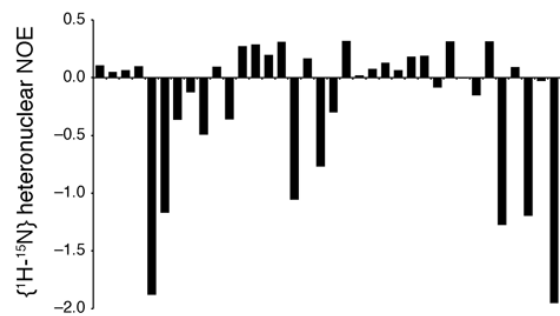
Figure S5. Determination by isothermal titration calorimetry of docking affinity. (A) Analysis of binding between wild type VirA^{CDD} and VirFG^{NDD}. (One of three measurements is shown, but the K_d value represents the average of three independent experiments). (B) Analysis of binding between wild type VirA ACP_{5b}-^{CDD} and VirFG^{NDD}. (C) Analysis of binding between wild type VirA^{CDD} and VirFG^{NDD}-linker. (D) Dilution analysis of VirA^{CDD}. No further change in heat signal is observed after VirA^{CDD} reaches a concentration of 30 μM , indicating that it is completely homodimeric above this concentration. (E) Dilution analysis of VirFG^{NDD}. No evidence for homodimerization is observed across the full range of concentrations (to 78 μM). (F) Analysis of binding between VirA^{CDD} D6940A and wild type VirFG^{NDD}. (G) Analysis of binding between VirA^{CDD} D6940R and wild type VirFG^{NDD}. (H) Analysis of binding between VirA^{CDD} N6949K and wild type VirFG^{NDD}. (I) Analysis of binding between VirA^{CDD} E6954A and wild type VirFG^{NDD}. (J) Analysis of binding between VirA^{CDD} E6954K and wild type VirFG^{NDD}. (K) Analysis of binding between VirA^{CDD} D6940A + E6954A and wild type VirFG^{NDD}. For (B)–(K), the indicated error is that on the individual measurement.

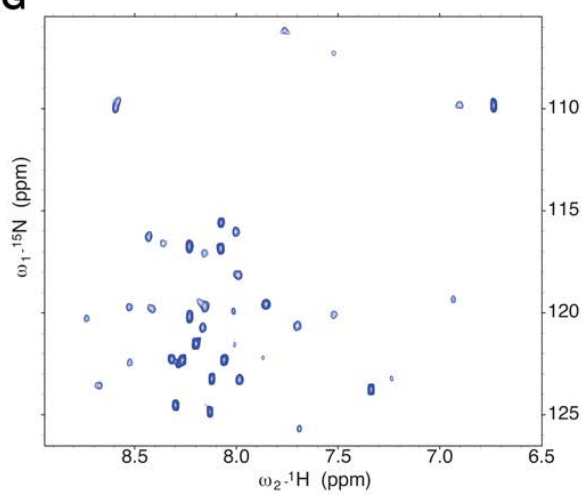
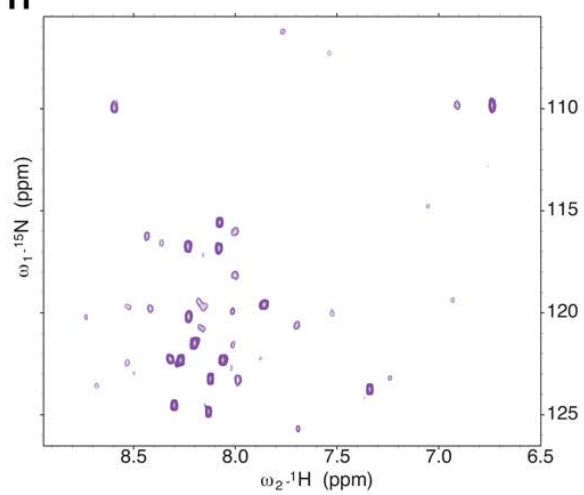
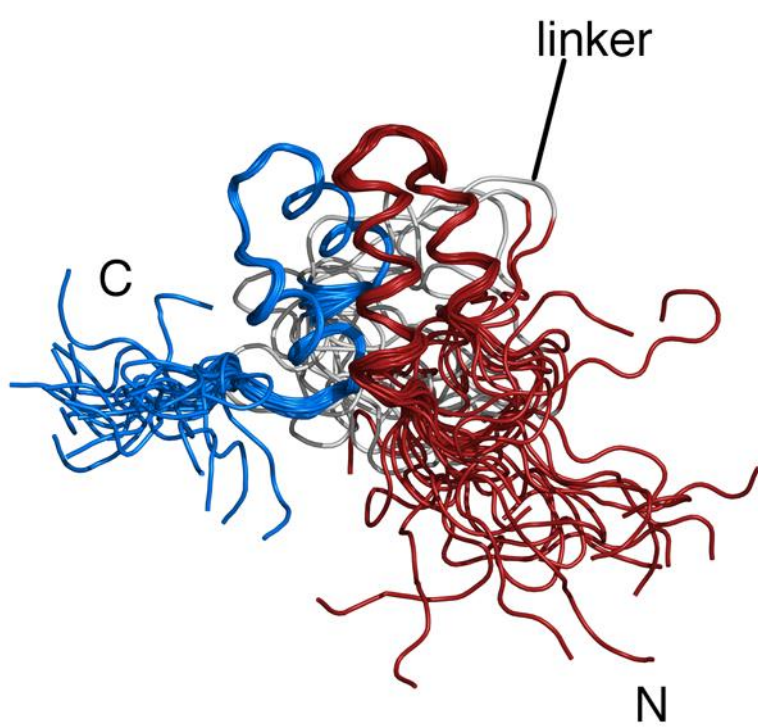
A



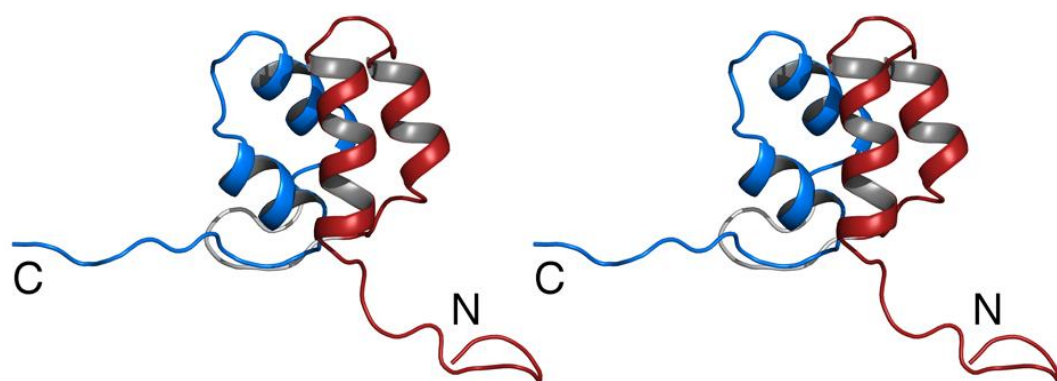
B



C**D****E****F**

G**H****I**

J



K

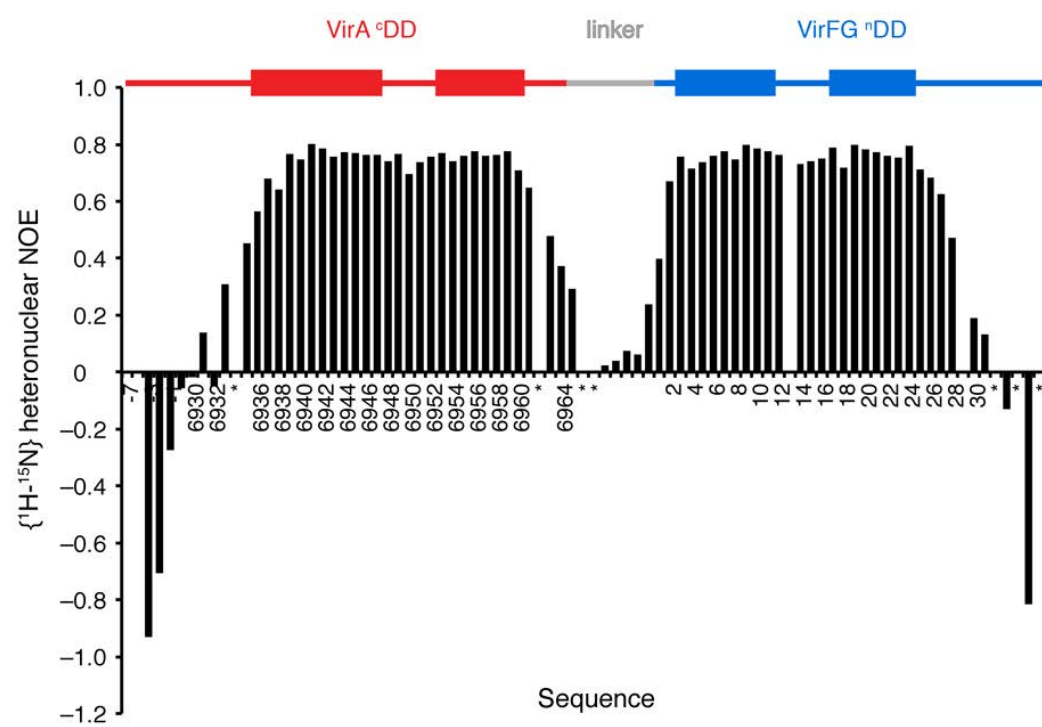


Table S5. Summary of SAXS and SEC-MALLS data.

Construct	R_g (Guinier) (Å) ^a	R_g (GNOM) (Å) ^a	D_{max} (GNOM) (Å) ^b	χ^2 (DAMMIN) ^c	χ^2 (CRY SOL) ^c	Theoretical MW (kDa)	SAXS MoW (kDa) ^d (q range used)	MW SEC-MALLS (kDa)	by
VirA ^C 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 ^N 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 ^N DD-linker	n.d.	n.d.	n.d.	n.d.	n.d.	6.87	n.d.	6.7 ± 0.6	
VirA ^C DD/VirFG ^N 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 ^C DD–VirFG ^N 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 _{5b} – ^C DD/VirFG ^N 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 _{5b} – ^C DD ³	37.0 ± 0.3	37.7	140	1.231	n.d.	14.3	31.4 ³		Rapid monomer/dimer exchange
VirA ACP _{5b} – ^C DD– ^N 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 _{5b} – ^C DD– ^N DD–KS ₆ VirFG fusion	48 ± 2	50.5 ± 0.1	200	1.76	n.d.	171.6 (dimer)	205 (0.2)	n.d.	

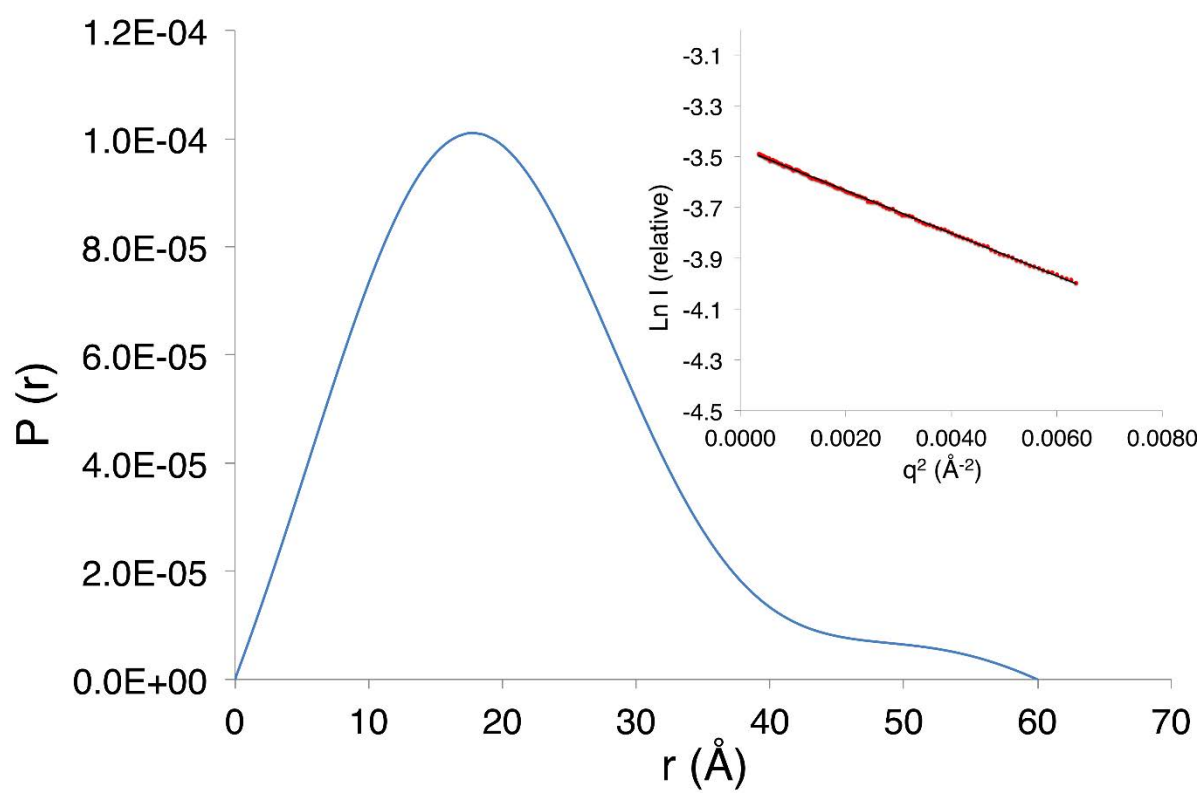
^a R_g is the radius of gyration derived from the atomic models using the program PRIMUS.⁵¹ For experimental data, R_g is given by the Guinier approximation.

^b D_{max} is the maximum particle diameter derived from the distance distribution function ($P(r)$) using the program GNOM.⁵²

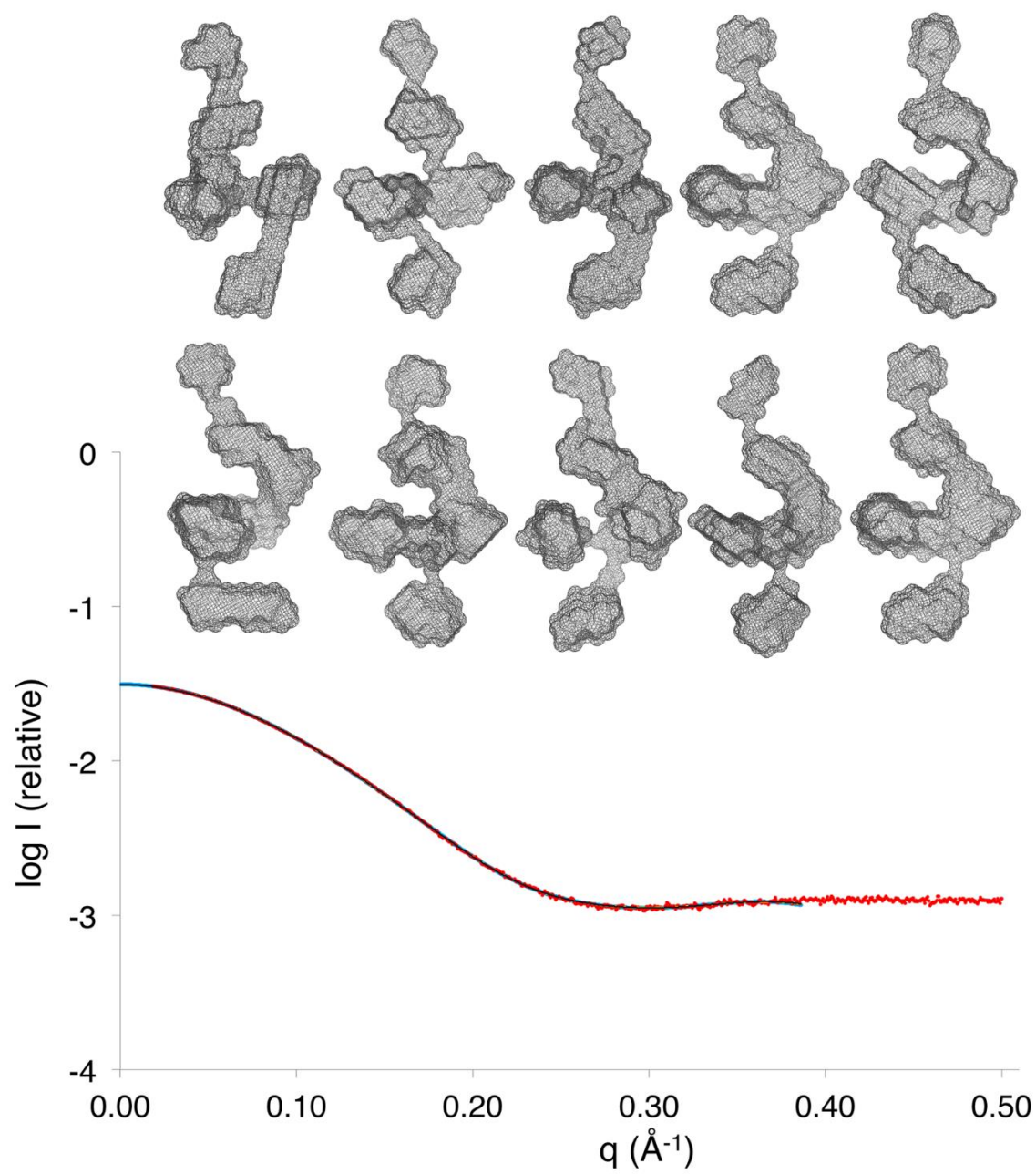
^c χ^2 is the discrepancy between the experimental SAXS profile and the theoretical curve calculated using the programs DAMMIN⁵³ or CRY SOL.⁵⁴ n.d.: not determined.

^dThe theoretical error on a typical molecular weight measurement by SAXS MoW is 10%.⁵⁵

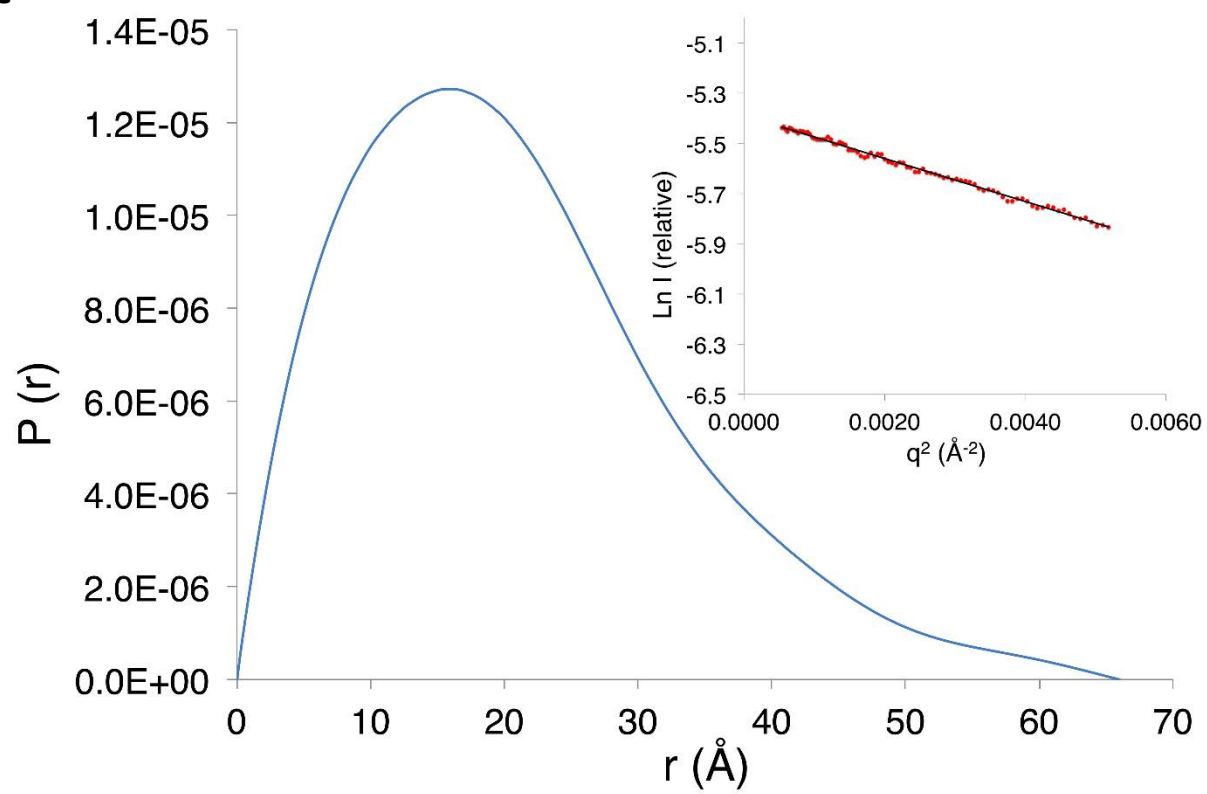
A



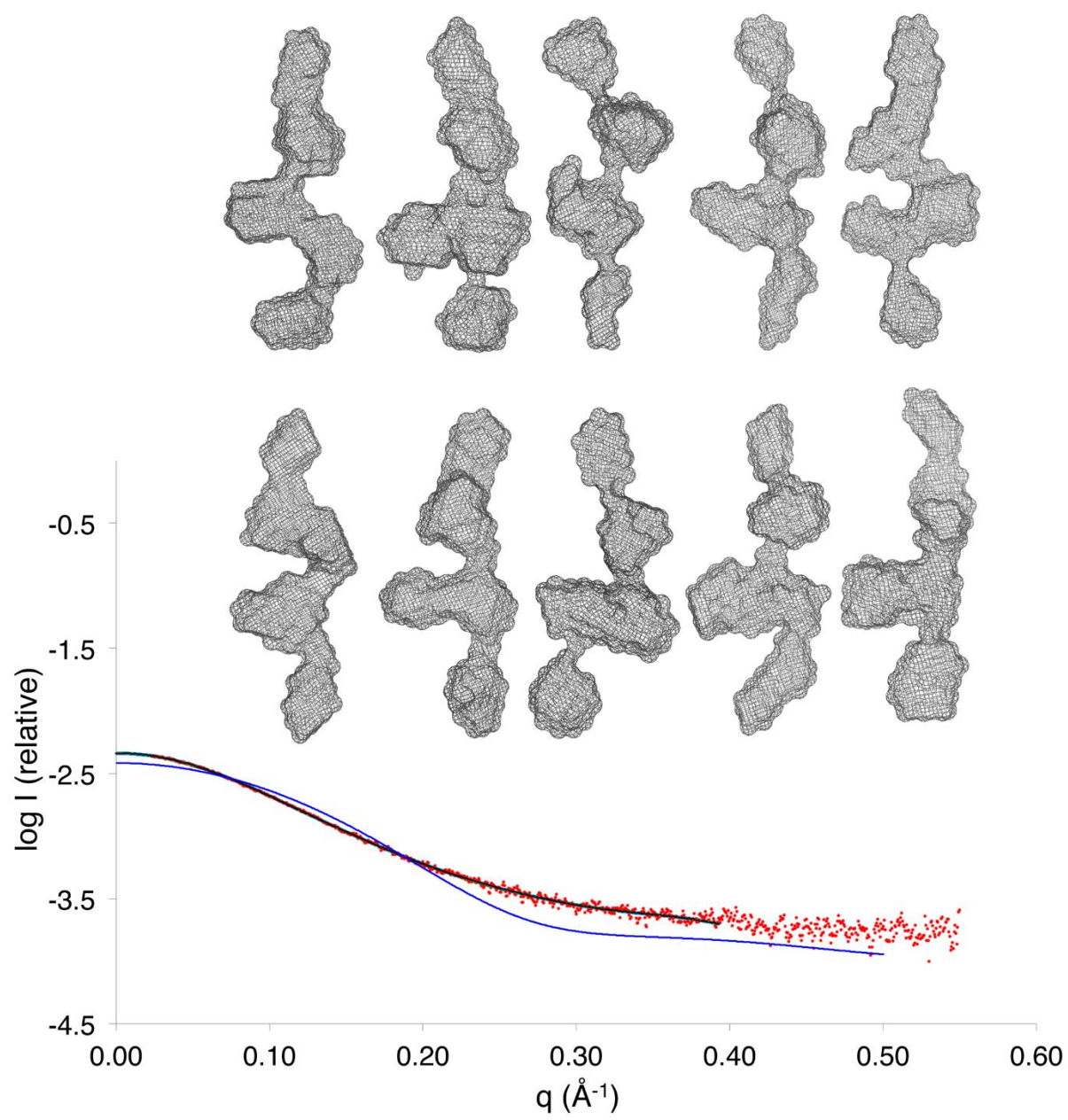
B

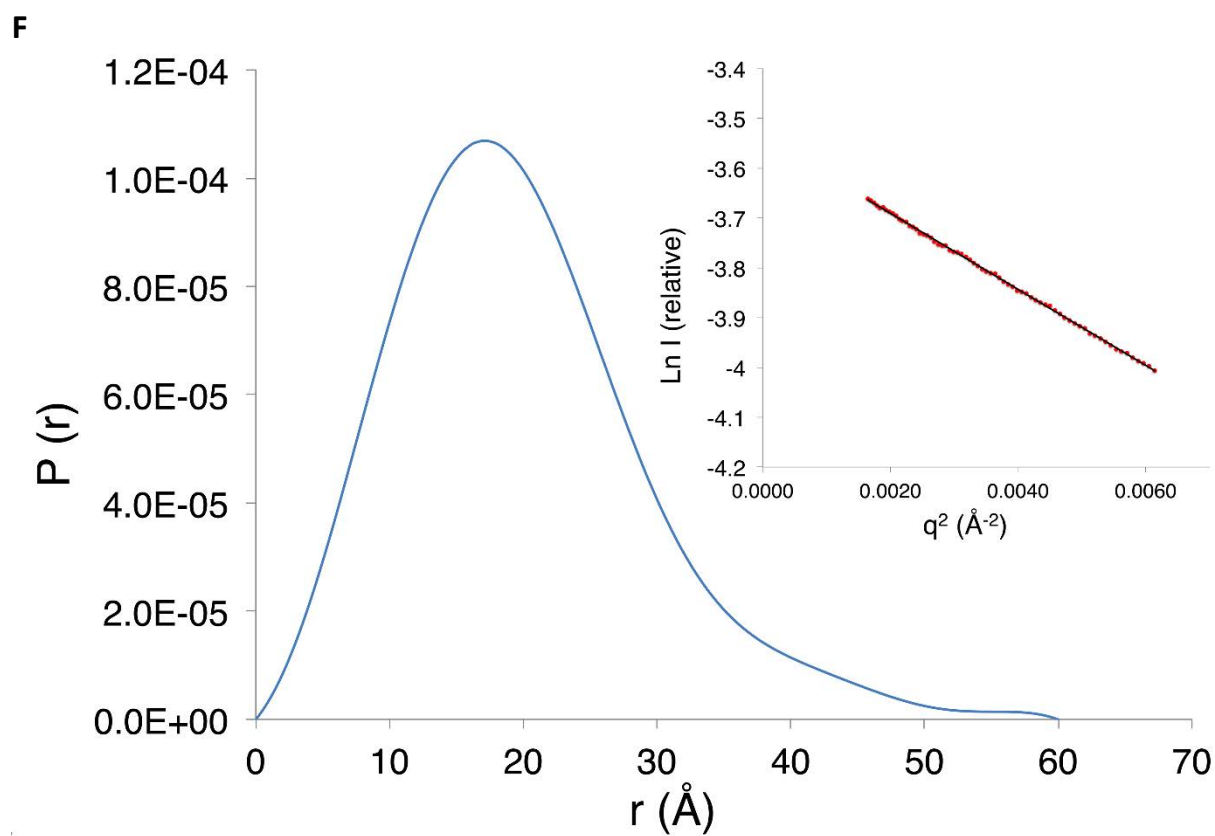
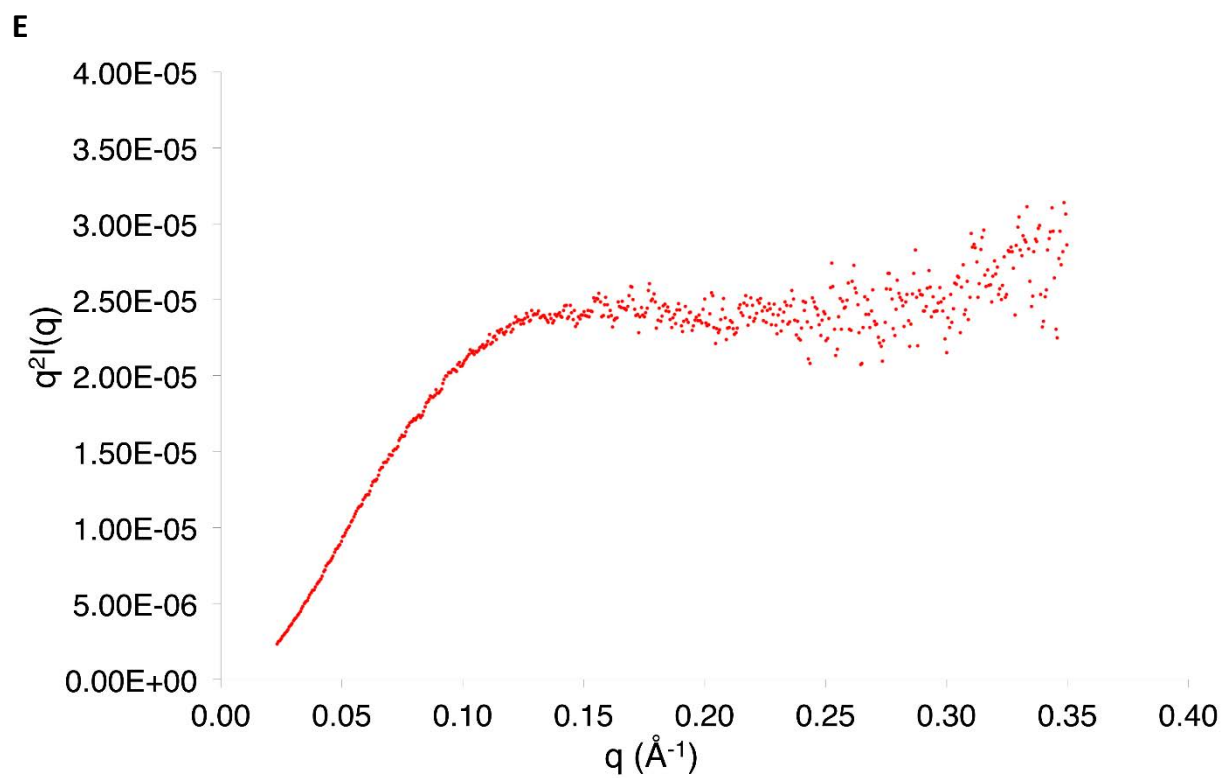


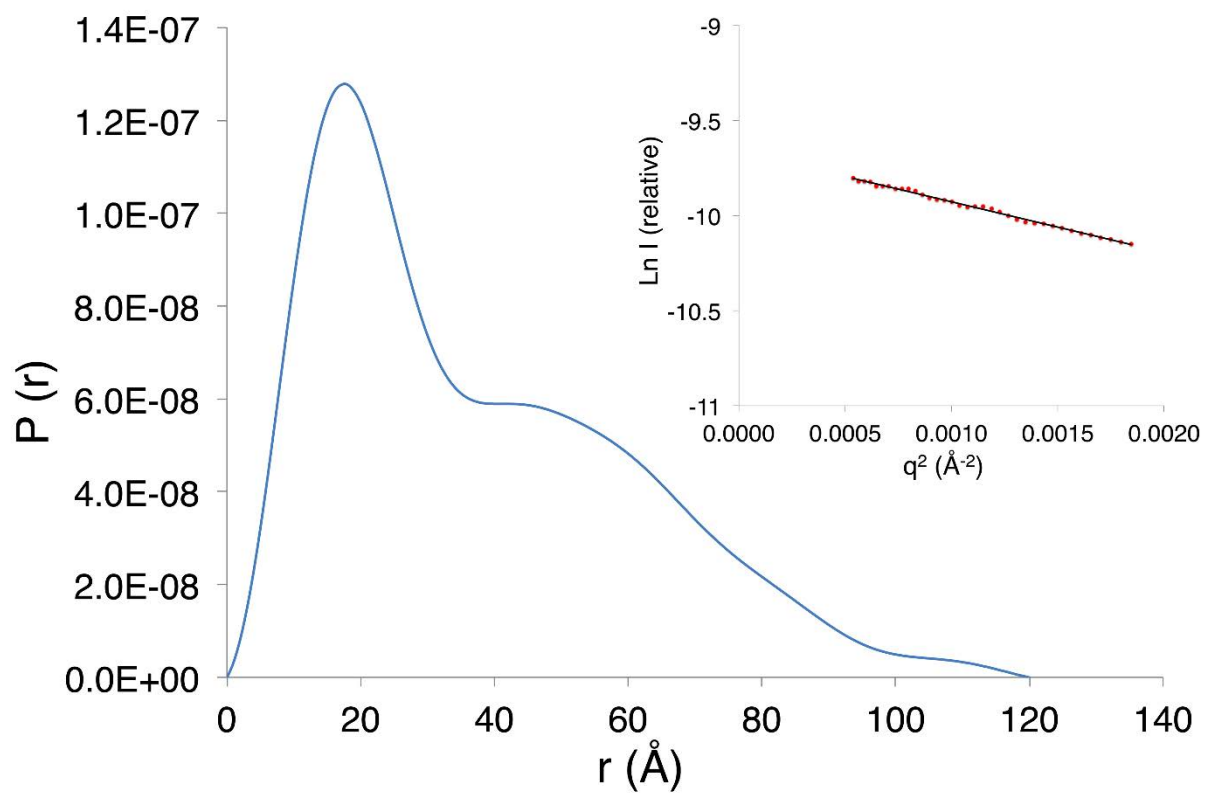
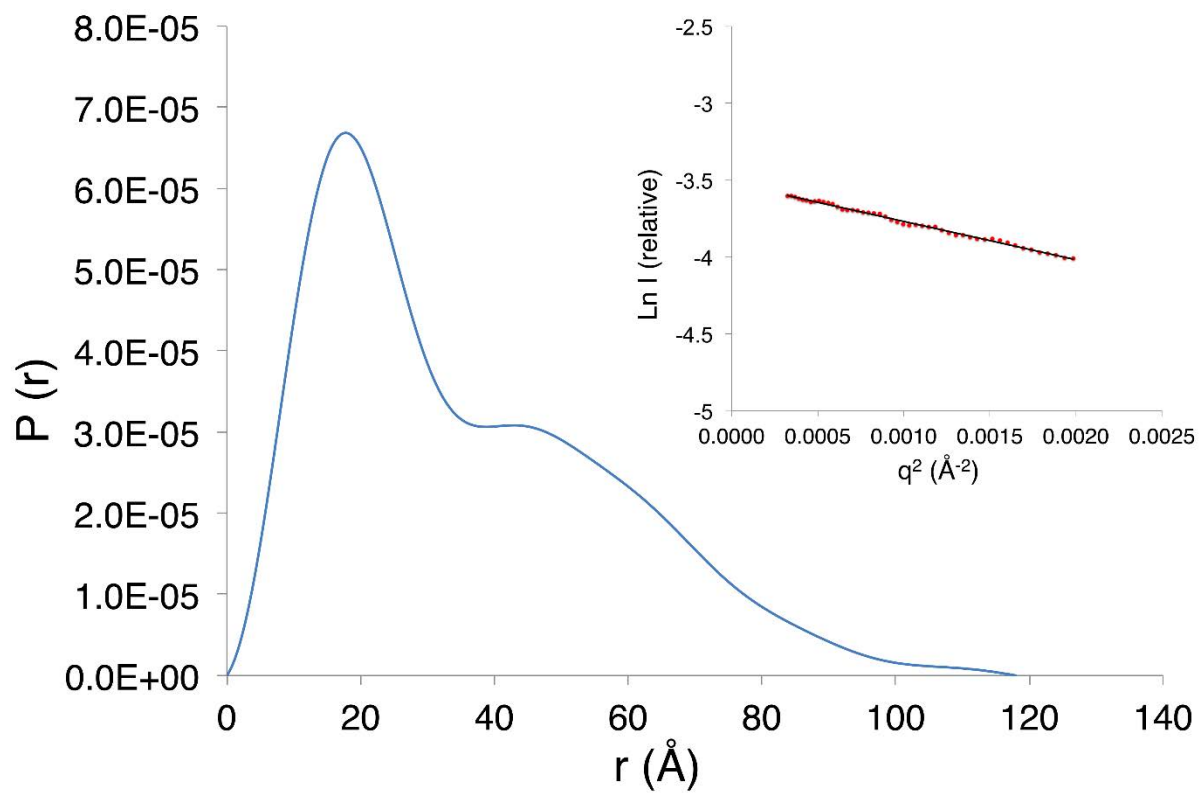
c



D





G**H**

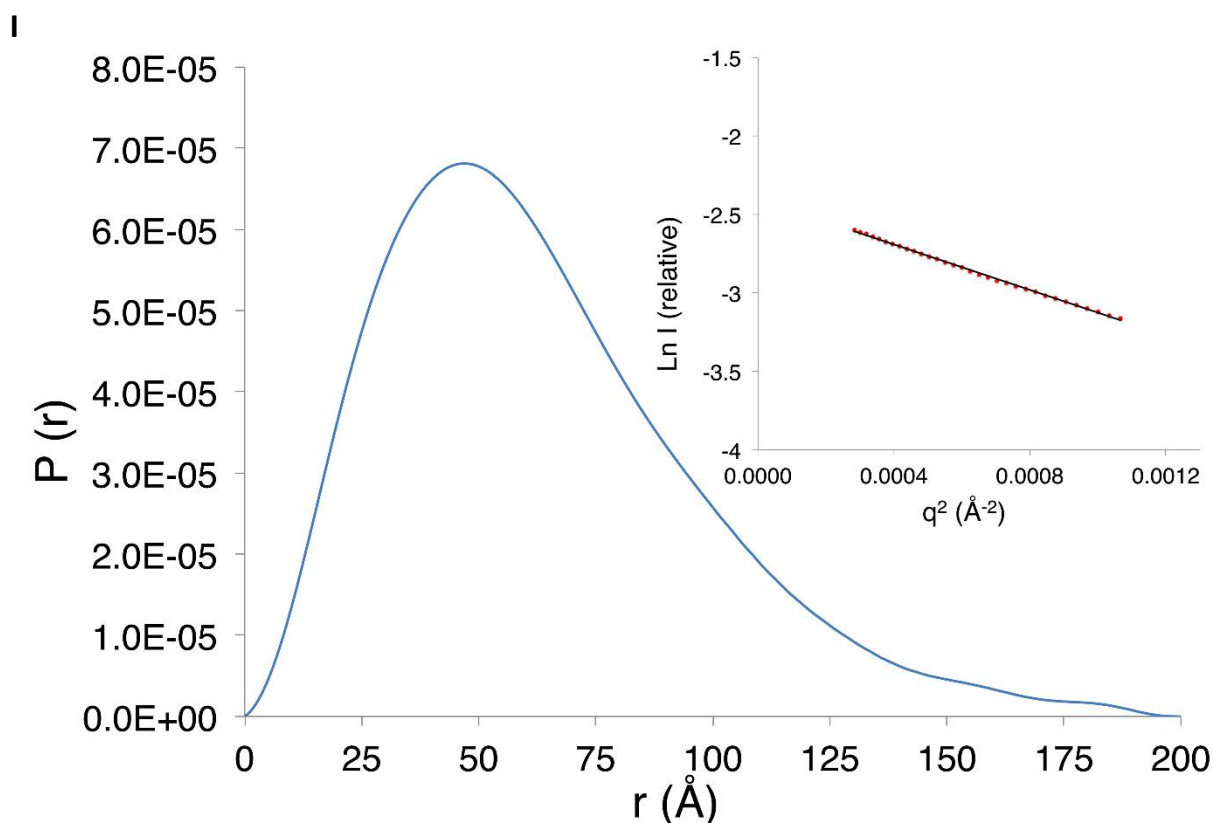
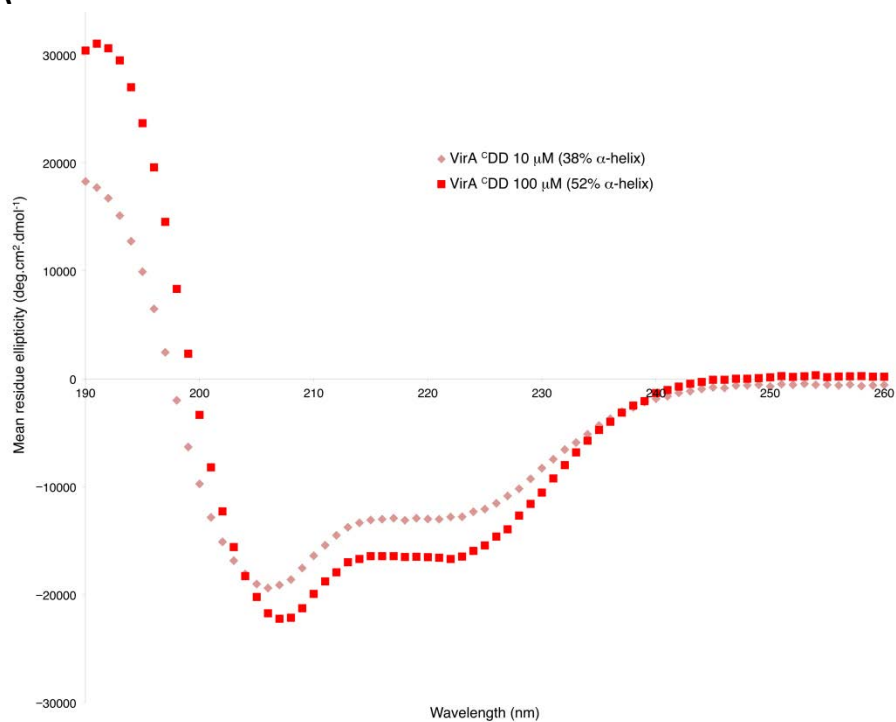
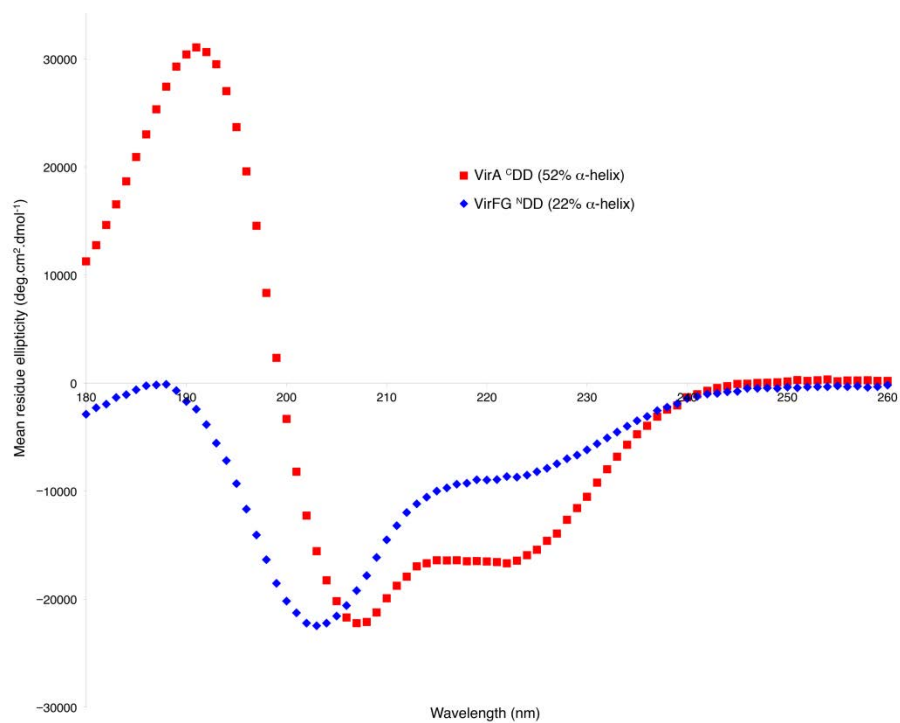
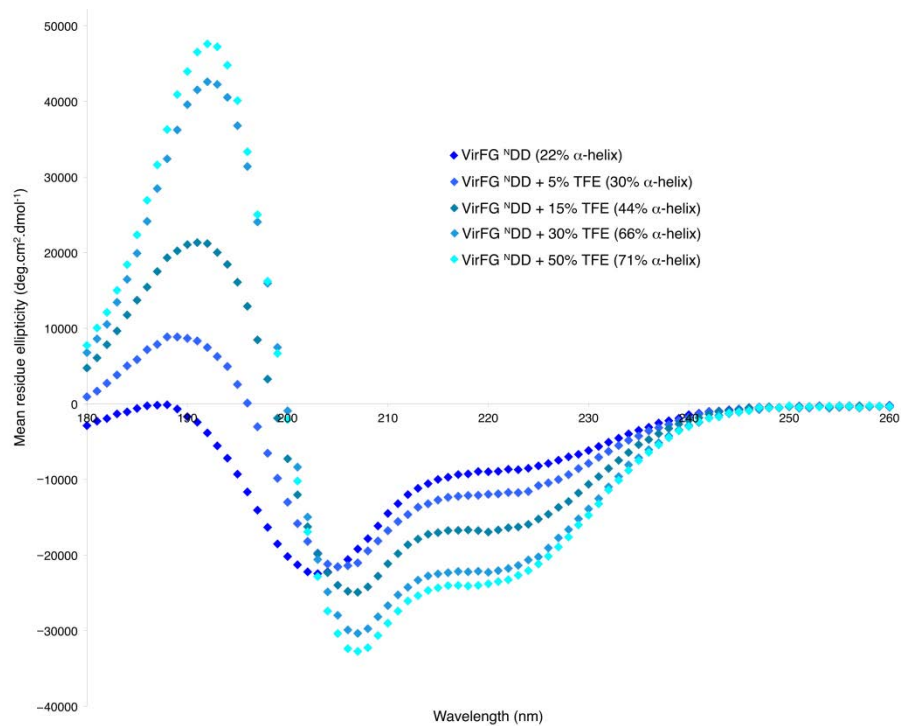


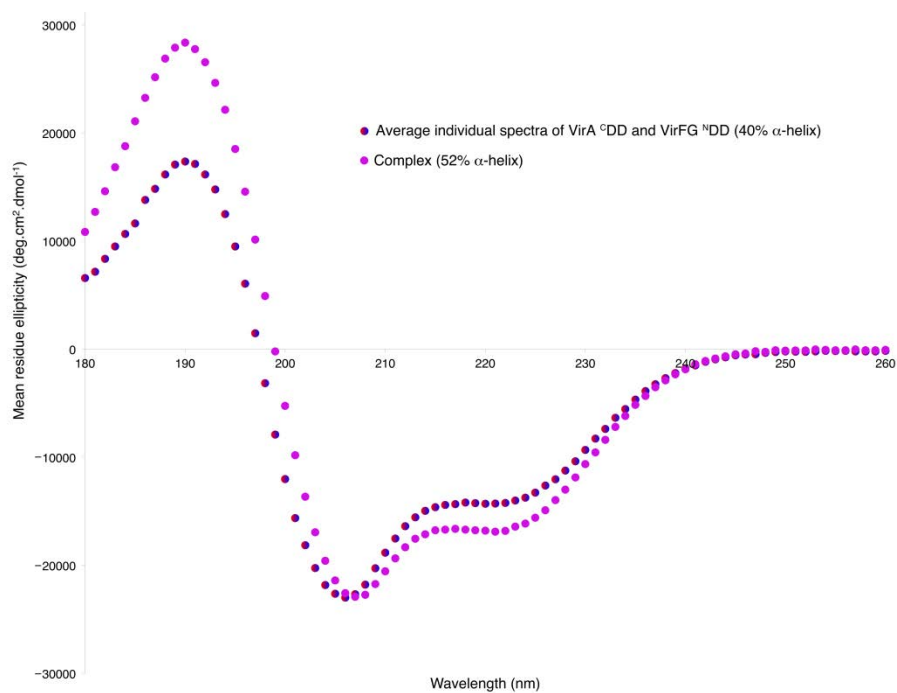
Figure S7. Analysis by SAXS of the various constructs. (A) The distance distribution function for VirA^{CDD} calculated with GNOM. The Guinier plot is inset. (B) Panel of 10 representative *ab initio* form calculations for VirA^{CDD} calculated with DAMMIN. The fit between the SAXS data and the 10 form calculations is shown ($\chi^2 = 1.732\text{--}2.543$). (C) The distance distribution function for VirFG^{NDD} calculated with GNOM. The Guinier plot is inset. (D) Panel of 10 representative *ab initio* envelopes of VirFG^{NDD} computed with DAMMIN. Superimposition of the experimental SAXS data acquired on VirFG^{NDD} (red dots) and the theoretical curve for VirFG^{NDD} calculated using CRY SOL from one of the VirA^{CDD}-VirFG^{NDD} 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 VirFG^{NDD} is excellent ($\chi^2 = 2.321\text{--}2.489$). (E) The Kratky plot (red dots) for VirFG^{NDD} calculated in the q range of $0.0005\text{--}0.35\text{ \AA}^{-1}$. (F) The distance distribution function for VirA^{CDD}-VirFG^{NDD} (covalent fusion) calculated with GNOM. The Guinier plot is inset. (G) The distance distribution function for VirA ACP_{5b}-^{CDD}/VirFG^{NDD} (non-covalent complex) calculated with GNOM. The Guinier plot is inset. (H) The distance distribution function for VirA ACP_{5b}-^{CDD}-VirFG^{NDD} (covalent complex) calculated with GNOM. The Guinier plot is inset. (I) The distance distribution function for VirA ACP_{5b}-^{CDD}-VirFG^{NDD}-KS₆ (covalent complex) calculated with GNOM. The Guinier plot is inset.

A**B**

C



D



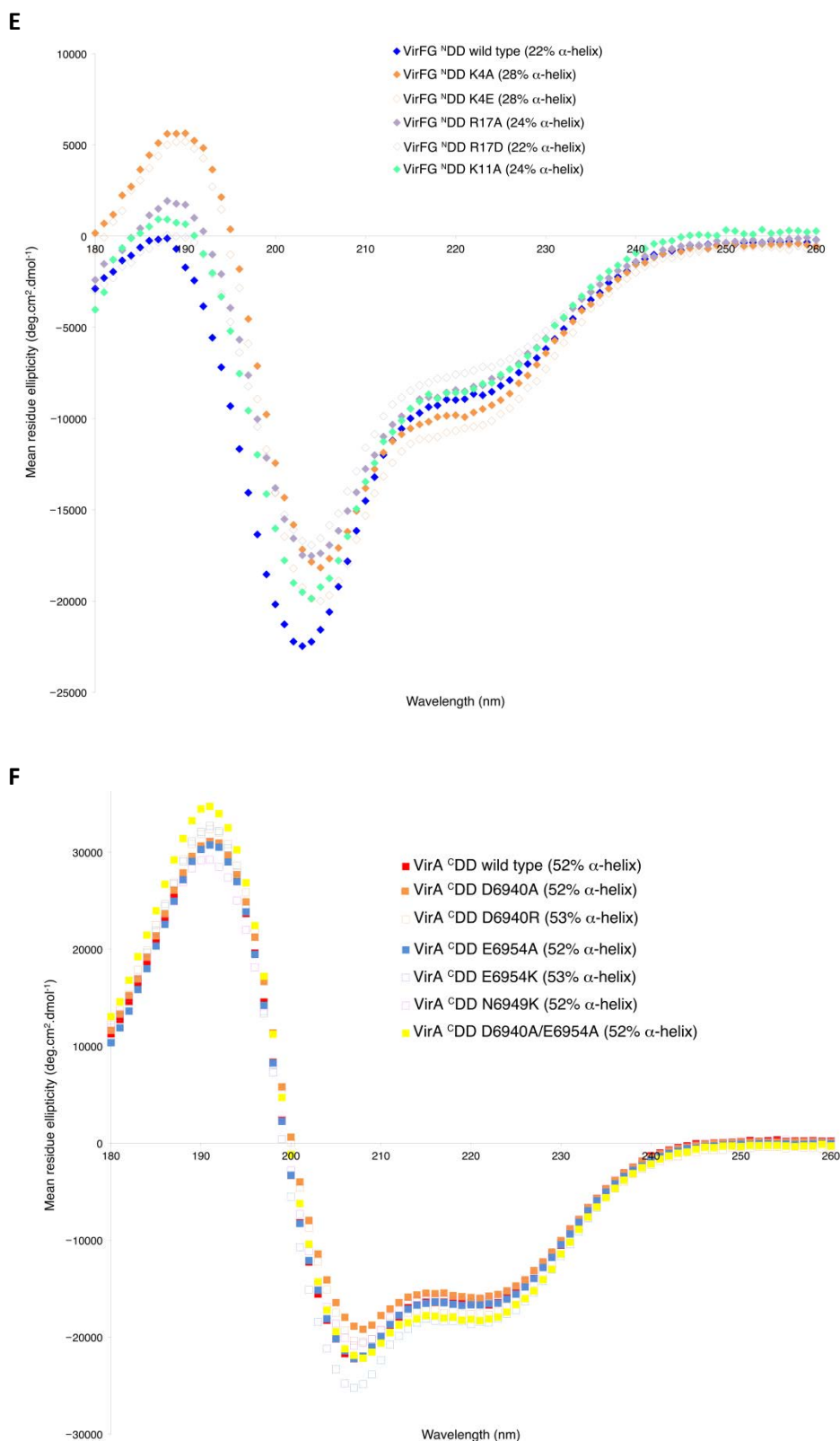


Figure S8. Analysis by circular dichroism of wild type and mutant docking domains. (A) Analysis of VirA ^CDD at concentrations above (100 μM) and below (10 μM) the point at which it homodimerizes completely. The percentage α-helix is indicated in each case. (B) Analysis of wild type VirFG ^NDD and for comparison, VirA ^CDD (both at 100 μM). 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.

Table S6. NMR and refinement statistics for VirA ^CDD–VirFG ^NDD.

	Protein
NMR distance and dihedral constraints	VirA ^C DD–VirFG ^N DD
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 ± 0.008
Max. distance constraint violation (Å)	0.12
Dihedral angle constraints (°)	0 ± 0
Max. dihedral angle violation (°)	0
Deviations from idealized geometry	
Bond lengths (Å)	0.011 ± 0.001
Bond angles (°)	1.947 ± 0.024
Impropers (°)	n.d.
Average pairwise r.m.s. deviation* (Å)	VirA ^C DD 6936–6960, VirFG ^N DD 3–25
Heavy	0.99 ± 0.13
Backbone	0.38 ± 0.09

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

Table S7. Putative interface residues in the VirA^CDD/VirFG^NDD docking domain families.

	C-terminal docking domain ➡														N-terminal docking domain ➡													
Interface	6939	6940	6942	6943	6945	6946	6949	6951	6953	6954	6956	6957	6959	6960	3	4	6	7	10	11	15	17	19	20	23	24		
VirA/FG	L	D	L	L	A	V	N	L	I	E	A	V	L	L	A	K	I	L	F	K	L	R	A	A	L	L		
SnaE2/E3	L	D	L	L	A	V	N	L	V	E	A	L	L	L	A	A	L	L	V	A	V	R	E	A	L	L		
ChiB/C	L	W	L	L	G	V	G	L	V	N	A	C	L	L	P	R	I	F	L	K	I	R	E	A	L	L		
ChiE/F	V	E	L	L	W	L	G	A	R	E	E	R	R	L	Q	D	M	M	K	-	-	-	E	S	R	L		
ChiC/D	L	D	L	F	A	L	R	E	L	E	M	V	-	-	L	Y	M	E	V	T	F	R	E	K	V	F		
KirAV/AVI	F	A	L	L	A	L	A	P	D	D	V	F	L	I	E	D	L	R	L	R	A	L	V	T	R	L		
KirAII/AIII	D	A	L	L	G	L	G	V	V	D	A	L	E	L	L	R	V	L	V	A	A	L	D	I	A	L	L	
OzmN/H*	-	-	S	L	R	Q	S	L	K	Q	E	Q	-	-	-	-	-	-	-	-	R	M	D	K	L	L		
OzmH/I	P	D	L	E	R	F	G	L	A	A	V	L	L	L	F	R	I	L	Y	V	V	A	Q	A	L	L		
OzmK/L	E	A	L	I	R	I	-	L	E	E	V	S	L	L	S	R	D	L	S	R	L	-	Q	K	L	L		
TaiE/K	T	D	L	L	A	I	G	L	A	G	A	D	I	W	S	K	I	F	L	R	L	R	E	A	A	L		
TaiL/M	L	D	V	L	R	V	G	L	V	E	A	E	L	L	L	R	V	L	L	K	I	P	E	A	M	L		
TaiD/E	L	A	L	L	R	F	G	M	L	D	I	V	L	-	K	K	I	L	Y	R	L	T	S	A	V	L		
5923/2	L	A	A	A	D	A	G	H	A	A	L	A	I	L	K	D	L	L	T	Q	-	I	E	L	R	L		
5925/4	L	D	E	L	E	I	-	L	D	D	L	A	L	L	L	Y	L	R	L	L	S	L	A	A	L	K	Q	
BryA/B	L	D	L	L	K	I	G	V	I	E	A	D	L	I	Y	R	I	L	L	Q	I	F	E	A	K	L		
BryB/C	L	K	I	L	K	V	N	L	I	D	A	E	L	F	A	K	I	F	L	K	I	I	Q	A	M	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: ABS90470.1)).

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/Green = inter-docking domain hydrophobic interface

Blue/Blue = Intra-docking domain hydrophobic interface

Pink/Pink/Purple = inter-domain polar interactions

Residue 6956 participates in two types of interactions, as indicated.

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.; Krug, D.; 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.; Gölz, 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.; Krug, D.; 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.; Boaisa, 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, Ó. *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.; Platzter, 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.; Arikan, 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.; Rajski, 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.; Pieretti, 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-Suwailem, 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 Annaïval,[†] Fanny Risser,[†] Sabrina Collin,[†] Pierre Roblin,^{‡,§}
Christophe Jacob,[†] Arnaud Gruez,^{*,†} Benjamin Chagot^{*,†} and Kira J. Weissman^{*,†}

[†]UMR 7365, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA), CNRS-Université de Lorraine, Biopôle de l'Université de Lorraine, Campus Biologie Santé, 9 Avenue de la Forêt de Haye, CS 50184, 54505 Vandœuvre-lès-Nancy Cedex, France

[‡]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

We aimed to strengthen the case that our observations concerning the VirA^{CDD}/VirFG^{NDD} 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^{CDD} and VirFG^{NDD}, 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^{CDD} 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^{NDD} (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^{CDD} and VirFG^{NDD} from the homologous pristinamycin PKS (SnaE2^{CDD} and SnaE3^{NDD}, respectively (page 3)), despite overall low levels of sequence identity between the docking regions (53% between VirA^{CDD} and SnaE2^{CDD} and 39% between VirFG^{NDD} and SnaE3^{NDD}¹¹).

We next extended this analysis to the remaining 25 putative VirA^{CDD}/VirFG^{NDD}-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^{CDD} and VirFG^{NDD} (analysis of each pair relative to VirA^{CDD} and VirFG^{NDD} is presented on each and highlighted in pink), while for 3 others, only the N-terminal docking domain shares predicted properties with VirFG^{NDD}. 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	KirAVI (coiled coil?)
SnaE2	SnaE3	MInD	MInE
AtcE	AtcF	OzmN	OzmH
BaeM	BaeN (coiled-coil?)	OzmH	OzmJ
Bat 1	Bat2	OzmK	OzmL
BryA	BryB (coiled-coil?)	RhiC	RhiD (coiled coil?)
BryB	BryC (coiled-coil?)	RizC	RizD (coiled-coil?)
ChiB	ChiC	Ta-1	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
KirAII	KirAIII		

Key:

Red = For VirA^CDD/VirFG^NDD observed α -helices in the complex; for all others, α -helices predicted by PsiPred

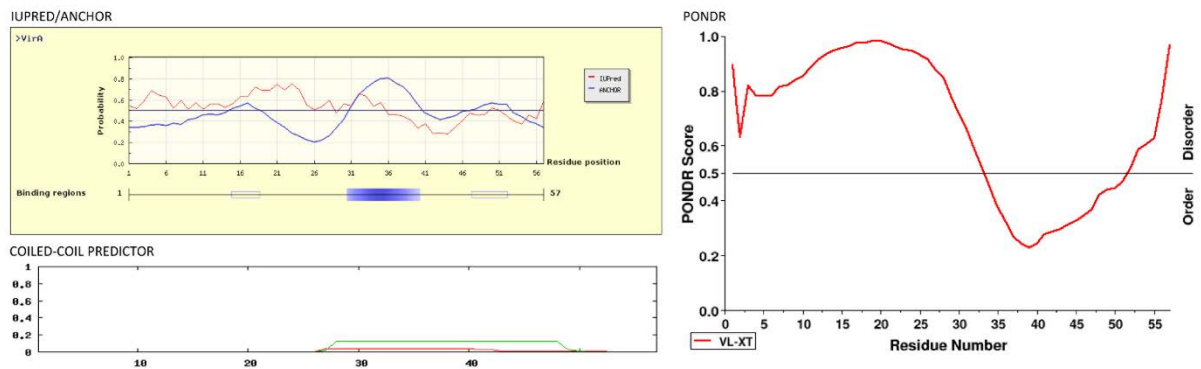
Blue = start/end site of expression constructs for VirA^CDD/VirFG^NDD, respectively

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^CDD

29 40 46 53

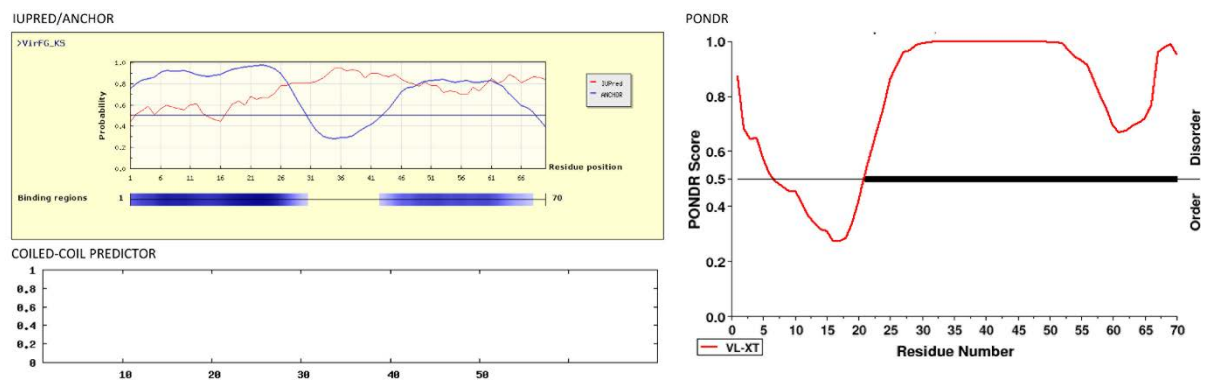
AGRAAPVPAAAAEPAALPLPVTGAGEPSQADLDALLSAVRDNRLSIEQAVTLLTPRR



VirFG^NDD

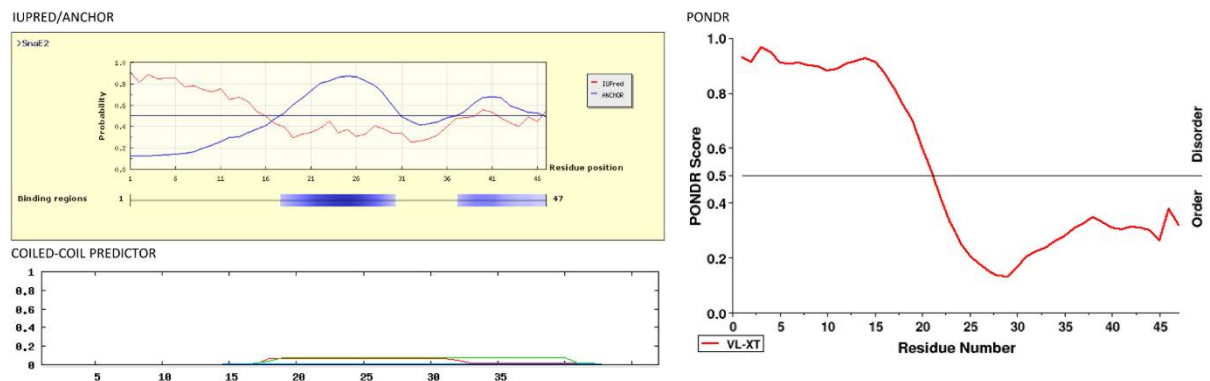
3 11 15 25

MDAKEILTRFKDGGLDRAAAQALLAGRTPAAAPRPSEPAAPTRPAVPAPVPAVEPAAGTTVAAEGTAGRP



SnaE2^CDD

TSRTAPQPAGAVPAGSVSAADLDALLAAVRDNRLTVEQALALLPQHT



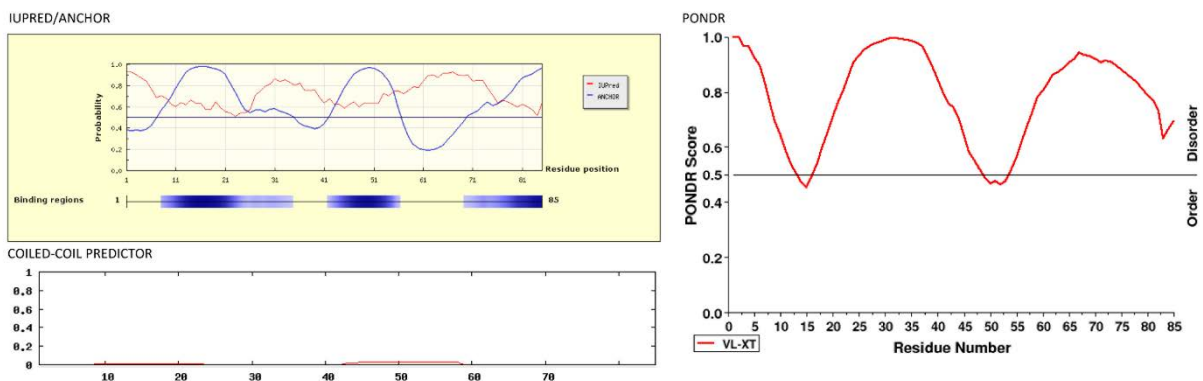
OBSERVATIONS

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^CDD

SnaE3^NDD

MPAEPDAVDRDRAAALLRAVAGQRGPAAGAATASARPAPGGVDRAEAAALLRAAAGERRPGGDGPAGG
SMGRTLPAEAGSAGRVL



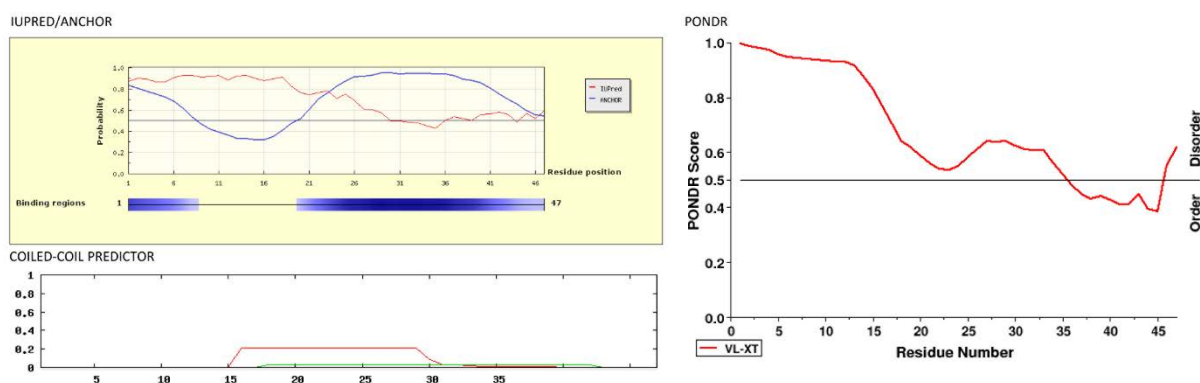
OBSERVATIONS

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^NDD

AtcE^CDD

HAGHAAAPPGADGAGPERATDVEDMLRRLAAGELTVAEAYTNLGGPA



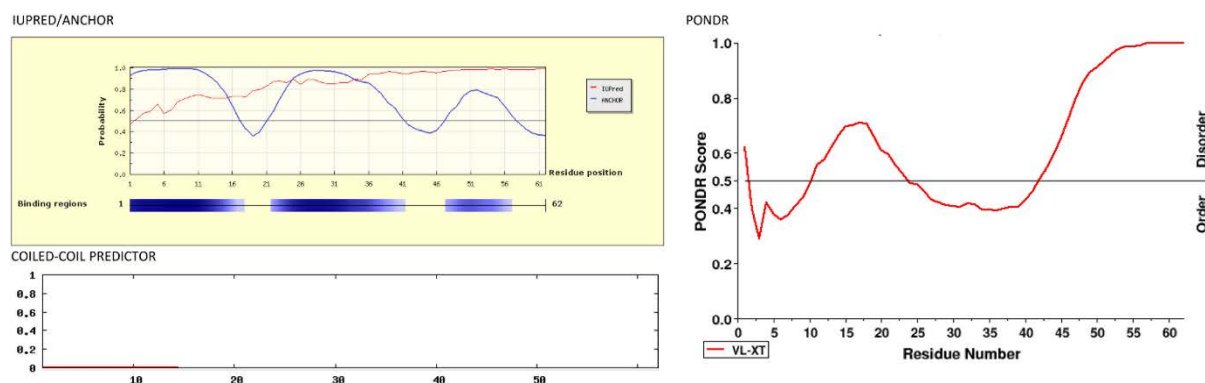
OBSERVATIONS

Predicted interaction region (ANCHOR) overlaps well with predicted α -helices. The docking region is predicted to be borderline disordered (IUPred/PONDR) and not to form a coiled-coil.

Agreement with VirA^CDD

AtcF^NDD

(M) NSAVKRVLELVETRQISPEEGRRRLTALGTAGDRPVAGEAGDRPRPDGRAAAPPASEGPPSD



OBSERVATIONS

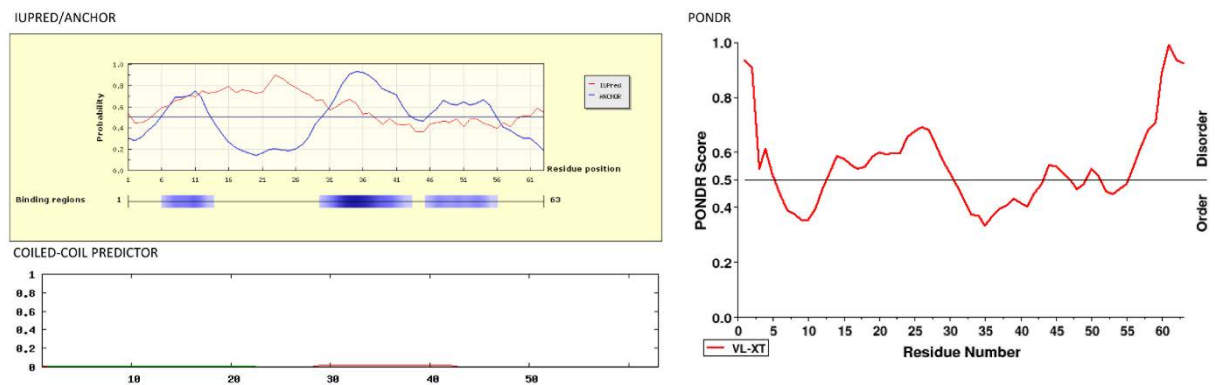
Predicted interaction regions (ANCHOR) overlap well with predicted α -helices. The docking region is predicted to be largely disordered (IUPred/PONDR), but not to form a coiled-coil.

Agreement with VirFG^NDD

BaeM^CDD

31 42 48 56

LTEHKETIETFYRTEETETEAAAPESKEYTDQEIIAMMKQVSEGTLD~~FKRVQDI~~IEGSKTYES



OBSERVATIONS

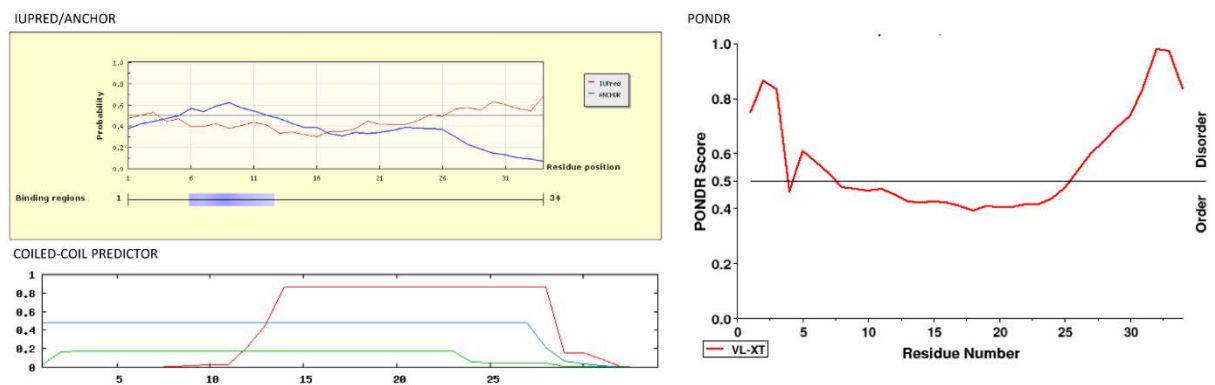
Predicted interaction regions (ANCHOR) overlap well with the predicted α -helices. The docking region is predicted to be borderline disordered (IUPred/PONDR), and not to form a coiled-coil.

Agreement with VirA^CDD

BaeN^NDD

2 10 15 31

(M) N~~HKELL~~DAYRSGTLTIAEVEQKLQAFKRTTAKRP



OBSERVATIONS

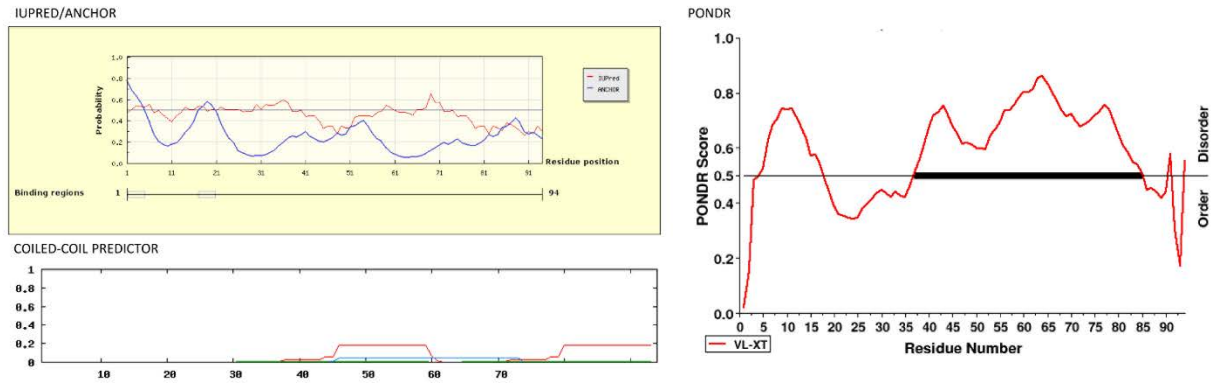
The first predicted α -helix is predicted to be an interaction region (ANCHOR). The docking region is predicted to be borderline disordered (IUPred/PONDR), but to form a coiled-coil.

Does not agree with VirFG^NDD; may be an alternative type of docking domain?

Bat1^{CDD}

65

VRIVSAQPREASSARTLDVVVGERLDTQSSDDRTMKMLEDFRDGNVKIEEVEGILDRWTKTQSSDDRT
 MKMLEDFRDGNVKIEEVEGILDRWTK
 76 82 92

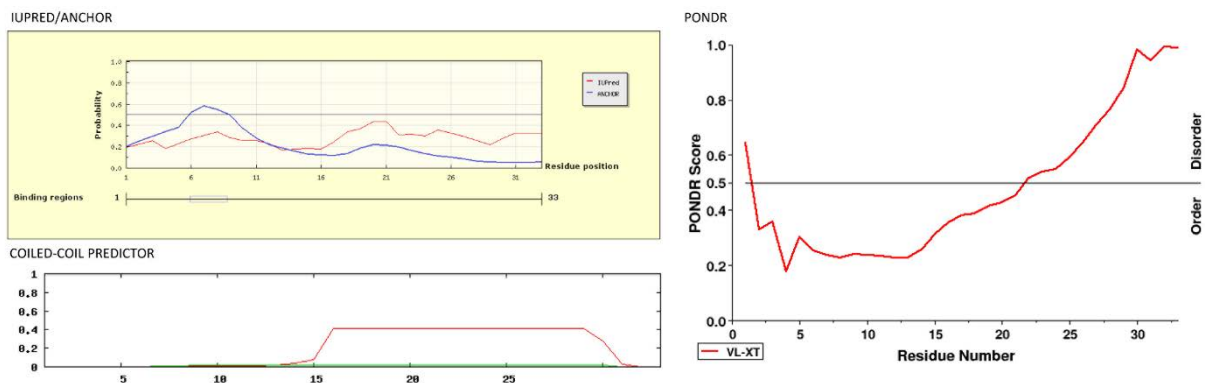


OBSERVATIONS

The predicted α -helices are not predicted to be interaction regions (ANCHOR). The docking region is predicted to be disordered (IUPred/PONDR), and not to form a coiled-coil. Poor agreement with VirA^{CDD}

Bat2^{NDD}

(M) TRRDLLFAYKNRSLTQELVAALTELNTERRF5
 5 9 16 25

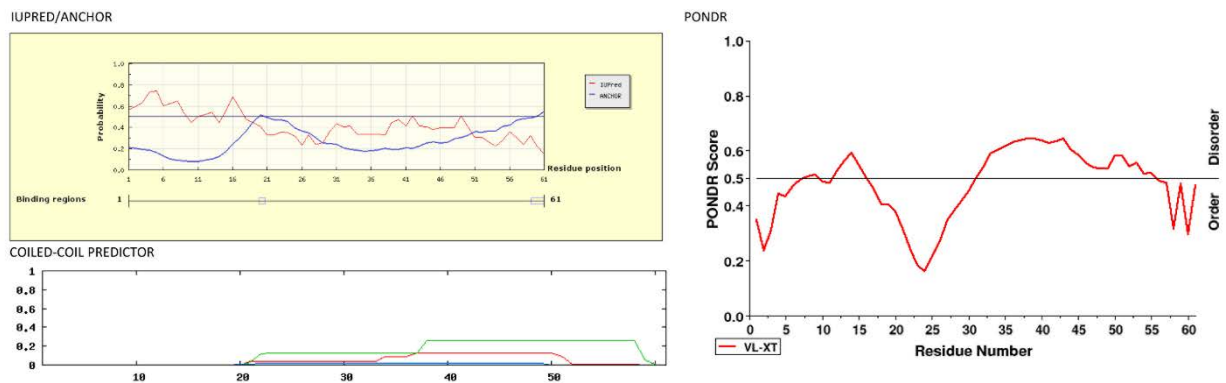


OBSERVATIONS

The secondary structure elements are not predicted to be interaction regions (ANCHOR). The docking region is not predicted to be disordered (IUPred/PONDR), and not strongly predicted to form a coiled-coil. Poor agreement with VirFG^{NDD}

BryA^{CDD}

24 33 38 47
 RKQIISNDEKQHQPSTIFTPTSLDELLKKIQEGTLGIEEADQLIDELPDYHLDMEHELL



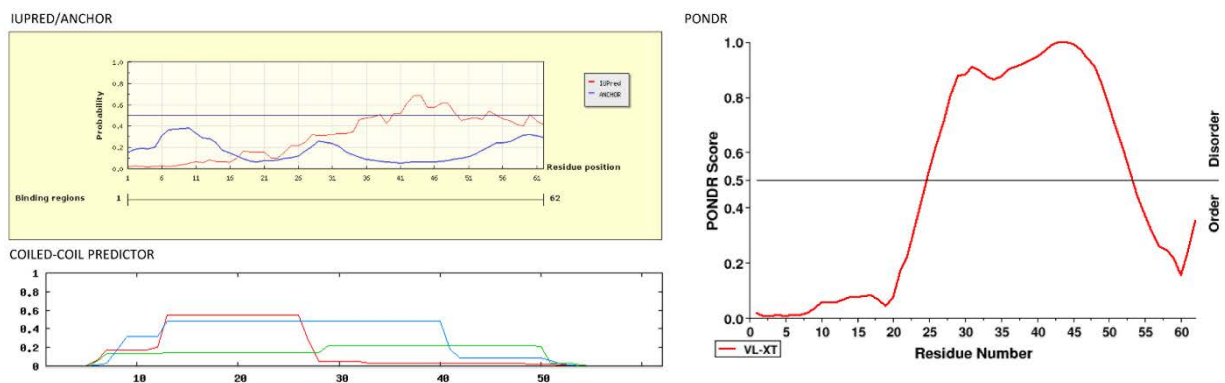
OBSERVATIONS

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 (PONDNR), and not to form a coiled-coil.

Poor agreement with VirA^{CDD}

BryB^{NDD}

1 19 24 43
 (M) VVVEEFFVSYRDLKALQDEKISFEEAKYKLIK RKDKKSKQRLNHDRELNRSMKITPKIVNN



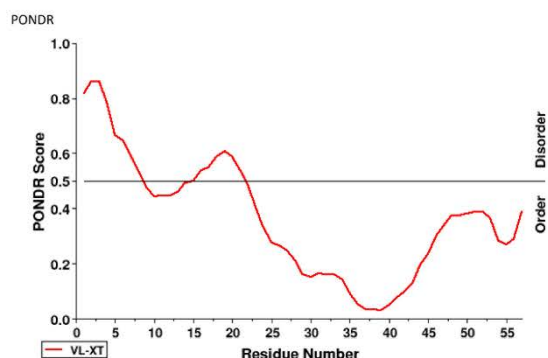
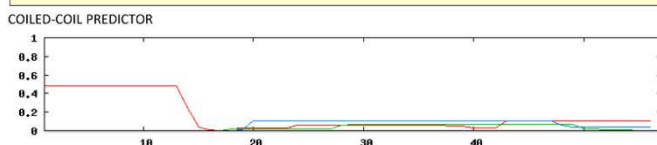
OBSERVATIONS

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/PONDNR), and the overall domain to form a coiled-coil.

Does not agree with VirFG^{NDD}; may be an alternative type of docking domain?

BryB^CDD

KQKIEQKNDIQNLSSSSYFSEDNQEVALSLKEILKKVEGNTLSIDKAEKLFQFSLK

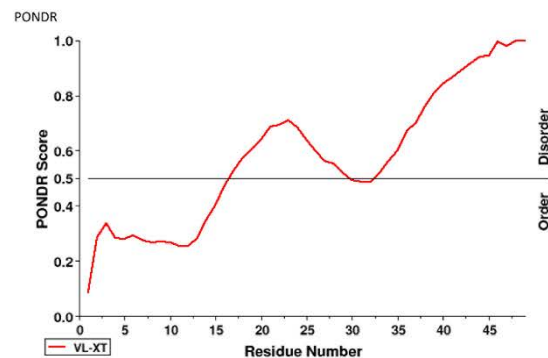
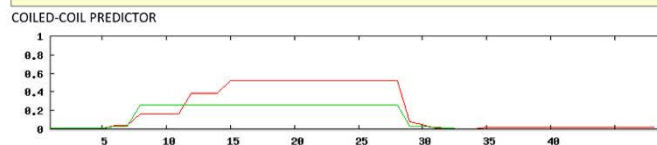


OBSERVATIONS

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^CDD

BryC^NDD

(M) KAKVIFQGLKSNEISIEQAEKMLLSLDTESKHGHPESKLSDIKSDLHDD



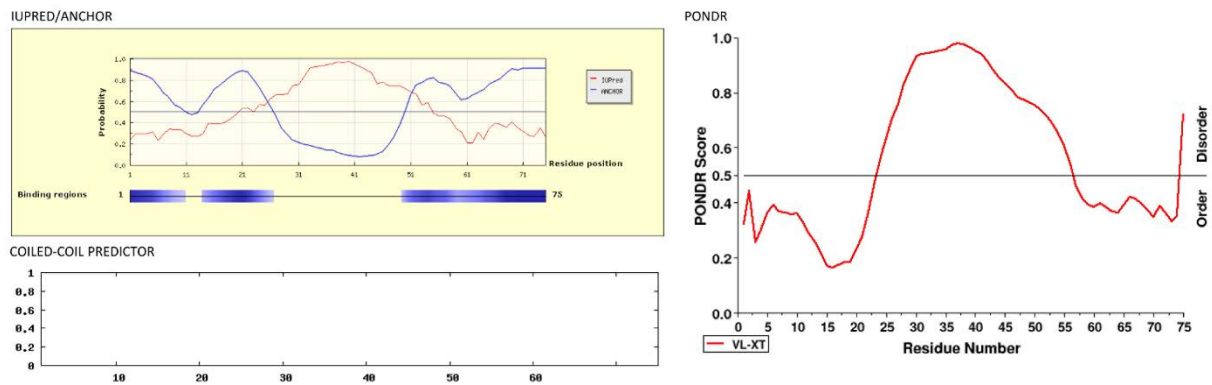
OBSERVATIONS

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^NDD; may be an alternative type of docking domain?

ChiB^CDD

LSSFAADVRLPGAAKIDTAALFGDGAAPAPAPAPAPVDEPRGGEPEGDEPLWRLLRGVESGALDVNEA
 CRLLELP
 70



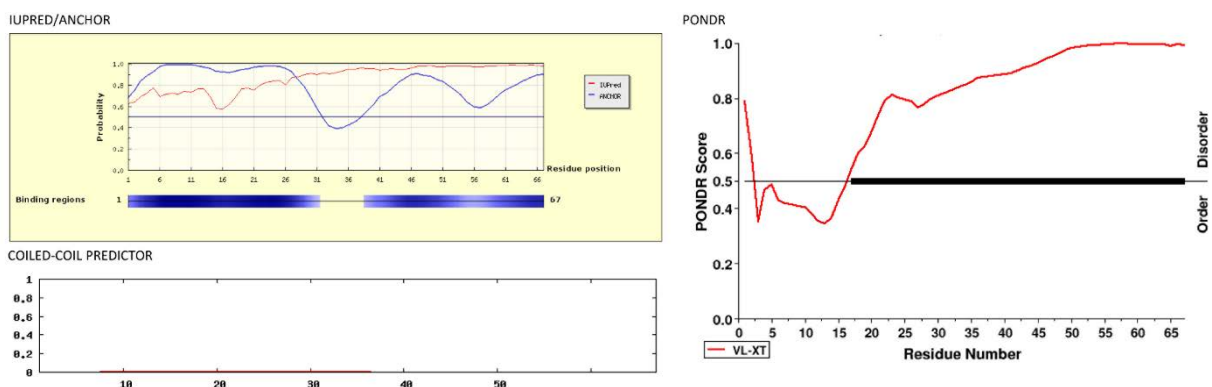
OBSERVATIONS

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^CDD

ChiC^NDD

(M) KRDPRI⁴FEQLK¹²NGQISRAEAHRL¹⁶LKARS²⁸DGREQEATGPGAAEGAAAEARAPSAPFPQPASSDSD
 VG



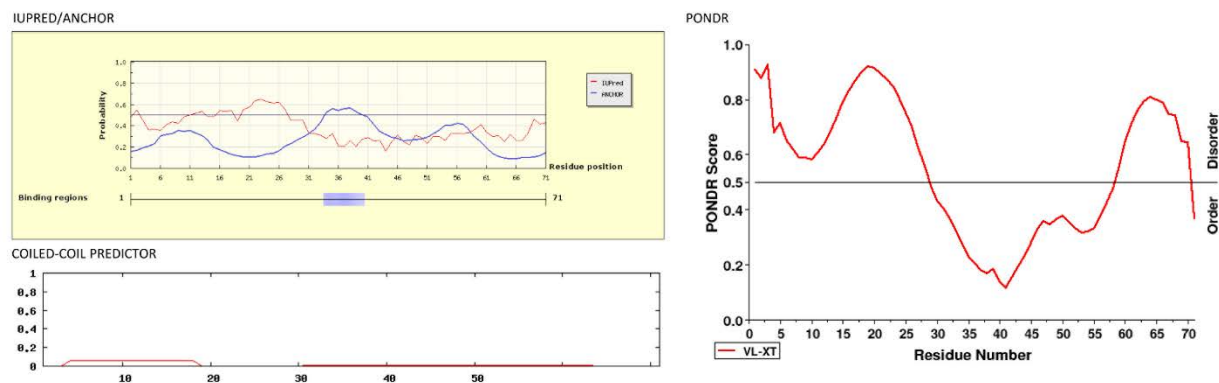
OBSERVATIONS

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^NDD

ChiC^CDD

ERDPALAEIKAQVLRAQGAPGPGAPELAVG³¹HDELWRL⁶²LGAL⁶⁵GDAELDALGDAELDALF⁶⁸NALRSRPELE
GMV



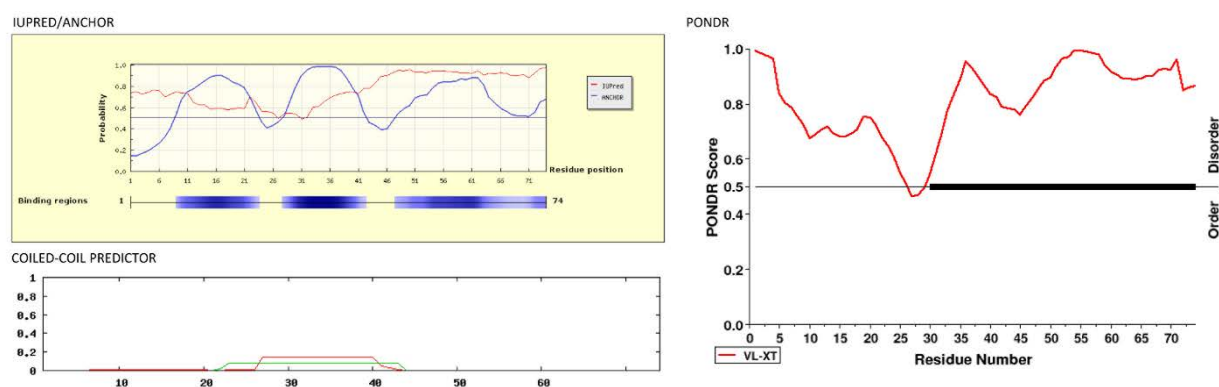
OBSERVATIONS

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^CDD

ChiD^NDD

MQRP⁵DRGR⁴³LYRMEKAVTDAMF⁴³DE⁴³RSREEKLSVFRALRDRLRRTPPQHGGAADGAGPRAPVLAPDPQN
RHRPF⁴³P



OBSERVATIONS

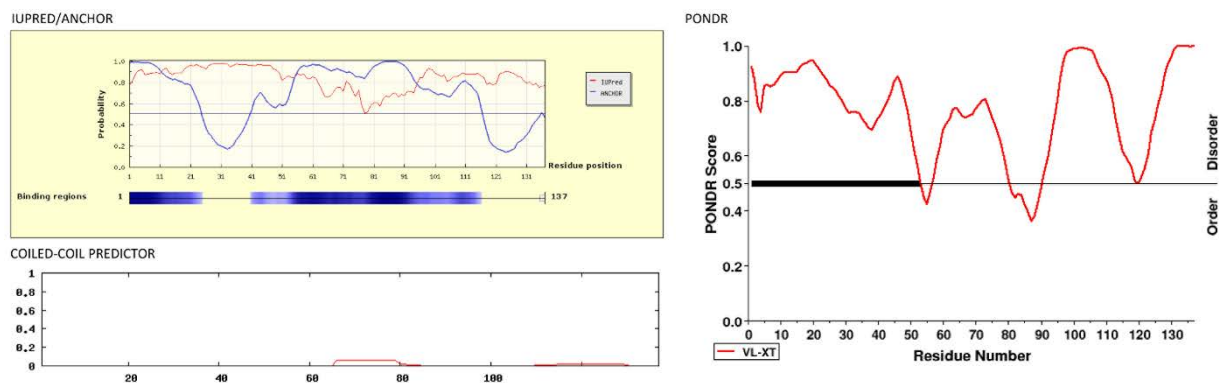
Predicted interaction regions (ANCHOR) overlap well with predicted α -helix. The docking region is strongly predicted to be disordered (IUPred/PONDR), but not to form a coiled-coil.

Strong agreement with VirFG^NDD

ChiE ^CDD

57

VDHHEAALRASEPPGVGGGAEPVRPRANPERPERADPPERAGVAPGRRAPSPPAAAVERFDGAVP**ED**
LAGEVARMSDAEVERLLGWLQGGGAGGAGGAGGAGATNGSSAVRASLDD**GQEREAEERRRLD**RGPSGEW
 R 79 99 105 108 119

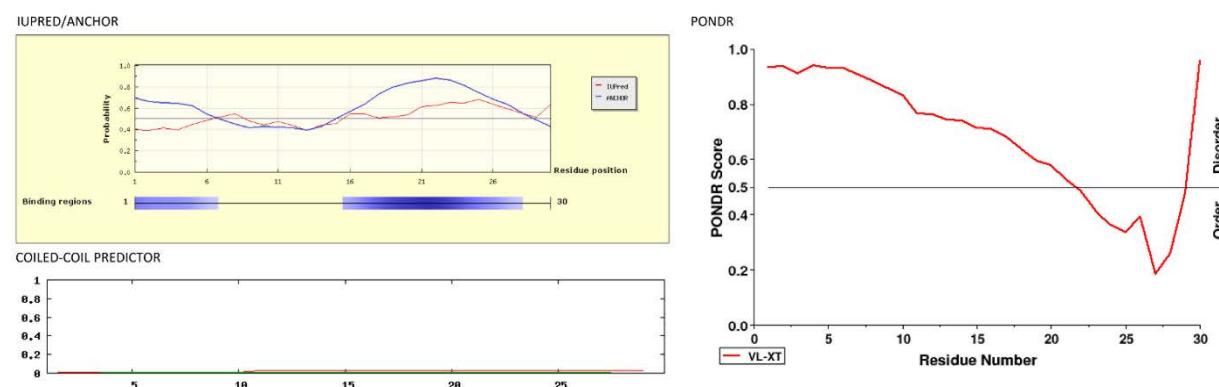


OBSERVATIONS

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 ^CDD

ChiF ^NDD

5 28
 MAMT**LRSQDL**MMDAK**RES**LVRL**LHQK**ANR

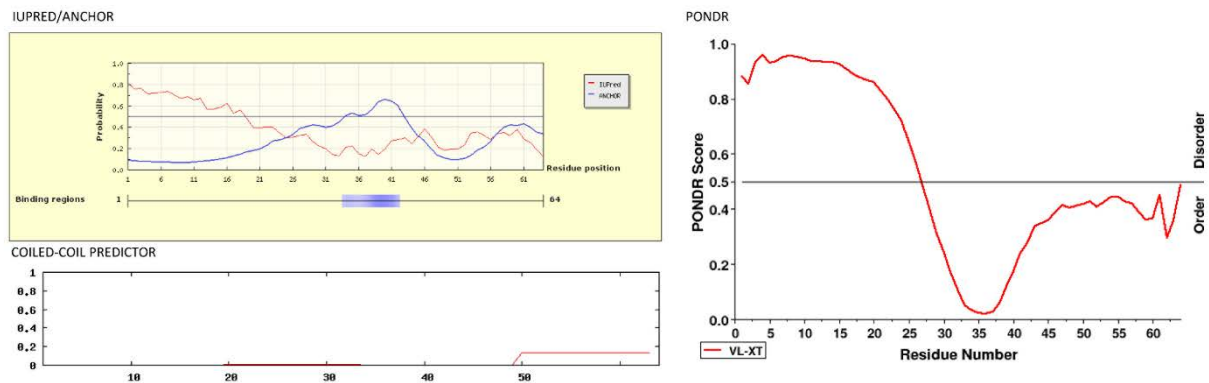


OBSERVATIONS

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 ^NDD

DisB^{CDD}

20 46 51 62
 SRRGARAPESRAKARSGAAGADLSTSLKALSGAVLREQFLAFGHDLAGVPGEELTRLYAILQEE



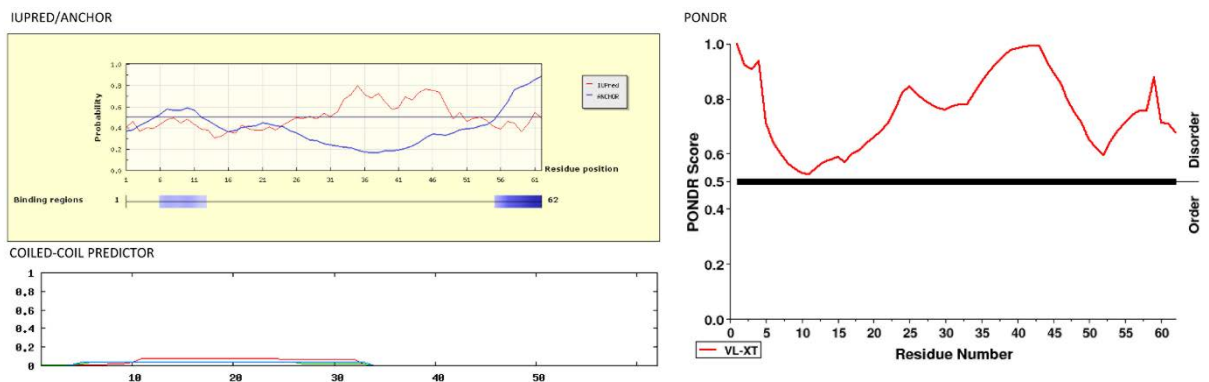
OBSERVATIONS

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^{CDD}

DisC^{NDD}

6 11 15 30
 MESAMTIQEFANLSAEEKVQVLRLRDRRASWQAAPGPAASAQPSLRPVITARPGDRFLPFP



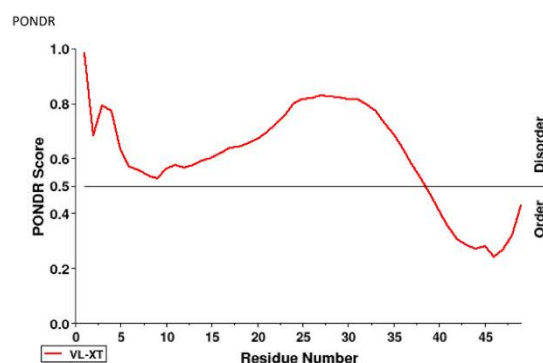
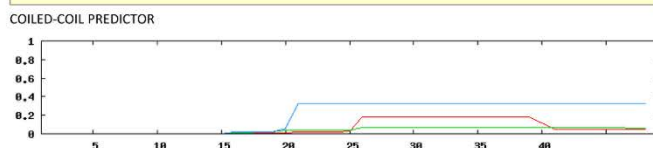
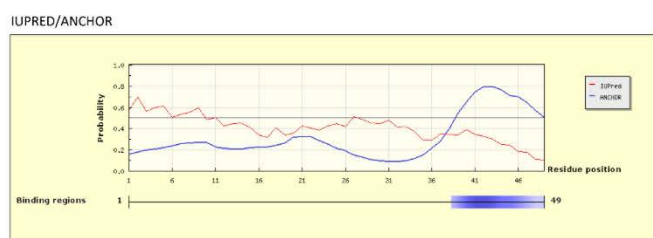
OBSERVATIONS

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^{NDD}

ElaK^CDD

22 31 33 47
SSEMTNISTYEPVQSLQQEFSLDDVVRQVEKGE~~MD~~AEANLLAVLSKQ



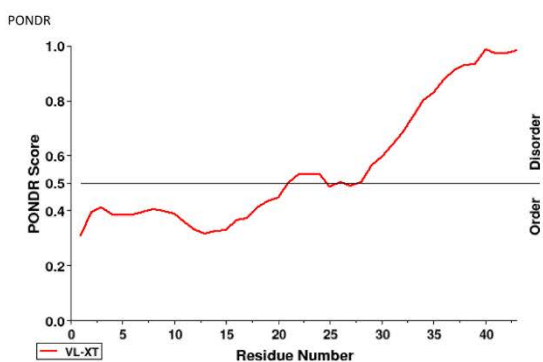
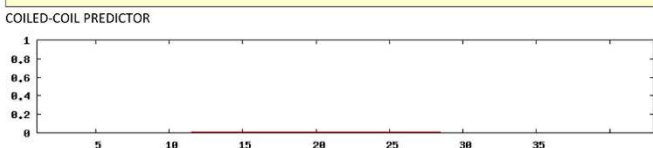
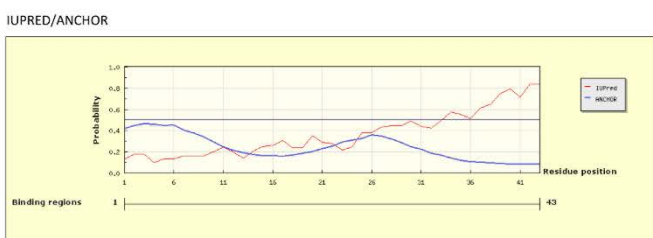
OBSERVATIONS

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^CDD

ElaO^NDD

2 10 17 26
(M) D~~SLAILKALS~~TGAI~~SAGAA~~REALKQLAVAPVSVKPGHDARATP



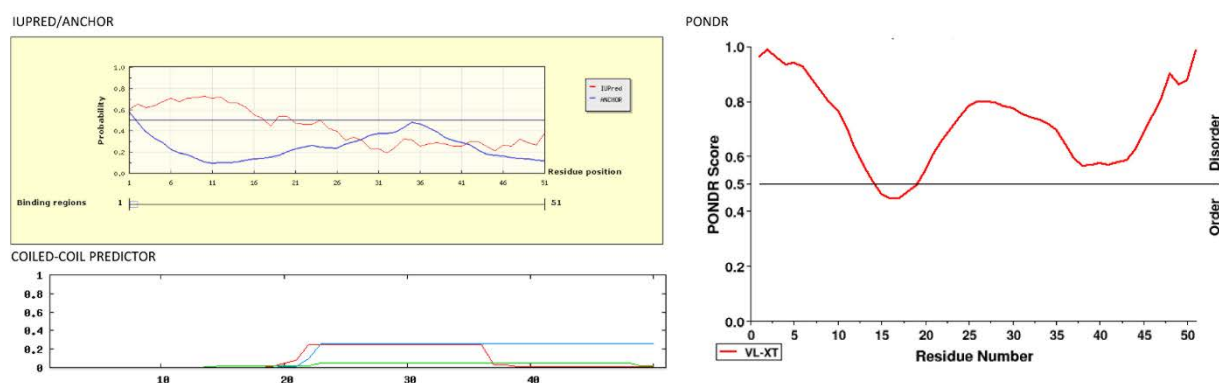
OBSERVATIONS

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^NDD

EtnH^CDD

20
CAEHPDDLARRQAGTQANGGAVLEERRAVEALAEELALGRISLDQTLDRRLRA

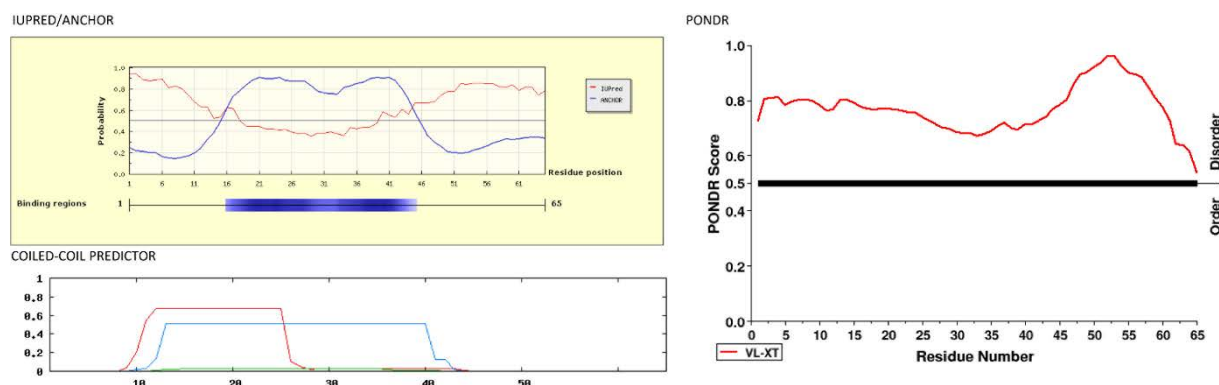


OBSERVATIONS

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^CDD

EtnI^NDD

10
(M) PEHSYEPRGESTVKSQIEALLREYQSGLLSEAELRARYAQLVASGSDGAPARPADRAGATQTVRR



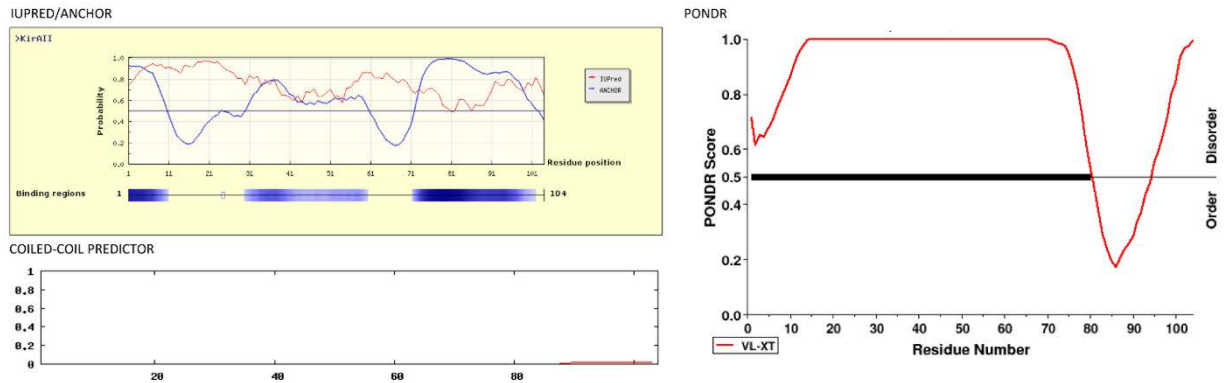
OBSERVATIONS

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^NDD; may be an alternative type of docking domain?

KirAII^{CDD}

HDKYGPLPATATPEPPAPVTSSATAPEPTSVFVASSAAAPAPIAPVAPSAAPAPVASSAAPEPP
 PAPVAADAVGALLAGLRDGSVTVDDALAE LRQGNVR
 72 85 91 99



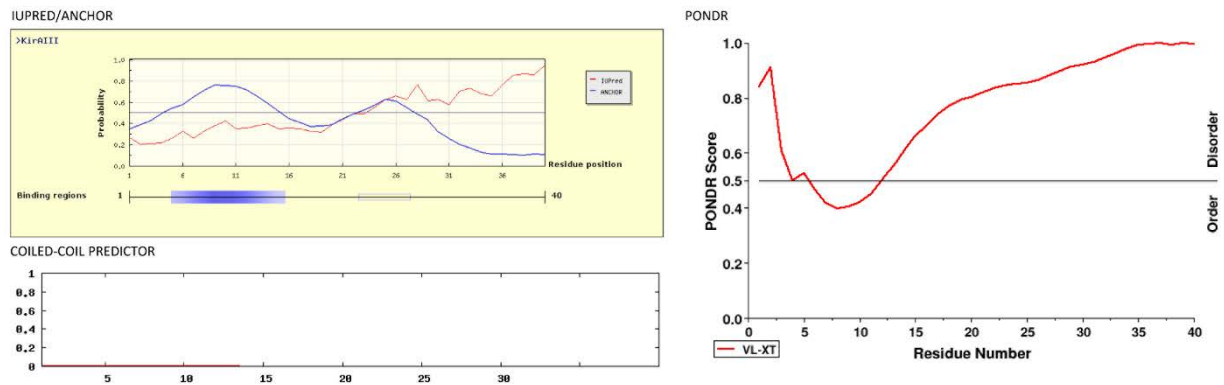
OBSERVATIONS

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^{CDD}

KirAIII^{NDD}

3 13 18 29
 MSRALRDVLA AVAAGELPDEIAESLLRGLSDFVPEPGYP



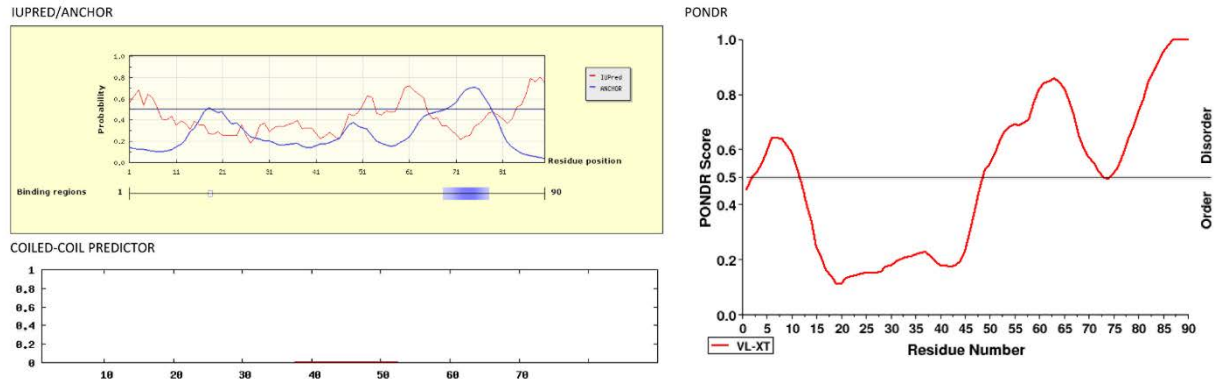
OBSERVATIONS

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^{NDD}

KirAV^CDD

33 52 61 66
 AGRFDGGTGGADDLPALFADFDRDLAAGLTEAATDEVFRDRVVERVRGLLTALSPATPEAP**PDLFAAAT**
 DD**EVFALIDA**ELGSPVRGQEDR
 71 79

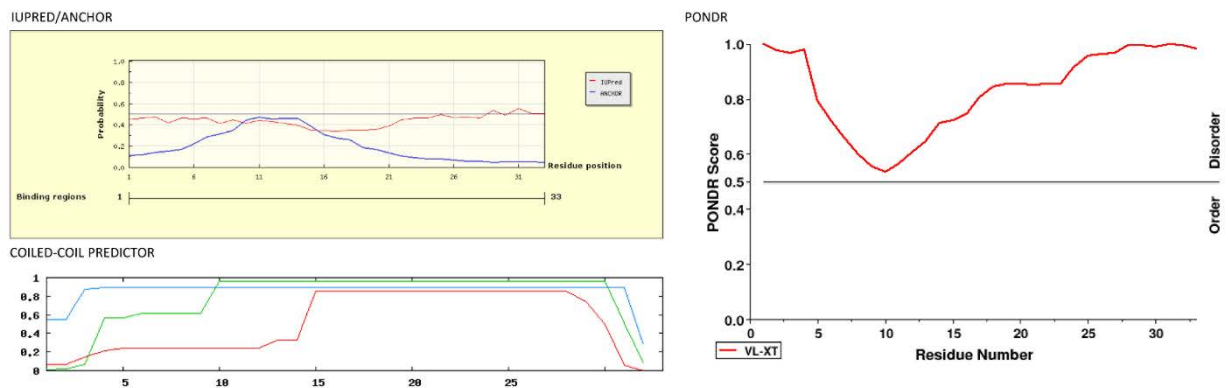


OBSERVATIONS

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^CDD

KirAVI^NDD

4 31
 MSS**EDKLRDYLRRVT**AELQV**T**RRRL**REVE**SRSG

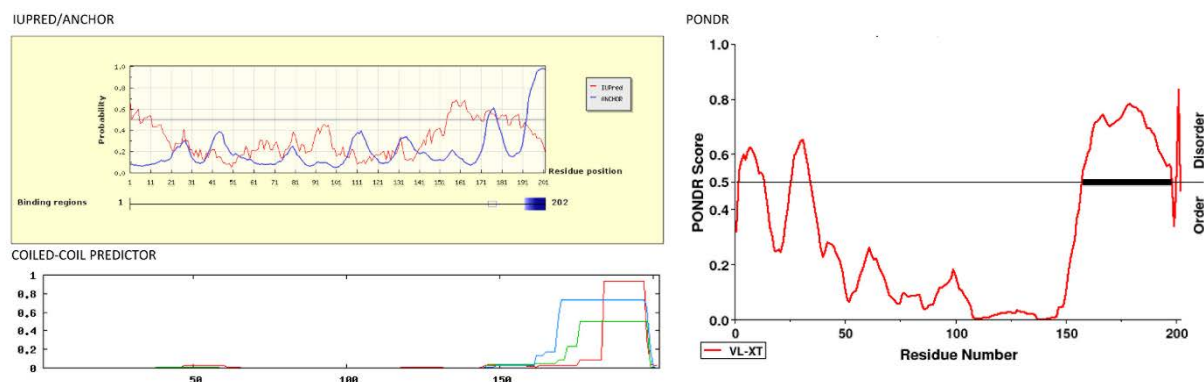


OBSERVATIONS

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^NDD; may be an alternative type of docking domain?

MInD ^CDD

49 60
 SSIRNRSADKPYNGNFREKSGFGLLEEIAYTLQTGREPFECRLAITCSNIGDLLEKLNQYIQGSRLQGV
 YEGNVTDFDSL YAPLFGGPEGHAYVQS LIQQNPAKLAQLWTIGCEISWEDLYEELKPVKLSLPLYP
 96 110
 FSKNRHWIHLNKDEAQKAEAFSNTLHNSNKDTADGEIMSLLEKAQTGEMNMKETSELIEELLFHE

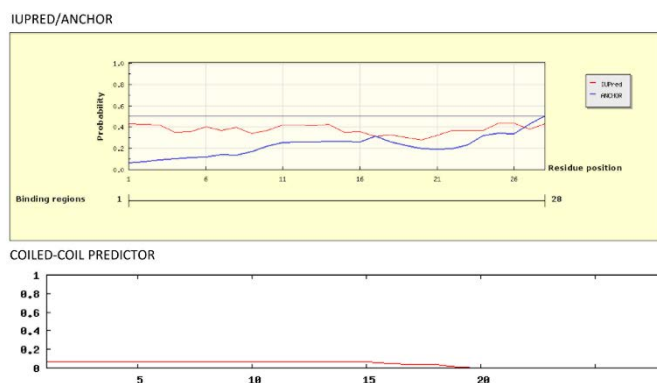


OBSERVATIONS

The predicted α -helices are not predicted to be interaction regions (ANCHOR). The docking region is predicted to be strongly ordered (IUPred/POND R), and not to form a coiled-coil. Poor agreement with VirA ^CDD

MInE ^NDD

3 16 21 23
 (M) NEKTMQIEQIVKKVKECSLTPEEGLELI



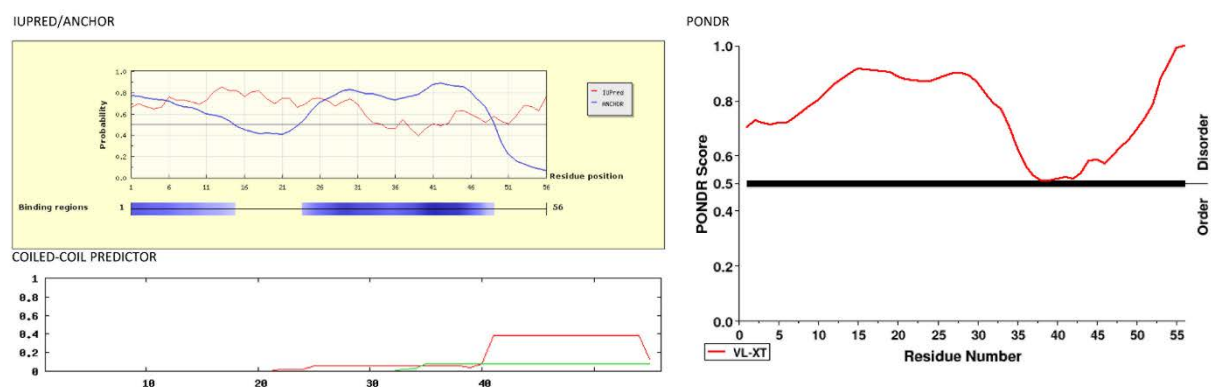
Sequence is too short for POND R

OBSERVATIONS

The predicted α -helices are not predicted to be interaction regions (ANCHOR). The docking region is predicted to be borderline ordered (IUPred/POND R), and not to form a coiled-coil. Poor agreement with VirA ^CDD

OzmN^CDD

AATGAAGTAAGTAAGPARAAEPEAGIEQLVDSLSDTEVDSLRLRLGSLVLOKQEEQR



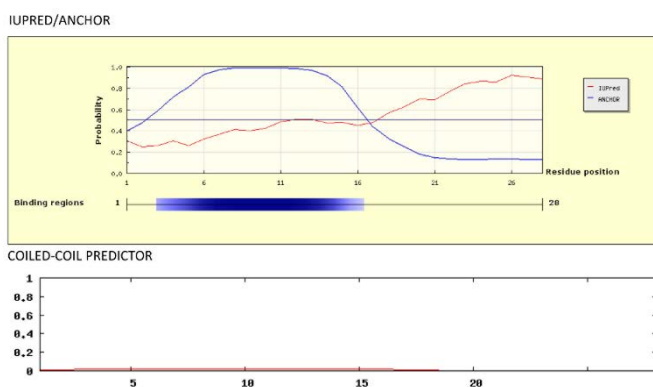
OBSERVATIONS

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^{CDD}

OzmH^NDD

6 24
MRRRS**MDDKRALLLKLLASVROQA**APATA**PRRTA**



Too short for PONDRA

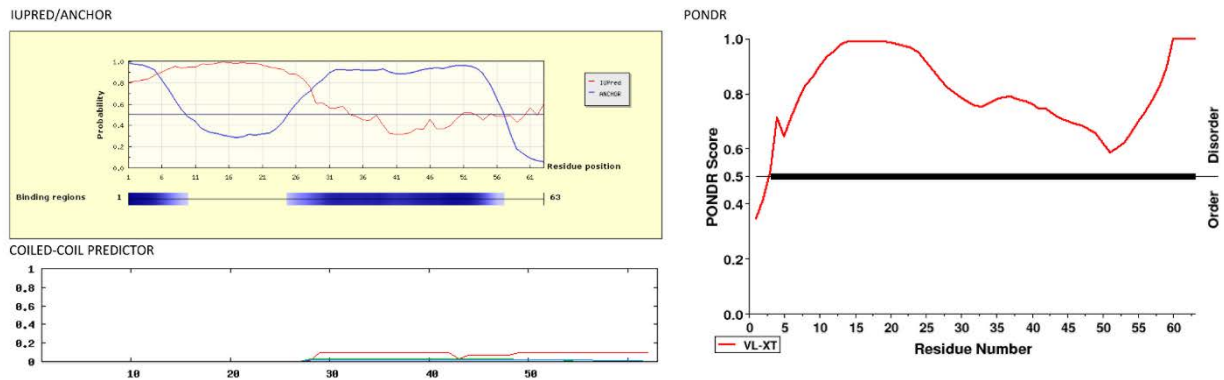
OBSERVATIONS

Predicted interaction regions (ANCHOR) overlap well with predicted α -helix. The docking region is predicted to be borderline disordered (IUPred), but not to form a coiled-coil.

Strong agreement with VirFG^NDD

OzmH^CDD

28 43 46 60
 IELVGPDEPTAGSGEPPADGTGDTAAPDELI²⁸AKLEARFAAGE⁴³LSAAEVLDLLDAELAT⁴⁶RE⁶⁰RR



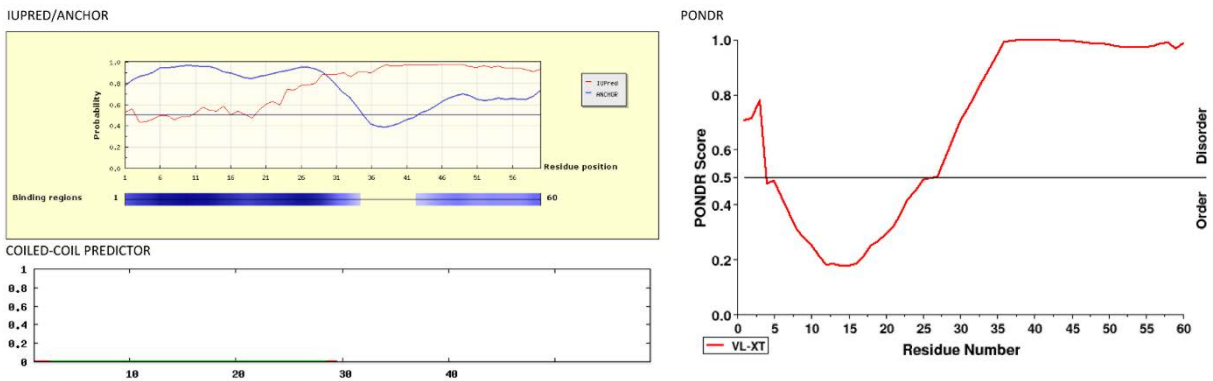
OBSERVATIONS

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^CDD

OzmJ^NDD

3 16 20 29
 MSTAEFRDILD³RYARGAVGAEQARDLLA¹⁶ARRPAQAPGRPTAPAPSGQDAATGPADRPETL²⁹



OBSERVATIONS

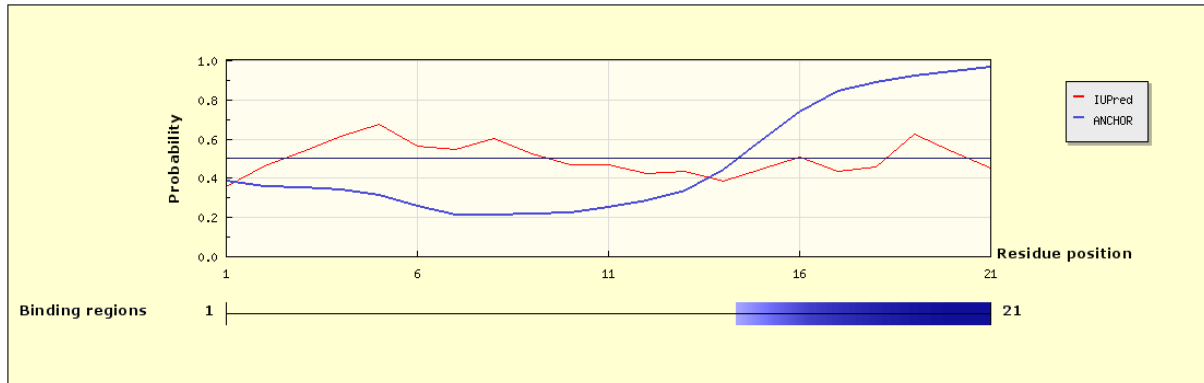
Predicted interaction regions (ANCHOR) overlap well with predicted α -helices. The docking region is predicted to be borderline disordered (IUPred) to somewhat ordered (PONR) but not to form a coiled-coil.

Agreement with VirFG^NDD

OzmK^CDD

2 8 12 19
EAE^{RED}LV^{RED}RIEELP^{ED}TVATLLA

IUPRED/ANCHOR



Sequence is too short for PONDR/COILED-COIL PREDICTOR

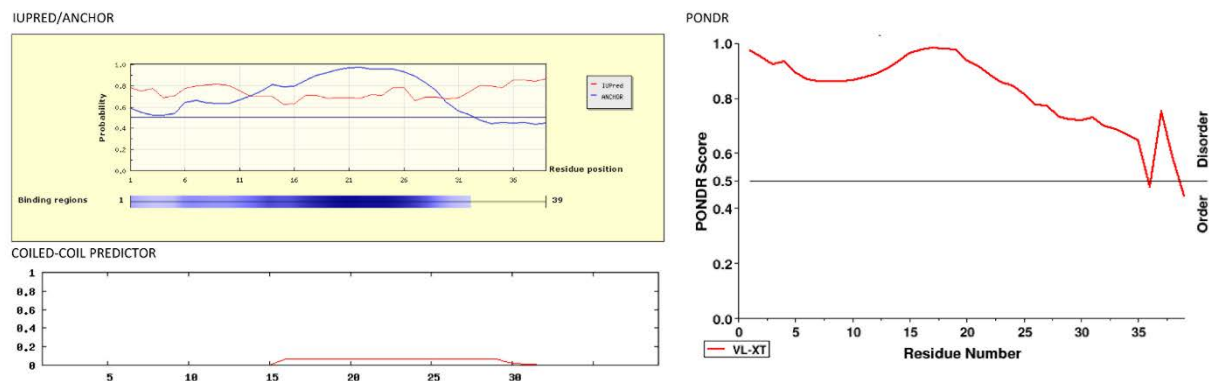
OBSERVATIONS

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^CDD

OzmL^NDD

3 13 16 29
MS^{RED}DLAT^{RED}SRKARLS^{PQ}KRALLEKLSGAARTSGTTRIPR



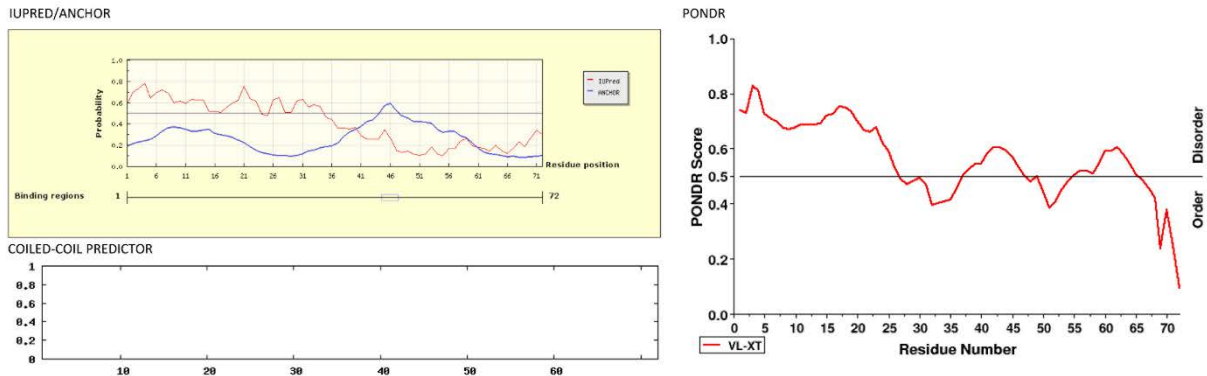
OBSERVATIONS

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^NDD

RhiC^CDD

ADNQRTAMAVSPNDLDHAVSAGNNVRGQSRPIEIHPPFADQLEGLFLSGELSIDSLNIVSPGVAEIR
EVSI



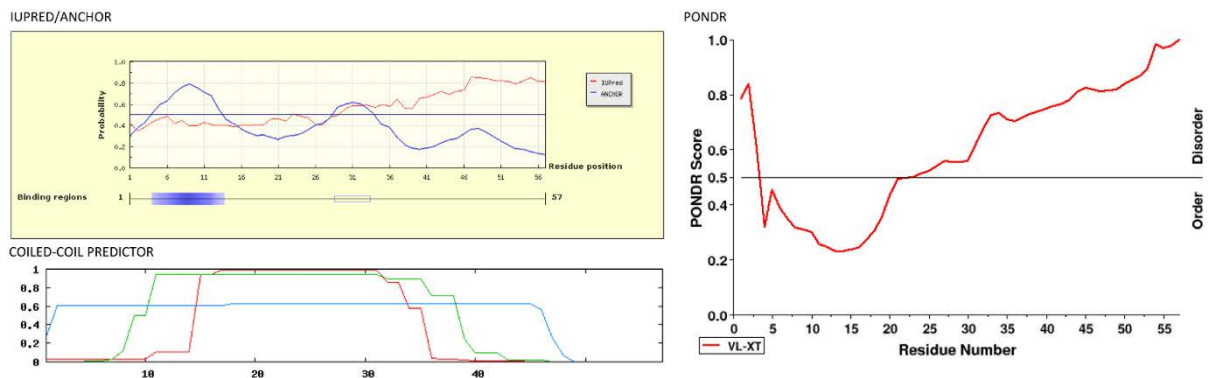
OBSERVATIONS

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^CDD

RhiD^NDD

(M) SSQEFQSI IAGLQNRQITVAEAKQRLQRLKAASATLGTETEAADKAVSGTGRAQEP



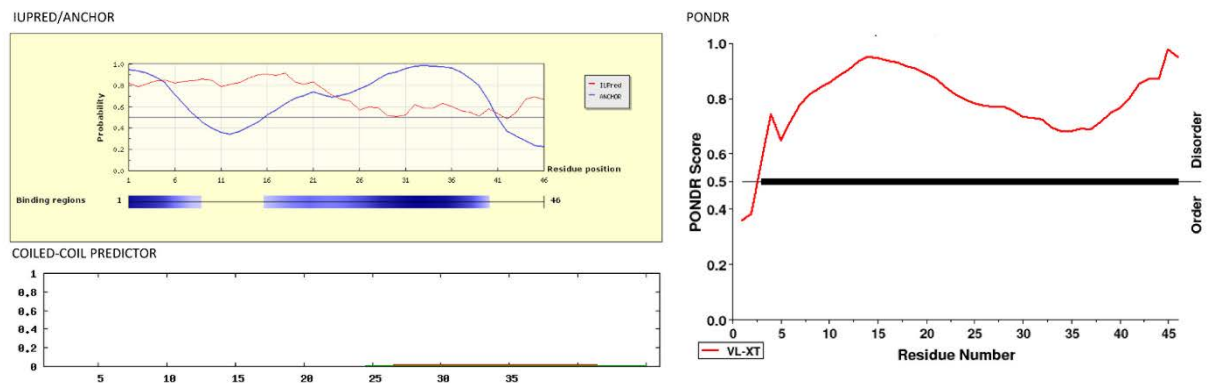
OBSERVATIONS

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^NDD; may be an alternative type of docking domain?

RizC^CDD

29 37 40
RALVTEAPPRGATQSANGHALEGMSDGNLEELFARLSGVSSQKGEA



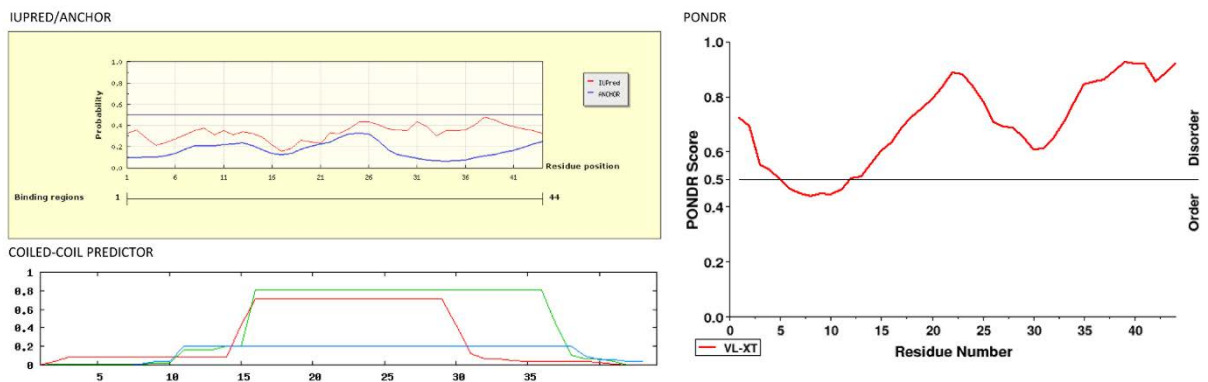
OBSERVATIONS

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^CDD

RizD^NDD

7 42
MPFKND~~SALNEFRLLSREEKLQRIAALIEKKSSAKS~~ASSEAF~~A~~



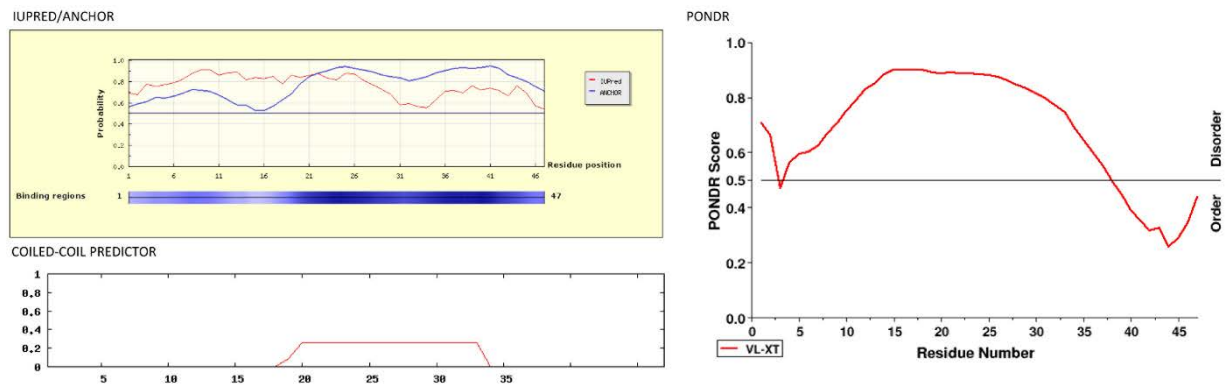
OBSERVATIONS

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^NDD; may be an alternative type of docking domain?

Ta-1 ^CDD

21 29 35 44
GNEIQKTVSAGAPAQGEPPSLDELLRQVEAGELDPSVAQQFLTNSQS



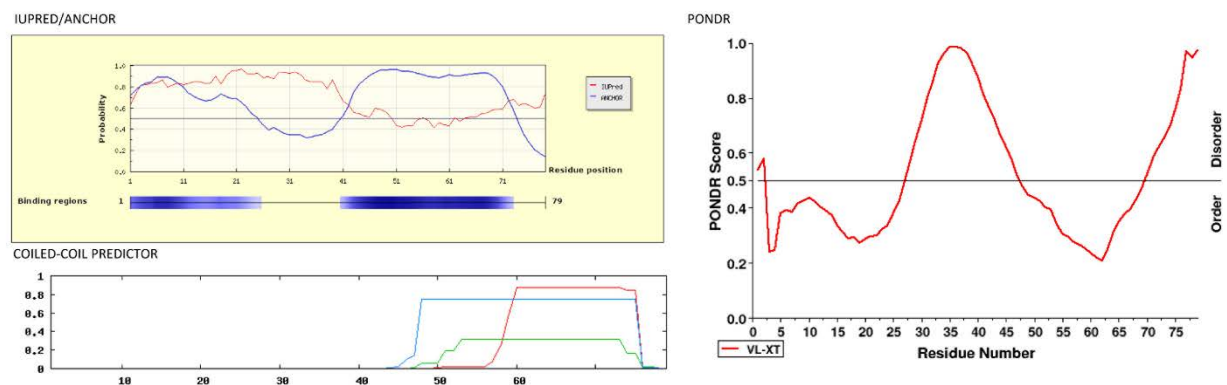
OBSERVATIONS

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 ^CDD

TaO ^NDD

43 55 58
(M) GEGHRIGVLQQPPRGRDCHEADDGSRRIPQPPVPQVGRPGRGAVMTRLEILQALQSGRMALADAK
KALAAIEAQVQPED
73



OBSERVATIONS

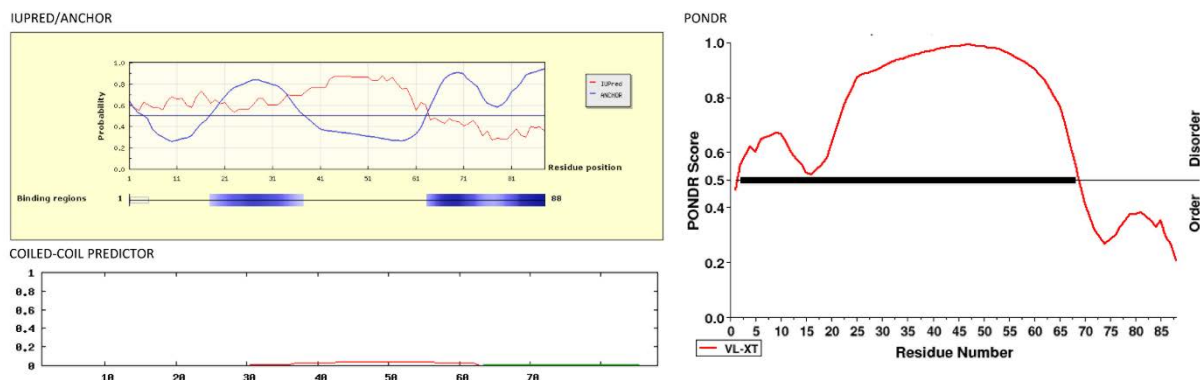
The predicted α -helices are predicted to be interaction regions (ANCHOR). The docking region is predicted to be borderline (IUPred) to mixed ordered/disordered (PONDR), and the second α -helix to form a coiled-coil.

Does not agree with VirFG ^NDD; may be an alternative type of docking domain?

TaiD^{CDD}

64

AERIARDGGRRAAEAAAPDAAMASDVAEASVSEATEASDASEASEASEASEASKAPADLA
ALLERFRAGQMDLDDIVDLV
74 81 87



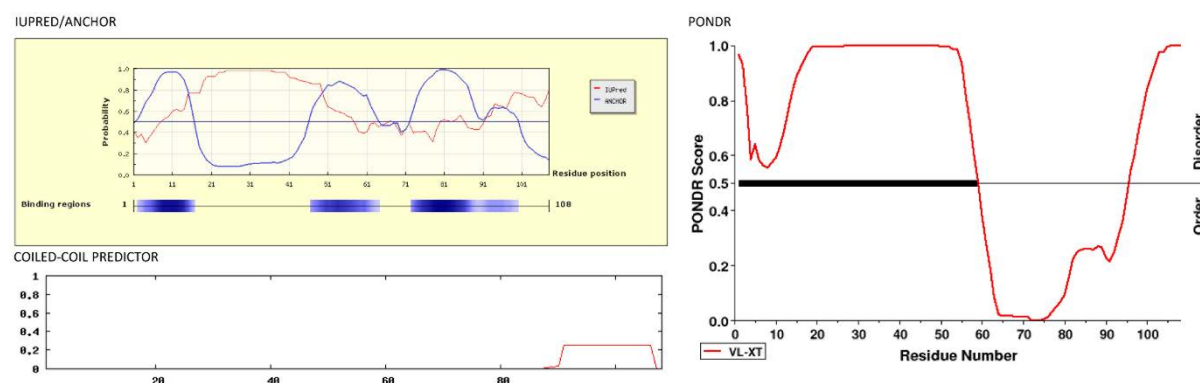
OBSERVATIONS

Predicted interaction region (ANCHOR) overlaps well with the predicted α -helices. The docking region is not predicted to be disordered (IUPred/PONDR), nor to form a coiled-coil.

Strong agreement with VirA^{CDD}

TaiE^{NDD}

MPRWRPTWLKRLSCPQPRHPHTPKHPHTPKHPKHPRHPKHPRISRRCSSASARGKWISTTSSTW
FDGDHVNKKDILLAYREGLLDTGSAQRVLDALRERSASAP
76 89 91 103



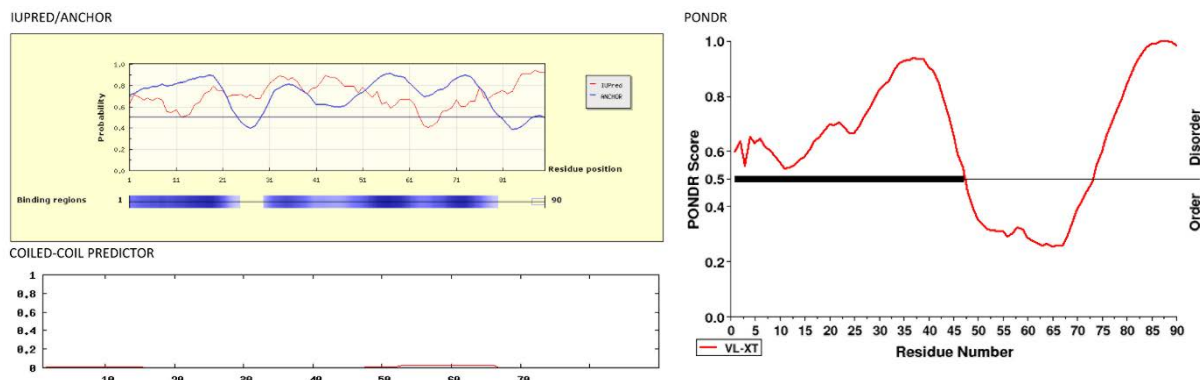
OBSERVATIONS

Predicted interaction regions (ANCHOR) overlap well with predicted α -helices. The docking region is predicted to be borderline disordered (IUPred) or ordered (PONDR) and not to form a coiled-coil.

Agreement with VirFG^{NDD}

TaiE^{CDD}

GTQLGARLPAAQAARAGAFAGVEPGEFDPARALPAAARAAAPPAHTDAAAHTD**TDALLRAIER**GELDAG
DADAIWRRMQSRAARPEPLAQP
 70 81



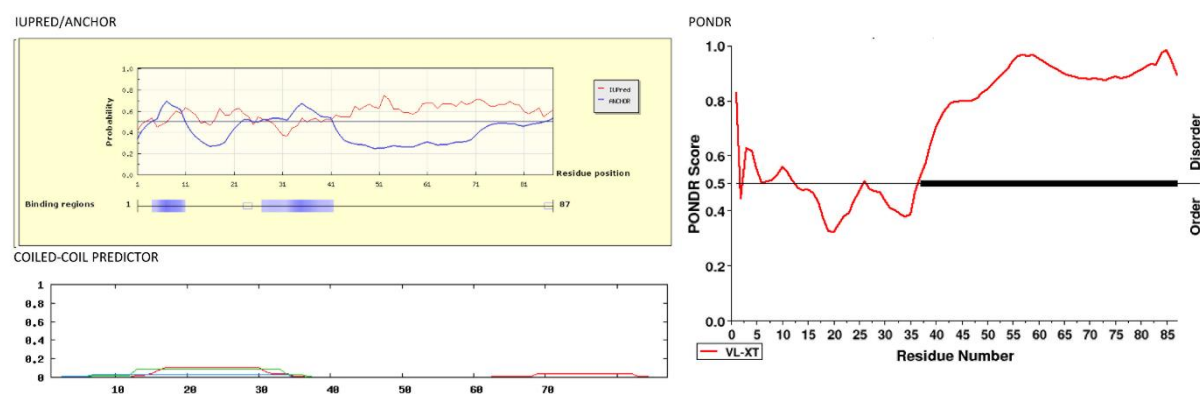
OBSERVATIONS

Predicted interaction region (ANCHOR) overlaps well with the predicted α -helices. The docking region is not predicted to be somewhat disordered (IUPred) to largely ordered (PONDR), nor to form a coiled-coil.

Agreement with VirA^{CDD}

TaiK^{NDD}

(M) RS**SKEIFEALR**DGRLS**REEAHAALRSARAAADAAL**GASGAFEPrgAEAAAGANGAADAGDAASGAS
 AANGERAHAAKTAESAAAADA



OBSERVATIONS

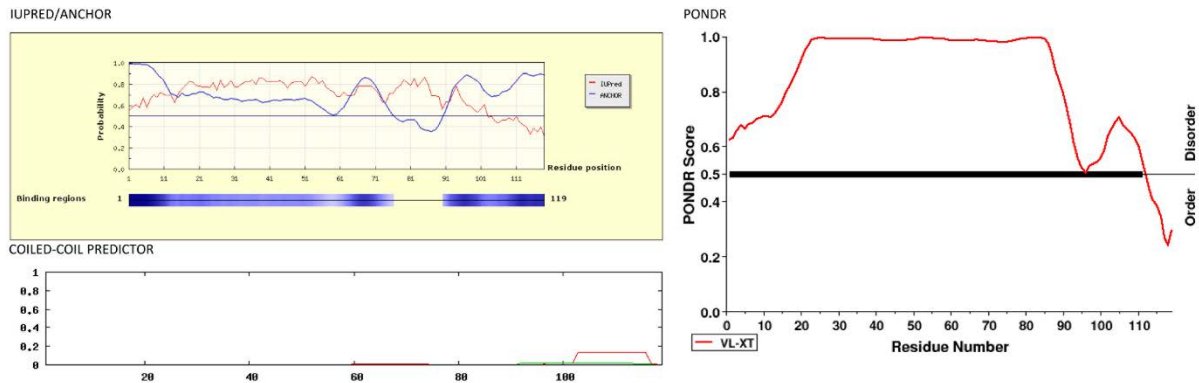
Predicted interaction regions (ANCHOR) overlap well with predicted α -helices (though predicted binding regions are weak). The docking region is predicted to be borderline disordered (IUPred/PONDR) and not to form a coiled-coil.

Agreement with VirFG^{NDD}

TaiL^CDD

ASELALARRRAGGAPLPDAEAASATTATTASASGTAHAPNAAGDAHAPNAAGARAATPAPAENAAGAV
HAARAANPANPATLANPAPGGGVPLDDV~~LARVHR~~GEL~~SV~~EAAEALLAGALG

93 102 105 117



OBSERVATIONS

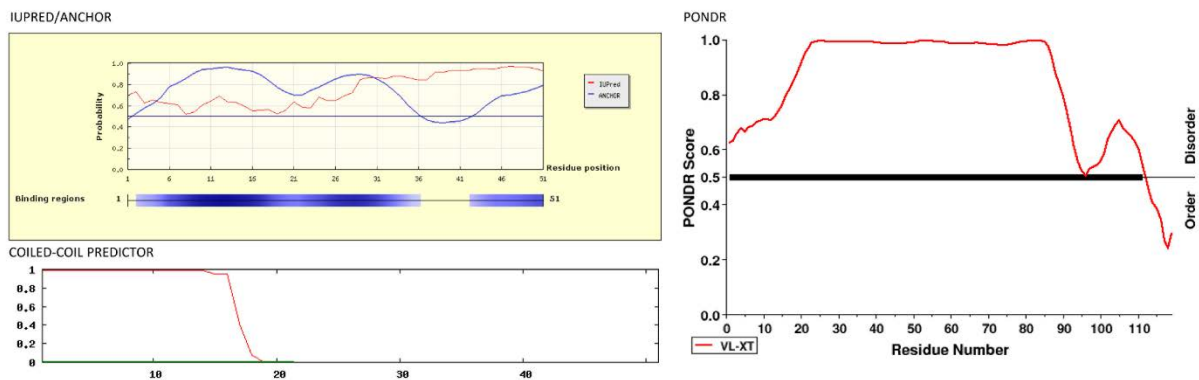
Predicted interaction region (ANCHOR) overlaps well with the predicted α -helices. The docking region is predicted to be mixed ordered/disordered (IUPred/POND), and not to form a coiled-coil.

Agreement with VirA^CDD

TaiM^NDD

3 16 20 32

MERQNL~~RNVLER~~LKAA~~AIGP~~DEAKRML~~SAARA~~GDAADKPQPGPAAAAPRDA



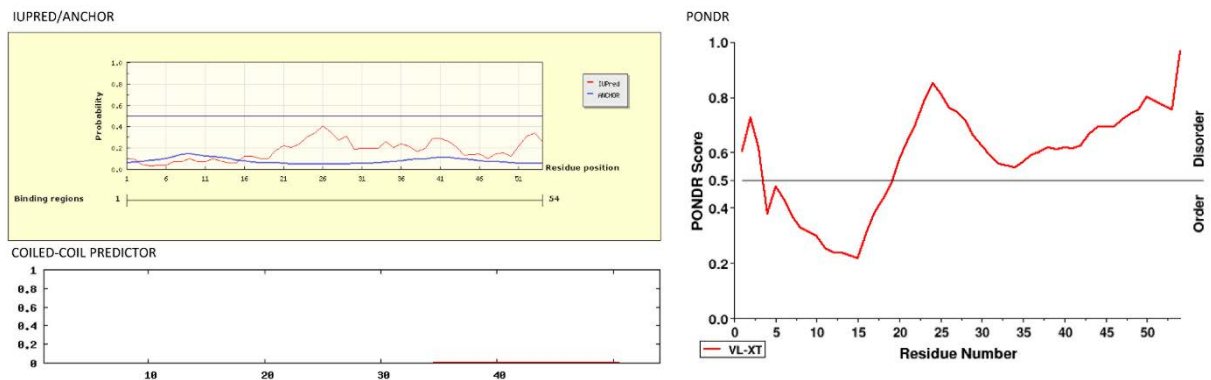
OBSERVATIONS

Predicted interaction regions (ANCHOR) overlap well with predicted α -helices (though predicted binding regions are weak). The docking region is predicted to be borderline (IUPred) to strongly disordered (POND), and not strongly predicted to form a coiled-coil.

Agreement with VirFG^NDD

5923 ^CDD

3 26 32 45 48 51
 ARGALAE^LFAPAIVAAPAQLANAAADASLGD^HSA^EELARIL^AHELGG^LES^RGAL

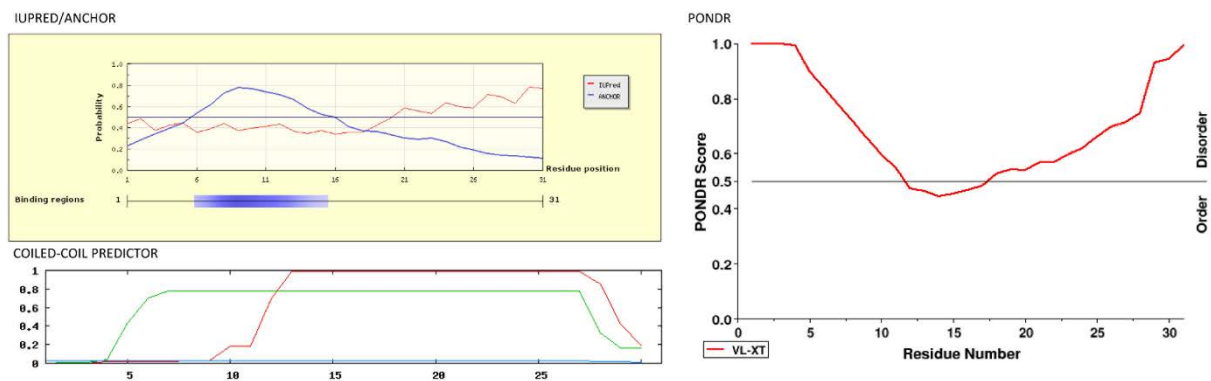


OBSERVATIONS

The predicted α -helices are not predicted to be interaction regions (ANCHOR). The docking region is predicted to be ordered (IUPred) to borderline disordered (PONDR), and not to form a coiled-coil. Poor agreement with VirA ^CDD

5922 ^NDD

4 28
 MTD^MDKDLLL^QSI^QTIRELK^TRLA^QAE^QGHH

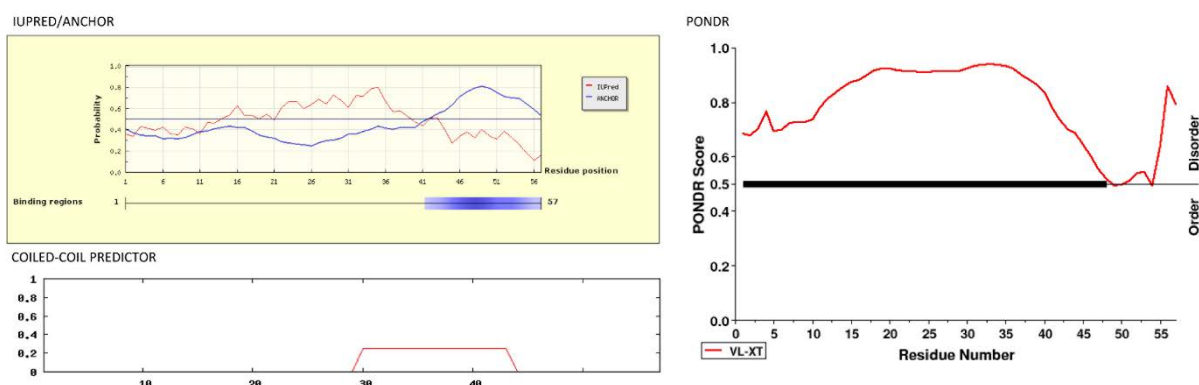


OBSERVATIONS

The predicted α -helix is predicted to be an interaction region. The docking region is predicted to be mixed ordered/disordered (IUPred/PONDR) but also predicted to form a coiled-coil. Does not agree with VirFG ^NDD; may be an alternative type of docking domain?

5925 ^CDD

LAANAPAGSAVAEAAASVAPAVEATEPAAAPLDTELSEIEGLQDDDLAALLGKEFIRE

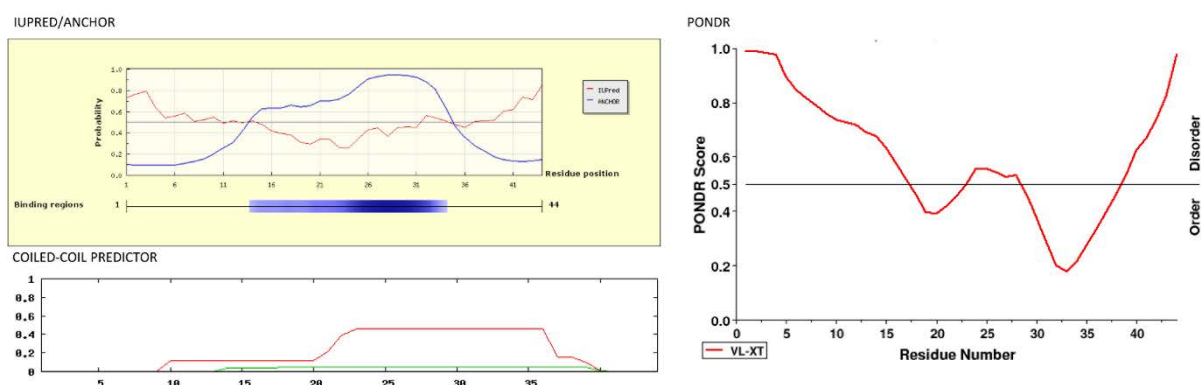


OBSERVATIONS

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 ^CDD

5924 ^NDD

MNKPTSSDGWKDDYLSRLSRLSKNQLMALALKLKQQQLEQGPAA



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 VirFG ^NDD

CONTROL: CLASS 1A DDs

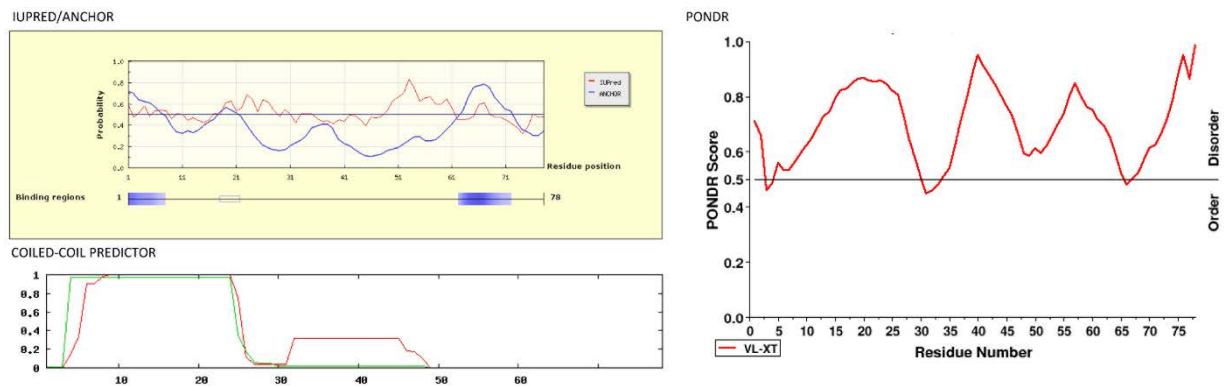
Key:

Red = α -helices observed in the NMR structure of the DEBS 2 C DD/DEBS 3 N DD complex

Blue = α -helices predicted by PsiPred

DEBS 2 C DD

8 20 28 42 61
 AASPAVD **IGDRLDELEKALE** **ALSA** EDG **HDDVGQRLESLLRRW** NSRRADAPSTSAISEDAS **DDELF** SML
 DQRFGGEDL
 70

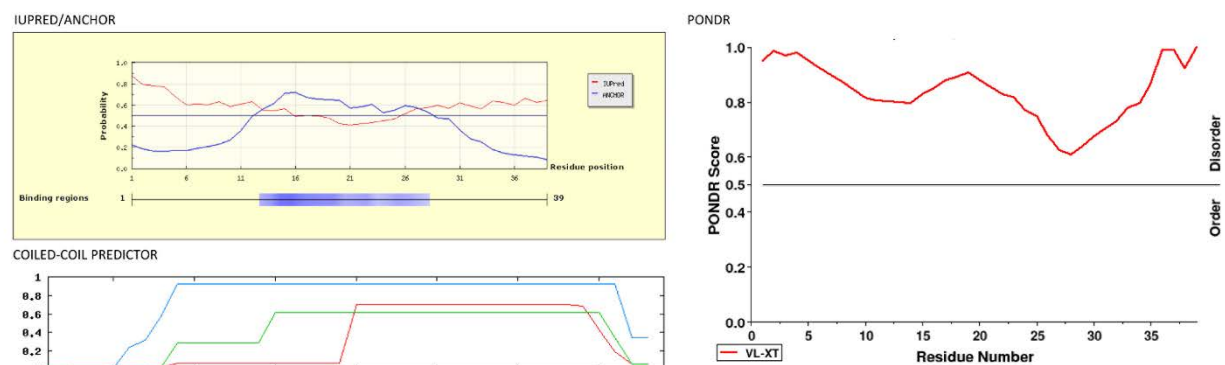


OBSERVATIONS

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 N DD

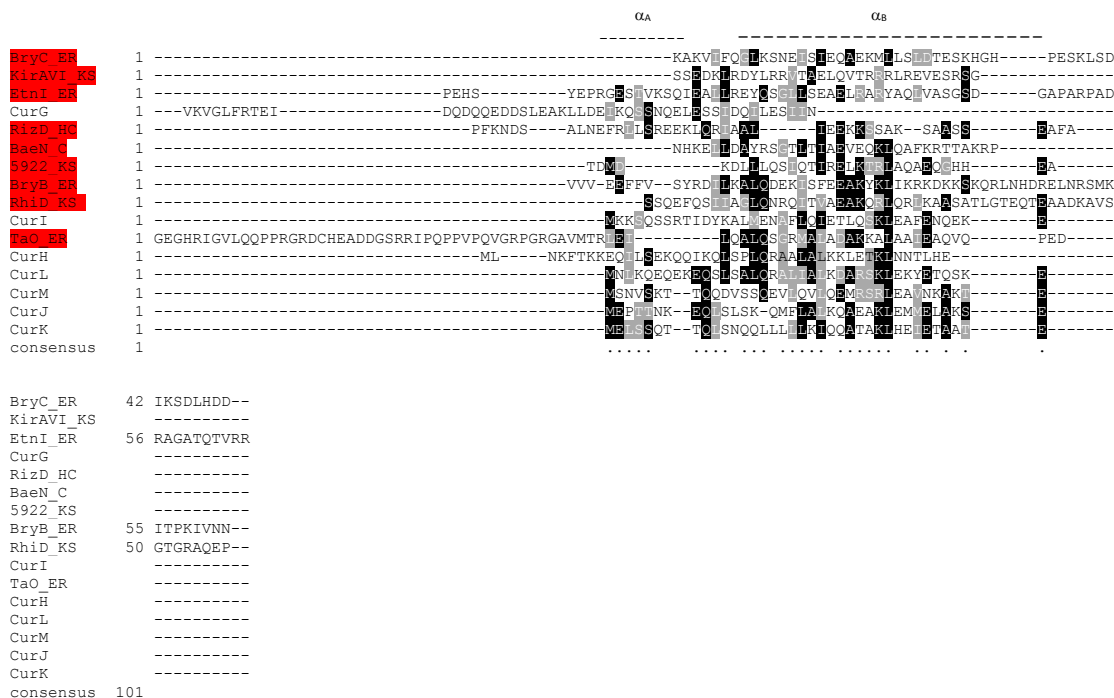
10 34
 MSGDNGMT **EKLRRYLKRTVTELD** SVTARL **REVEH** RAGE



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).

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 PKSs (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.

Caractérisation des interactions entre sous-unités au sein d'une PKS hybride, l'enacyloxine synthase

Risser, F., Collin, S., Dos Santos-Morais R., Gruez, A., Chagot, B. & Weissman, K. J.

Suite à la caractérisation de la nouvelle classe de DD présents entre les sous-unités VirA et VirFG de la virginiamycine synthase, une PKS *trans*-AT (Dorival et al., 2016), notre équipe s'est intéressée aux six interfaces de l'enacyloxine synthase qui est un système particulier : c'est en effet une PKS hybride *cis*-/*trans*-AT, les six premières sous-unités appartiennent à la classe des *cis*-AT et seule la dernière à la classe des *trans*-AT.

Dans un premier temps l'existence et les limites des DD putatifs, situés aux extrémités C_{ter} et N_{ter} des sous-unités du système, ont dû être confirmées. Pour ce faire, des alignements de séquences de ces régions ont été réalisés de manière à déterminer les limites des domaines conservés. Dans le cas des interfaces « classiques », situées entre des domaines ACP en C_{ter} et KS en N_{ter} de leur sous-unité (5925/5924, 5923/5922, 5922/5921, 5921/5920 et 5920/5919), des courtes séquences non assignées d'une cinquantaine et trentaine de résidus respectivement sont bien présentes. Seule l'extrémité C_{ter} de la sous-unité 5924 présente plus de 320 résidus en aval du dernier domaine conservé, une KS. L'interface formée entre cette sous-unité et la suivante est particulière puisqu'elle consiste en un module « splité ». Afin d'affiner ces premiers résultats, une série de logiciels a été utilisée sur les régions de *dockings* potentiels. Les régions N_{ter} et C_{ter} auraient la capacité à se structurer en hélice α (PSIPRED, Buchan et Jones, 2019). De plus les extrémités C_{ter} présenteraient un caractère désordonné, ayant la capacité à se structurer en présence d'un partenaire d'interaction (IUPred2A, Mészáros, Erdős, et Dosztányi, 2018), alors que les régions N_{ter} formeraient des *coiled-coil* assurant ainsi leur homodimérisation (prabi, https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_lupas.html).

Les caractéristiques de ces régions non annotées étant partagées avec d'autres classes de DD (Broadhurst et al., 2003 ; Whicher et al., 2013 ; Dorival et al., 2016), nous avons pu émettre l'hypothèse qu'elles correspondaient bien à des régions de *docking* entre sous-unités.

Ainsi, toutes les séquences en amont des domaines conservés en N_{ter} de leur sous-unité ont été clonées. Pour quatre ^cDD sur six, deux versions ont été clonées : l'analyse bio-informatique de ces régions prédit la présence d'une hélice α terminale avec un fort degré de confiance ainsi qu'une autre hélice α plus ou moins longue en amont. Une version courte comprend donc cette hélice terminale et une version longue comprend les deux hélices. Seuls

deux $^{\text{C}}\text{DD}$ sont clonés en une seule version courte : le 5925 $^{\text{C}}\text{DD}$ et le 5924 $^{\text{C}}\text{DD}$. Dans les deux cas, seule une petite hélice α , située à l'extrémité C_{ter} des sous-unités, est présente avec un degré de confiance important.

Au final, toutes les constructions sont obtenues à homogénéité à l'exception des DD du module « splité » et des deux versions du 5920 $^{\text{C}}\text{DD}$ que nous n'avons jamais réussi à purifier. Pour ces derniers (5924 $^{\text{C}}\text{DD}$, 5923 $^{\text{N}}\text{DD}$ et une version courte du 5920 $^{\text{C}}\text{DD}$), notre étude s'est faite avec des peptides synthétiques.

L'analyse biophysique des spectres CD révèle que tous les $^{\text{C}}\text{DD}$ comportent très peu de structurations secondaires alors que tous les $^{\text{N}}\text{DD}$ présentent des spectres caractéristiques de protéines structurées en hélice α . Par ITC, un K_d est déterminé pour tous les partenaires d'interaction, à l'exception des DD potentiels du module « splité ». Les affinités varient de $2,1 \pm 0,6 \mu\text{M}$ pour les partenaires de la première interface (5925/5924) à $57 \mu\text{M}$ dans le cas de l'interface 5922/5921, mais quelle que soit l'interface testée, la stœchiométrie est de 1:1. Ces expériences ITC permettent aussi de déterminer quelle version des $^{\text{C}}\text{DD}$, longue ou courte, est la plus pertinente. En effet, dans le cas de l'interface 5923/5922, aucune interaction n'est observée entre la version courte du $^{\text{C}}\text{DD}$ et son partenaire, alors que cela est possible avec la version longue : un K_d de $26 \pm 4 \mu\text{M}$ est calculé. Dans le cas de l'interface 5922/5921 l'affinité d'interaction est deux fois plus importante lorsque la version longue du $^{\text{C}}\text{DD}$ est testée ($57 \mu\text{M}$ contre $110 \mu\text{M}$). La version courte du 5921 $^{\text{C}}\text{DD}$ est utilisée par la suite puisqu'aucune différence significative de K_d n'est observée entre la version longue et courte de ce $^{\text{C}}\text{DD}$ avec le 5920 $^{\text{N}}\text{DD}$.

Pour caractériser plus finement les interactions mises en évidence par ITC, des spectres CD avant mélange des deux partenaires et après mélange, donc après formation des complexes, sont enregistrés. Pour chaque interface, à l'exception du module « splité », une augmentation des signaux spécifiques des hélices α est observée. Nous avons ainsi émis l'hypothèse que cette augmentation de structuration provient en majorité des $^{\text{C}}\text{DD}$ se repliant en hélice α lors de l'interaction avec leur partenaire. Les $^{\text{C}}\text{DD}$ de l'enacyloxine synthase seraient donc des régions intrinsèquement désordonnées (IDR) de la PKS. Alors que ce type d'interaction est déjà mis en évidence chez des PKS *trans*-AT (Dorival et al., 2016), c'est la première fois qu'elle est caractérisée au niveau de jonctions PKS *cis*-AT. C'est aussi la première fois que ce mécanisme d'interaction est démontré au niveau de plusieurs interfaces d'un même système.

Par une approche combinant ITC, CD, SLS et SAXS, l'état oligomérique des DD est étudié. Il s'avère que tous les ^CDD sont monomériques, alors que tous les ^NDD, à l'exception du 5921 ^NDD, sont homodimériques, et ce, probablement *via* la formation d'un *coiled-coil*. Un autre point commun entre toutes les interfaces réside dans le pI des partenaires : tous les ^CDD présentent un pI acide, compris entre 4 et 6 alors que la majorité des ^NDD présentent un pI basique, compris entre 9,7 et 10,6. Seuls le 5922 ^NDD et le 5920 ^NDD possèdent un pI de 5,4 et 6,4 respectivement. Toutes les interfaces caractérisées de l'enacyloxine synthase partagent donc de nombreux points communs : des ^CDD monomériques et chargés partiellement négativement, intrinsèquement désordonnés, interagissant avec des ^NDD structurés en hélices α , s'homodimérisant par l'intermédiaire d'un *coiled-coil* et pour la plupart chargés partiellement positivement. Des expériences de mésappariement menées par ITC montrent cependant que ce système PKS présente des interactions hautement spécifiques puisque le 5925 ^CDD et le 5921 ^CDD n'interagissent pas ou très faiblement, avec les autres ^NDD du système.

Dans le but de mieux comprendre d'où émerge la spécificité d'interaction entre chaque interface, des structures RMN de couples de DD en fusion ont été résolues. La première construction relie les DD de la première interface (5925/5924) par l'intermédiaire d'un *linker* (Gly₂Ser)₃. La structure du complexe présente de nombreuses caractéristiques de DD de classe II (Whicher et al., 2013) : les ^NDD homodimériques sont composés de deux hélices α chacun, les deuxièmes forment un *coiled-coil*. Les ^CDD monomériques sont aussi formés de deux hélices α , et interagissent avec un monomère de ^NDD chacun. Globalement la structure obtenue est très semblable à celle des DD de classe II (Whicher et al., 2013) puisque les monomères adoptent la même structuration. L'angle entre les deux hélices α d'un monomère de ^CDD est le même que dans la structure des ^CDD de cette classe, de même pour l'angle entre les deux hélices α d'un monomère de ^NDD. Toutefois la topologie du complexe diffère légèrement par rapport à la structure cristallographique (Whicher et al., 2013), notamment dans la position des hélices α des ^NDD formant le *coiled-coil*.

La dernière interface, permettant la communication entre une sous-unité *cis*-AT et la dernière sous-unité *trans*-AT, appartient quant à elle à la classe Ib de DD (Buchholz et al., 2009). Ces derniers ont été fusionnés par l'intermédiaire d'un *linker* (Gly₃Ser)₂ afin d'obtenir la structure du complexe par RMN. Ici, le *docking* est réalisé par un ^CDD formant une petite hélice α s'insérant dans le sillon formé par le *coiled-coil* de ^NDD. Une différence majeure est à noter avec les DD de classe I : le ^CDD ne présente pas le couple d'hélices α en amont de

l'interface de *docking* qui s'homodimérise par l'intermédiaire d'un fagot de quatre hélices (Broadhurst et al., 2003).

Les alignements de séquences suggèrent que les autres interfaces de l'enacyloxine synthase appartiendraient soit à la classe II pour les interfaces 5925/5924 et 5923/522 soit à la classe I pour les interfaces restantes : 5924/5923, 5922/5921, 5921/5920 et 5920/5919. Le mésappariement des DD appartenant à la même classe serait principalement empêché par un code électrostatique différent. Par exemple les pI des DD de l'interface 5925/5924 sont de 3,8 et 9,7 respectivement alors que ceux de l'interface 5923/5922 sont de 5,0 et 5,4 respectivement. En outre, il semble que les domaines adjacents aux DD pourraient jouer un rôle dans le cas du module « splité » puisqu'aucune interaction directe entre les DD potentiels n'est observée. Des constructions plus longues ont donc été clonées : KS-^CDD 5924, ^NDD-KR 5923 et ^NDD-KR-ACP 5923. Malheureusement ces protéines sont peu solubles et n'ont pas pu être purifiées.

Dans ce travail collaboratif au sein de l'équipe, j'ai cloné, produit et purifié les différentes constructions avec l'aide de Sabrina Collin. J'ai mené les expériences de CD et d'ITC et Raphaël Dos Santos-Moirais les expériences SLS. Avec l'aide du docteur Arnaud Gruez, j'ai enregistré et traité les données SAXS obtenues au synchrotron SOLEIL. Les enregistrements des spectres et la résolution des structures RMN ont été réalisés avec l'aide du docteur Benjamin Chagot.

Understanding Intersubunit Interactions in the Enacyloxin Mixed *cis*-/*trans*-acyltransferase Modular Polyketide Synthase

Fanny Risser,[†] Sabrina Collin, Raphael Dos Santos-Morais,[‡] Arnaud Gruez,^{*} Benjamin Chagot,^{*} and Kira J. Weissman^{*}

UMR 7365, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA), CNRS-Université de Lorraine, Biopôle de l'Université de Lorraine, Campus Biologie Santé, 9 Avenue de la Forêt de Haye, BP 20199, 54505 Vandœuvre-lès-Nancy, France
Supporting Information

ABSTRACT: Specific interactions between consecutive polypeptide subunits in modular polyketide synthases (PKSs) are critical for faithful biosynthesis in bacteria of a wealth of bioactive natural products. These contacts are mediated, at least in part, by matched pairs of C- and N-terminal ‘docking domains’ (DDs). Here, we have characterized the DDs operating at the six inter-PKS interfaces within the unusual mixed *cis*-/*trans*-acyltransferase (AT) PKS responsible for assembling the promising antibiotic enacyloxin. Taken together, our data reveal that specificity is achieved in the enacyloxin PKS by deploying multiple functionally orthogonal classes of DDs, but also by modulating the electrostatic drivers of association within single classes. Our results also reinforce the utility of intrinsically disordered regions (IDRs) for achieving both medium affinity and high specificity at PKS intersubunit junctions. Overall, this work substantially increases the number of orthogonal DDs available for engineering novel, highly-specific interfaces within *cis*-AT and *trans*-AT PKSs and their hybrids, and defines a clear set of design rules for their usage.

■ INTRODUCTION

Modular polyketide synthases (PKSs) are bacterial mega-enzymes responsible for synthesis of complex polyketide natural products, many of which have found clinical application.¹ A striking feature of these systems is their assembly-line (division-of-labor) organization, as each cycle of chain building is carried out by a specific module of enzymatic domains (**Figure 1**).^{2,3} To date, two classes of modular PKSs have been identified: the ‘*cis*-acyl transferase (AT)’ and ‘*trans*-AT’ systems, which appear to have distinct evolutionary origins.^{4–6} The defining difference between the two families is that in *trans*-AT PKSs, the eponymous ATs are not present within the multienzymes, but are discrete proteins which act iteratively in *trans*.⁷ *Trans*-AT PKSs also exhibit a broader array of module architectures, and the majority of systems contain modules (nonribosomal peptide synthetases (NRPSs)) which incorporate amino acid monomers into the chains, although these are also present in many *cis*-AT PKSs.⁸ An additional common feature is that following release of the chain extension intermediates from the last multienzyme, they are elaborated to their final, bioactive forms by so-called ‘post-assembly line’ enzymes (**Figure 1**).^{9,10}

In both *cis*-AT and *trans*-AT PKSs, the multiple modules that constitute the assembly lines are distributed among several gigantic polypeptides called ‘subunits’ (**Figure 1**). As a result, the fidelity of the biosynthesis depends on specific interactions between the various subunits which allow them to adopt the correct sequence. Conversely, rational modification of interprotein communication by genetic engineering to enforce unnatural subunit ordering is a promising strategy towards generating

novel, high-value polyketide analogues, as subunits/modules are employed as intact units.¹¹ This approach has recently gained traction with the discovery that the functional domains within PKS modules are highly interdependent.^{12,13}

For both *cis*-AT and *trans*-AT PKSs, communication between subunits is mediated, at least in part, by short sequences at the extreme ends of the polypeptides called ‘docking domains (DDs)’ (where ^CDD refers to a C-terminal docking domain, and ^NDD to an N-terminal DD).¹⁴ Multiple structural classes of DDs have been identified to date, typically by solving the NMR or crystal structures of artificially-fused native ^CDD–^NDD pairs in order to stably capture the interactions. Collectively, these efforts have revealed four distinct types of DDs within *cis*-AT PKSs and their hybrids with NRPSs (1a (ref. ¹⁴)/1b (ref. ¹⁵), 2 (ref. ¹⁶) and 3 (ref. ¹⁷) (**Figure 2A**)). For classes 1a/1b and 2, formation of the docking domain complexes relies on contacts between α -helices, while for class 3, the ^CDD, which is a short linear motif (SLiM) within the overall disordered C-terminus of the subunit,¹⁸ inserts as a β -strand into a β -hairpin formed by the N-terminal domain, with additional contacts to an N-terminal α -helix¹⁹ (this class of DDs has thus recently been renamed ‘SLiM– β -hairpin docking (BHD) domains’¹⁸). There are, in addition, fundamental differences in oligomerization state of DDs, as the ^NDDs are largely homodimeric, and the interaction portions of the ^CDDs uniformly monomeric (the C-termini of type 1a/1b docking domains include an additional four α -helix bundle dimerization motif upstream of the docking α -helix, but it is not involved in docking (**Figure 2A**)).^{14,20}

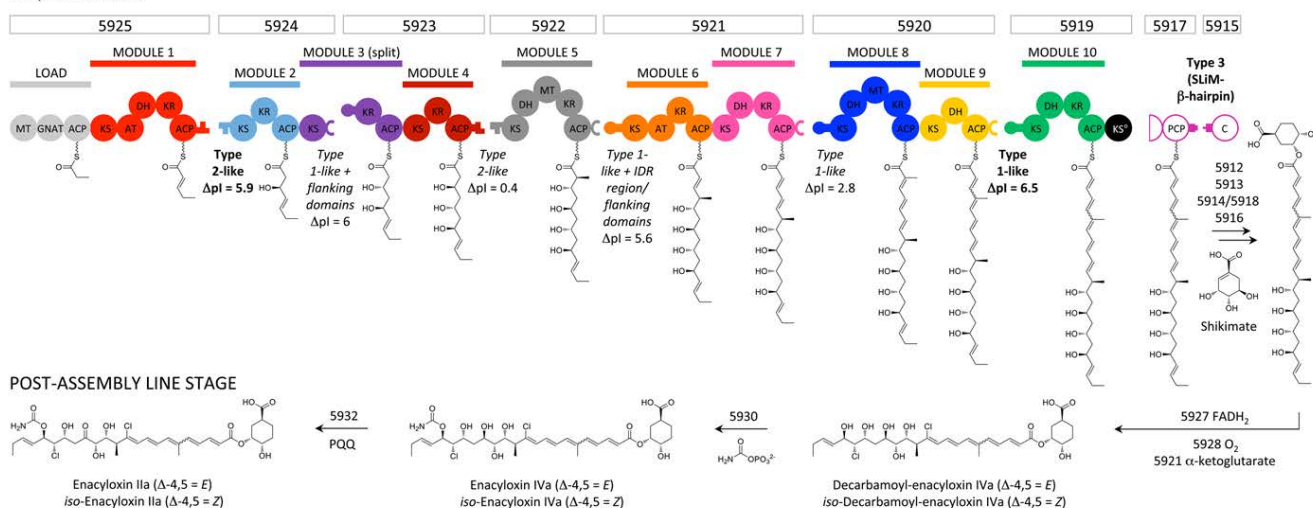


Figure 1. The enacyloxin biosynthetic pathway. The biosynthesis occurs in two stages: assembly of the core by a hybrid polyketide synthase (PKS) (solid spheres)/nonribosomal peptide synthetase (NRPS) (hollow spheres), followed by elaboration of the released chain by a series of post-assembly line enzymes. The PKS portion of the pathway encompasses seven large subunits (Bamb_5925–5919, boxed), the first six of which belong to the *cis*-AT PKS class, and the last of which is a *trans*-AT PKS. Each subunit contains one or more functional modules (one loading and 10 chain extension), which are themselves composed of catalytic and acyl carrier protein (ACP) domains. As determined in this work, communication across the PKS intersubunit interfaces is mediated by a collection of docking domains belonging to two different types: 1-like and 2-like (DD types were assigned based on structure elucidation at high resolution (those in bold), or detailed biophysical and bioinformatics analysis), while a type 3 DD (or SLiM-β-hairpin¹⁸) was recently shown to operate at the PCP (Bamb_5917)/C (Bamb_5915) interface involved in chain off-loading. Key: MT, C-methyl transferase; GNAT, Gcn5-related *N*-acetyl transferase; KS, ketosynthase (the ° indicates a condensation-inactive version which serves as an ACP transacylase); AT, acyl (malonyl) transferase; DH, dehydratase; KR, ketoreductase; PCP, peptidyl carrier protein; C, condensation.

More recently, *trans*-AT PKSs were also demonstrated to incorporate DDs, with the two classes described to date (**Figure 2B**) differing from those in *cis*-AT PKSs. In the first published example of the 4 α-helix bundle type of DDs,^{21,22} monomeric ^CDD and ^NDD which both encompass two consecutive α-helices, were shown to interact to form a four α-helix bundle with a unique topology for a protein-protein interaction.²¹ Another notable feature of this complex is that the ^NDD exhibits many convincing characteristics of an intrinsically disordered region (IDR),²³ and was in fact the first IDR to be shown to operate in bacterial secondary metabolism. Subsequently, this type of docking interaction was identified within a much more extensive range of interfaces, including those between domains which belong to the same module, but which are located on adjacent polypeptides (so-called ‘split modules’, which are common in *trans*-AT PKSs).²² The second class of DD, dehydratase docking (DHD),²⁴ is located C-terminal to KS domains. As with 4 α-helix bundle ^CDDs, DHDs are IDRs, but they do not associate with another docking domain, but with a downstream DH domain situated on a separate multienzyme – the first example of a DD interacting directly with a catalytic domain to ensure productive intersubunit contacts.²⁴

In terms of the strength of docking, measured affinity between matched DD pairs in both *cis*- and *trans*-AT PKSs is remarkably consistent and in the low/medium μM range (2–24 μM),^{15,16,21,25} arguing that docking interactions between subunits are transient.^{26,27} Specificity appears to arise from strategically-positioned polar/charge:charge interactions at the DD interfaces,^{16,17,19} but for many systems, it has not yet been possible to elucidate a clear amino acid ‘code’ correlating with partner

choice and which could be targeted for genetic engineering to rationally redirect communication.

An attractive target for synthetic biology is the PKS responsible for biosynthesis of enacyloxin in several strains of *Burkholderia* (**Figure 1**).²⁸ Enacyloxin is a polyene antibiotic which exhibits useful activity against both Gram-positive and Gram-negative bacteria. Susceptible strains include multi-drug resistant *Acinetobacter baumannii*, among the 12 families of bacteria recently identified by the World Health Organization (WHO) as posing the greatest threat to human health.²⁹ However, enacyloxin in its native form is unsuitable for clinical application due notably to its chemical instability,²⁸ and therefore it is of substantial interest to generate structural analogues by genetic engineering.

Towards this end, we aimed to elucidate the molecular basis for docking within the enacyloxin PKS in order to enable subunit/module swaps. In addition, the pathway comprises several unusual features which make it an attractive model system for understanding protein-protein interactions in modular PKSs (**Figure 1**). Firstly, it includes seven PKS subunits, a relatively large number compared to other assembly lines.^{7,30} In addition, DDs have also recently been shown to mediate enacyloxin’s unusual chain off-loading mechanism, which involves the last PKS Bamb_5919, and two discrete NRPS domains (a peptidyl carrier protein (PCP, Bamb_5917) and a condensation domain (C, Bamb_5915)) (**Figure 1**).^{18,31} Given the resulting large number of interprotein interfaces in the assembly line (five of which involve flanking acyl carrier protein (ACP) and ketosynthase (KS) domains), studying the pathway was anticipated to illuminate how strict specificity is achieved in such situations.

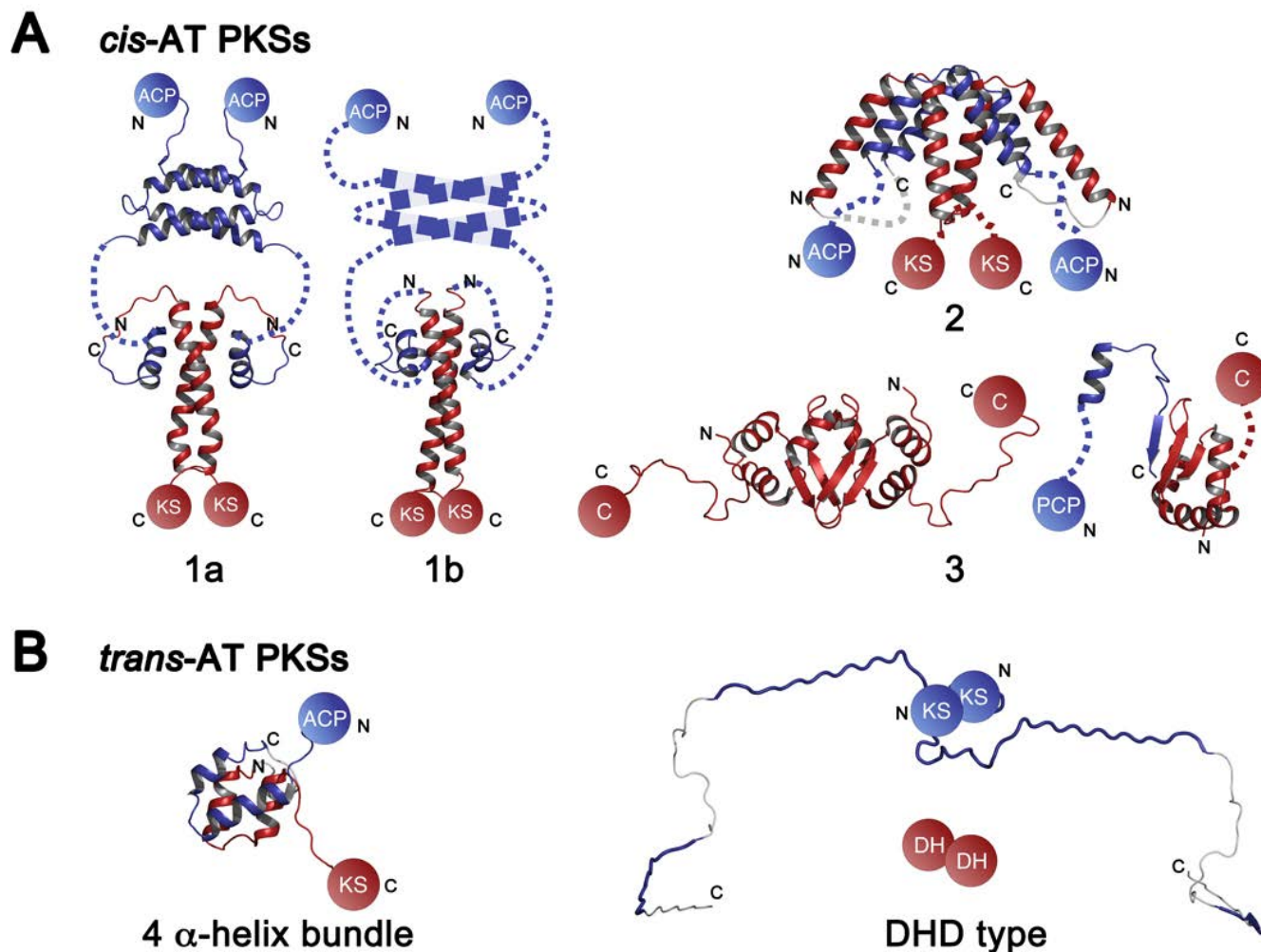


Figure 2. Docking domains (DD) in modular PKSs. (A) The four structural classes of DDs characterized to date from *cis*-AT PKSs,^{14–16} are shown in the presence of typical up- and downstream flanking domains. The C-terminal DDs (^CDD) are illustrated in blue, and the ^NDDs in red. Class 3 is notably present in hybrid PKS/NRPS systems (where the ^NDD is homodimeric),¹⁷ as well as pure NRPSs (where the ^NDD is monomeric).¹⁹ (B) Classes of DDs identified to date in *trans*-AT PKS systems. The 4 α-helix bundle type is formed by two α-helices contributed by each of the ^CDD and ^NDD. The DHD domains (for ‘dehydratase (DH) docking’) (shown in thick blue lines) which are present at KS/DH junctions, are intrinsically disordered regions (IDRs) which directly contact the downstream dehydratase domain. The ^NDD of the four α-helix bundle type of docking domain (shown in red) is also an IDR. Key: ACP, acyl carrier protein; KS, ketosynthase; PCP, peptidyl carrier protein; C, condensation.

Secondly, while the first six multienzymes (Bamb_5925–5920) are of the *cis*-AT type, the last (Bamb_5919) is a *trans*-AT PKS (in the interest of brevity, the subunits will be referred to hereafter by their numbers).²⁸ Determining how communication occurs across this atypical hybrid interface was thus expected to be useful for attempts to create artificial chimeric *cis*-AT/*trans*-AT PKS junctions via synthetic biology.

Using a multidisciplinary approach combining NMR, small angle X-ray scattering (SAXS), biophysical analysis and site-directed mutagenesis, we have comprehensively characterized docking within the enacyloxin PKS. Taken together, our results reveal similarities to previously-described DDs, but also several features which are unique to this system. They also identify multiple additional orthogonal DDs pairs and guidelines for their usage, further enriching the ‘toolbox’ of such elements available for genetic engineering.

RESULTS

In Silico Analysis of Docking in the Enacyloxin PKS. The starting point for this work was to identify putative docking domains in the enacyloxin PKS using bioinformatics, although such *in silico* analysis is insufficient for conclusive assignment either of DD extremities or structural classes, particularly in the case of *trans*-AT PKSs.²¹ For this, we analyzed the sequences of the C- and N-termini of all PKS subunits within the system. In the case of subunits 5925 and 5923–5920, the C-terminal domain is an ACP, while for subunits 5924 and 5922–5919, the N-terminal domain is a KS. The exceptional split module interface in the system (between subunits 5924 and 5923) is flanked by KS and KR domains. Thus, in order to identify candidate docking domains, we compared the terminal domain of each subunit to the sequences of appropriate homologous domains whose structures have been solved at high resolution, in order

Table 1. Summary of findings on the enacyloxin docking domains.

Subunit terminus ^a	Length (aa)	pI	Coiled-coil predicted ^c	IDR Predicted ^d / evidence obtained	Oligomerization state	Docking affinity (K _d)	DD class assignment or prediction
5925 ^C DD short ^b	29	3.8		N/Y	Monomer	1.1 ± 0.1 μM	2-like
5924 ^N DD	43	9.7	Y (2 nd half)	Y (weak)/N	Homodimer		
5924 ^C DD	22	4.6		N/Y	Monomer	No interaction	<i>Truncated* 1-like</i> , and junction likely involves up/downstream domains
5923 ^N DD	34	10.6	Y (v. weak)	N/N	Homodimer		
5923 ^C DD long ^b	54	5.0		N/Y	Monomer	26 μM	2-like
5922 ^N DD	32	5.4	Y (2 nd half)	Y/N	Homodimer		
5922 ^C DD long	80	4.1		Y/Y	Monomer	57 μM (long ^C DD); 110 μM (short ^C DD)	1-like (interaction helix, and may involve additional DD IDR region, flanking domains)
5922 ^C DD short	21	3.9		Y/Y	Monomer		
5921 ^N DD	31	9.7	Y (weak)	N/N	Monomer		
5921 ^C DD short	20	3.6		Y/Y	Monomer	8.5 ± 0.5 μM	<i>Truncated 1-like</i>
5920 ^N DD short	31	6.4	Y	N/N	Homodimer		
5920 ^C DD short	20	3.7		Y/Y	Monomer	43 ± 3 μM	Truncated 1-like
5919 ^N DD	29	10.2	Y	N/N	Homodimer		

^aMatched pairs of docking domains are indicated in the same color.

^bThe designation ‘long’ indicates that all sequence downstream of the last C-terminal domain (^CDDs) or upstream of the first N-terminal domain (for ^NDDs) was investigated; in the case of ^NDDs, the first M has been excluded; the designation ‘short’ refers to a version of the region that was investigated, whose sequence was informed by secondary structure analysis.

^cPredicted using: https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_lupas.html; Y if >50% probability over the majority of the sequence. As only ^NDDs have been shown to form coiled-coils, this analysis was not carried out with the ^CDDs.

^dPredicted using: <https://iupred2a.elte.hu/> (note: monomers of coiled-coil can predict to be IDRs); Y if >50% probability over the majority of the sequence.

^e‘Truncated’ refers to the fact that these ^CDDs lack the two α-helix dimerization motif found in classical type 1 DDs.^{14,15}

to determine the sequence regions in which DDs might be located (**Figure S1**). This was straightforward for all subunits ending in ACP domains (5925, 5923–5920). On the other hand, close analysis of the C-terminus of 5924 revealed a large unassigned region downstream of the conserved KS domain (**Figure S1**). To ascertain what portion of the extreme C-terminus of this sequence might contain a docking domain, we subjected it to secondary structure prediction (**Figure S1**). This analysis revealed a series of putative secondary structure elements, followed by a 40+ amino acid gap before the last strongly-predicted α-helix. We therefore focused on this extreme C-terminal region as potentially containing a ^CDD.

A series of predictive software was then used to analyze all of the putative DD regions for: i) secondary structure (PSIPRED³²); ii) propensity for disorder/acting as an interaction motif (IUPred2A³³); and in the case of the N-terminal DDs (because such elements have only been described for ^NDDs), iii) coiled-coil regions (coiled-coil prediction, NPS@ web server)³⁴ (**Figure S2**). We also determined the pIs of the putative DDs³⁵ (**Table 1**). Based on this analysis, we decided to express two versions of all but the 5924 ^CDD (full-length and a shorter version based on the secondary structure analysis) (**Tables S1** and

S2). In the case of the ^NDDs, we uniformly expressed all of the sequences upstream of the first conserved catalytic domains. However, for 5920 ^NDD, we additionally expressed a longer version, as a new enacyloxin PKS sequence annotation was released during the course of this work, which suggested an alternative start site (**Tables S1** and **S2**).

Expression of the Putative Enacyloxin DDs. Initially, eighteen constructs (5925 ^CDD long and short versions, 5924 ^NDD, 5924 ^CDD, 5923 ^NDD, 5923 ^CDD long and short versions, 5922 ^NDD, 5922 ^CDD long and short versions, 5921 ^NDD, 5921 ^CDD long and short versions, 5920 ^NDD long and short versions, 5920 ^CDD long and short versions, and 5919 ^NDD) were amplified from *B. ambifaria* genomic DNA and cloned into vector pBG102 or pLM302 for expression in *E. coli* BL21 (DE3) as His₆-SUMO-tagged or His₆-MBP tagged proteins, respectively. Based upon initial results, we then cloned ten additional constructs: three covalent fusions of 5925 ^CDD and 5924 ^NDD containing a (G₃S)₂, (G₃S)₃ or a zero-length linker; a fusion of 5920 ^CDD and 5919 ^NDD via a (G₃S)₂ linker; 5924 KS-^CDD; 5923 ^NDD-KR; 5923 ^NDD-KR-ACP; 5922 ACP-^CDD, 5921 ^NDD-KS and 5920 ACP-^CDD. (The sequences of all constructs are given in **Table S2**.)

To allow both efficient monitoring of the DDs by UV during purification and dosing for binding studies by ITC, we added a tyrosine residue at the N-terminal ends of certain C DDs and N DDs (**Table S2**). Following purification by nickel chromatography, and cleavage of the affinity/solubility tags using human rhinovirus 3C (H3C) protease (**Figure S3**), the N-termini of all of the proteins also harbored short, non-native sequences (GPGS for all but 4 constructs, **Table S2**). The GPGS sequence is highly similar and occupies the same position relative to the N DDs as the $(G_3S)_n$ sequence used to fuse certain N DDs covalently to their upstream C DDs, which we also show by NMR (*vide infra*), to have no effect on the docking interaction. In the case of the C DDs, the added residues lie upstream of the additional Y (**Table S2**). Unfortunately, despite extensive effort, we were unable to obtain the following constructs as recombinant proteins from *E. coli*: 5924 C DD, 5924 KS- C DD, 5923 N DD, 5923 N DD-KR-ACP, 5921 N DD-KS and 5920 C DD, so we sourced 5924 C DD, 5920 C DD short and 5923 N DD as synthetic peptides. In the following sections, we detail sequential analysis of each of the DD interfaces in the enacyloxin PKS.

Docking at the 5925 C DD/5924 N DD interface. Analysis of the 53-residue C-terminus of 5925 suggested the presence of up to three α -helices, while the 43-amino acid 5924 N DD was predicted to contain one long α -helix, with some propensity to form a coiled-coil within the second half of the sequence (**Figure S2**). While a construct encompassing the full 5925 C-terminus aggregated, a shorter version containing just the C-terminal 29 residues (5925 C DD short) proved well-behaved, and thus we judged it to contain the C DD (this construct is referred to hereafter as 5925 C DD). To investigate the interface, we initially subjected the recombinant DDs (5925 C DD and 5924 N DD (**Table S2**)) to circular dichroism analysis (**Figure S4**). This revealed that while the C DD was unstructured on its own, the N DD was well-folded into α -helix. We next analyzed both DDs by small-angle X-ray scattering (SAXS) and size exclusion chromatography-static light scattering (SEC-SLS), in order to determine their molecular weights in solution and therefore their oligomeric states. While 5924 N DD was found to be homodimeric, 5925 C DD is monomeric and highly extended with an R_g of 15.6 Å and a D_{max} of 62 Å, which is consistent with a disordered protein (as shown by a Kratky plot³⁶ (**Figure S4** and **Table S3**)). Acquisition of CD spectra of 5924 N DD at multiple concentrations (100, 10 and 1 μ M) did not reveal any substantial loss in secondary structure (**Figure S4**), suggesting the high stability of the homodimer, a result further confirmed by isothermal titration calorimetry (ITC), as little change was observed when the DD was titrated on its own (i.e. the monomers remained stably associated) (**Figure S4**). Analysis by ITC of 5925 C DD alone also yielded virtually no signal (**Figure S4**), but in this case, this behavior is consistent with its monomeric nature.

As we had previously observed induced structuration of an N DD in the presence of its partner,²¹ we next determined whether overall structuration of the DDs increased when they were combined. For this, we carried out differential CD analysis, acquiring spectra of 5925 C DD and 5924 N DD alone and together (**Figure S4**), as described previously.²¹ This analysis showed that the spectrum of the putative complex does not correspond to a simple average of the spectra of the two individual partners (α -helical content increased from 35.5% to 46.2%).³⁷ As 5924 N DD is already partially folded at the concentration

used in this experiment (60% α -helix), we thus attribute the structuration observed when the domains are mixed to induced folding of 5925 C DD (although the secondary structure content of 5924 N DD may also increase). Taken together with the SAXS and ITC results, these data are consistent with 5925 C DD being an intrinsically disordered region (IDR).³⁸

We next measured the affinity of the interaction by ITC (**Figure S4**). For this, we determined the K_d for binding when either 5925 C DD or 5924 N DD was placed into the cell and titrated with partner. The measured K_d s were similar and in the low μ M range ($1.1 \pm 0.1 \mu$ M for N DD in the syringe and C DD in the cell, and $3 \pm 1 \mu$ M for C DD in the syringe and N DD in the cell (experiments repeated in triplicate)) consistent with previous measurements of DD affinities,^{15,16,21,25} but the binding stoichiometry was different depending on which partner was in the cell: $N = 1.2$ for 5925 C DD in the cell and 0.6 for 5924 N DD in the cell. (A stoichiometry of 0.5 was also obtained by SEC-SLC when the 5925 C DD was in excess (**Figure S4**)). As it has previously been observed that titrations of IDRs into their partners can yield anomalous results,³⁹ we presumed that the stoichiometry of approximately 1:1 is correct, a result supported by subsequent NMR analysis (*vide infra*).

To further investigate the docking interaction, we analyzed 15 N-labeled 5925 C DD in the absence/presence of 14 N 5924 N DD and 15 N 5924 N DD in the absence/presence of 14 N 5925 C DD by NMR, so that in each case all signals arising could be attributed to the 15 N-labeled partner (**Figure S4**). These experiments provided clear evidence for stable complex formation between the discrete DDs, as well as confirming the induced structuration of the 5925 C DD seen by CD, and therefore encouraged us to elucidate the structures of both the well-folded 5924 N DD and a covalent complex of the two DDs by multidimensional heteronuclear NMR spectroscopy. For this, structure calculation was performed using a two-step procedure consisting of initial structure generation using CYANA,⁴⁰ followed by restrained molecular dynamics refinement within Amber.^{41,42} The quality of the structures was assessed using PROCHECK-NMR⁴³ and the MolProbity⁴⁴ server (**Table S4**).

The solved structure of 5924 N DD (**Figure 3A**) closely resembles that of type 2 N DDs¹⁶. Specifically, the N DD encompasses two consecutive α -helices, α A (W10–L18) and α B (K23–E39). α B from the two monomers forms a parallel coiled-coil, while each copy of α A contacts the α B of the opposite monomer, to form an overall 4 α -helix bundle. The first 13 residues of the construct (including four non-native residues) are unstructured, as predicted (**Figure S3**). Dimer formation is ensured by predominantly hydrophobic interactions between the second helix of one monomer and both helices of the other monomer. Hydrophobic residues involved in the packing are W10, Y14, L15 and L18 located in α A, L26, L29 and L33 from α B of the same monomer, and L26, M27, A30, L31, K34 and L38 within α B of the other monomer (**Figure 3B**). Residue K23 located at the beginning of α B of each monomer interacts with the carbonyl oxygen of S19 and L21 located within the loop between the helices of the other monomer. K11 of one monomer lies in the vicinity of Q35 of the other monomer, and may participate in a polar interaction. Overall, the structure of 5924 N DD resembles that of the prototypical class 2 N DDs from the curacin PKS¹⁶ (e.g. an RMSD of 0.55 Å between W10–Q36 of 5924 N DD and Q48–T74 of CurH (PDB code 4MYV)), but the

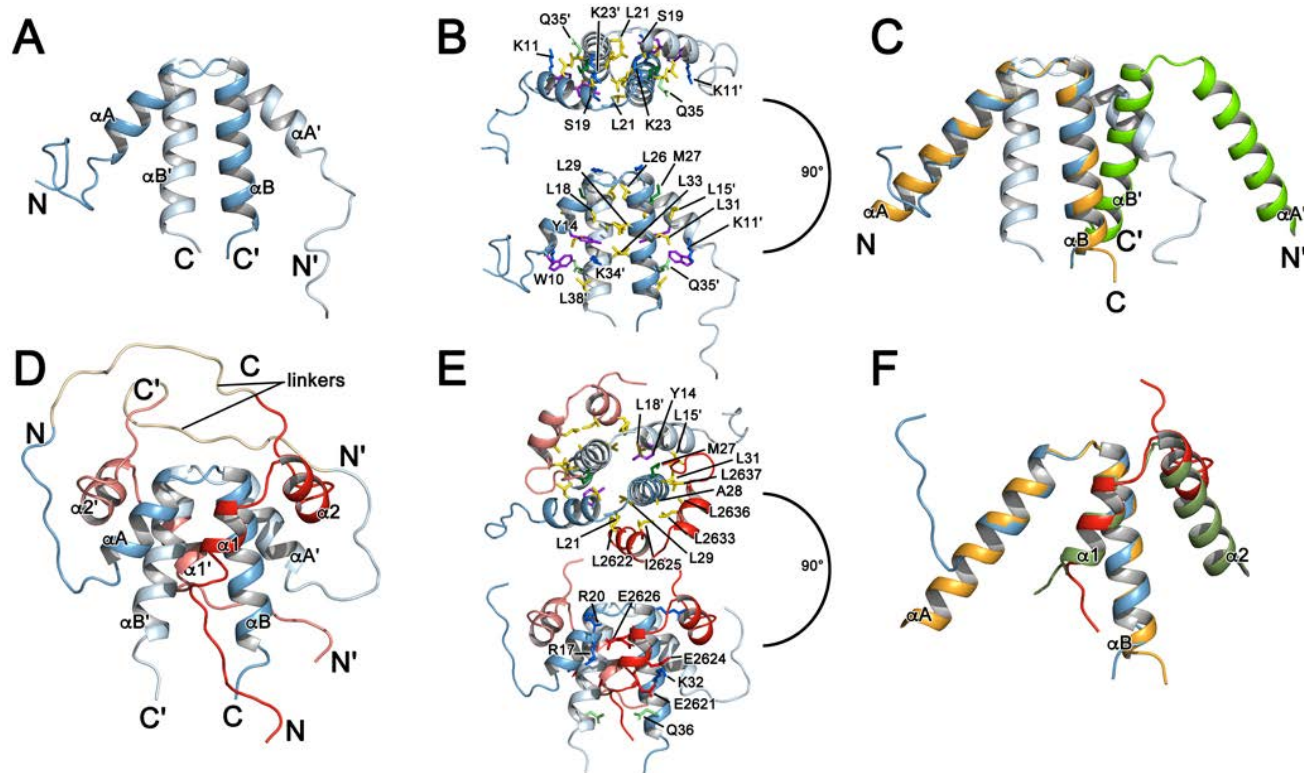


Figure 3. NMR structures of 5924 ^NDD and the 5925 ^CDD–5924 ^NDD complex. (A) Cartoon representation of the 5924 ^NDD dimer, with one monomer colored in blue and the other in light blue (residues indicated with primes). (B) Interactions between the protein chains. Top and side views of the residues (represented as sticks) involved in dimer formation. Basic residues are shown in blue, polar in light green, aromatic in purple and aliphatic in yellow. (C) Comparison of dimer topology between CurH ^NDD and 5924 ^NDD. The 5924 ^NDD dimer is represented as in (A). The CurH ^NDD chains are colored orange and green. To emphasize the overall difference in architecture, one monomer of CurH ^NDD is superimposed on one monomer of 5924 ^NDD. (D) Cartoon representation of the 5925 ^CDD–5924 ^NDD dimer. 5925 ^CDD is shown in red (dark and light for the two monomers), 5924 ^NDD in blue, and the (G₃S)₃ linker in beige. The α-helices present in 5925 ^CDD are indicated as α1 and α2, and those in 5924 ^NDD as αA and αB. The primes indicate the second monomer of each DD. (E) Interactions between protein chains in the 5925 ^CDD–5924 ^NDD dimer. Residues involved in hydrophobic interactions (top view) and in polar interactions (bottom view) are shown as sticks. Basic residues are colored blue, polar in light green, methionine in dark green, aromatic in purple and aliphatic in yellow. (F) Comparison of the mode of ^CDD/^NDD interaction between CurG ^CDD–CurH ^NDD and 5925 ^CDD–5924 ^NDD. The figure shows a monomer of CurH ^NDD (orange) interacting with a monomer of CurG ^CDD (green), superimposed on monomers of 5925 ^CDD and 5924 ^NDD, as shown in (D). For clarity of presentation, certain linkers in the structures have been removed.

mode of dimerization differs: the CurH ^NDD monomers self-associate purely via formation of a coiled-coil, whereas in the 5924 ^NDD structure, helices αA also contact the coiled-coil formed by αB to yield an interleaved 4-helix bundle (**Figure 3C**).

In order to stably capture the docked complex, we next investigated three constructs in which the 5925 ^CDD was covalently fused to the 5924 ^NDD: i) directly (via a zero-length linker), ii) using a (G₃S)₂ linker, as previously described,^{16,21} or, iii) with a (G₃S)₃ linker (**Table S2**), the last of which yielded a stably-docked DD complex. Furthermore, comparative analysis of a mixture of the two independent DDs and the (G₃S)₃ covalent complex by CD yielded essentially the same spectra, showing that the artificial fusion had not fundamentally altered the interaction (**Figure S4**). Full structure elucidation by NMR as described above revealed an overall topology again close to type 2 docking interactions (**Figure 3D** and **Table S4**). Within the complex, the structure of the 5924 ^NDD is equivalent to that of the isolated domain, while 5925 ^CDD also adopts two α-helices (α1 E2621–G2627 and α2 D2630–L2637). The second α-

helix binds in a perpendicular orientation to an αB from the coiled-coil dimer, while helix α1 contacts both αA and αB, to form an overall 8 α-helix bundle.¹⁶ We further confirmed the overall shape of the complex and its molecular weight by SAXS (see **Figure S4** for more detailed discussion of these results).

The docking interface is hydrophobic and comprises: residues L2622 and I2625 from α1 of 5925 ^CDD interacting with Y14, L18 from 5924 ^NDD αA, and L21, A28 and L29 from 5924 ^NDD αB, as well as L2633, L2636 and L2637 from α2 of 5925 ^CDD interacting with M27, A28 and L31 of 5924 ^NDD αB and L15 of αA of the other monomer (**Figure 3E**). The structure additionally identified multiple polar interactions potentially contributing to docking specificity (5925 ^CDD/5924 ^NDD): E2626/R17 and/or R20, E2621 and/or E2624/K32, and potentially E2621/Q36. More specifically, for two of the interactions, the NMR structure suggested that a charged residue of one partner was sandwiched between two oppositely-charged residues of the other partner (**Figure 3E**). Consistent with the important role of such contacts in docking, the overall pI for the ^CDD

shows it to be acidic (pI = 3.8) and the ^NDD to be basic (pI = 9.7) (**Table 1**).

In order to evaluate their roles in driving assembly of the DD complex (k_{on}),⁴⁵ we targeted ^NDD residues of each pair, as we anticipated that effects on the ^CDD IDR would be difficult to interpret (i.e. that we would not be able to distinguish mutations which impacted the docking interaction from those which influenced the induced structuration of the ^CDD, or both). Three ^NDD residues (R17, K32 and Q36) were mutated to alanine and to glutamate, but of the two K32 mutants, only the K32A construct proved soluble. Far-UV CD analysis of all of the obtained mutants confirmed that their structures were not perturbed significantly relative to wild type 5924 ^NDD, while SAXS and SEC-SLS analysis (of all constructs except R17E and Q36A) demonstrated that they remained stably homodimeric (**Figure S4**). Following these controls, we measured the binding affinity of each 5924 ^NDD mutant to wild type 5925 ^CDD by ITC (experiments in triplicate), as well as its ability to induce structuration of the 5925 ^CDD by CD. None of the mutants produced a substantial decrease in the measured K_d s for the interaction or a change in binding stoichiometry (**Figure S4**) (R17A: 2.8 ± 0.6 μ M, $N = 1.3$; R17E: 5 ± 1 μ M, $N = 1.2$; K32A: 6.3 ± 0.5 μ M, $N = 1.2$; Q36E: 2.7 ± 0.1 μ M, $N = 1.1$), while a slightly improved binding affinity was determined for the interaction between 5924 ^NDD Q36A and the 5925 ^CDD ($K_d = 0.5 \pm 0.1$ μ M, $N = 1$). In addition, all of the ^NDD mutants still provoked proper structuration of 5925 ^CDD, as measured by CD. These data support the idea that the docking interaction cannot be dissected into individual pairs of contacting residues, but rather arises from an ensemble of contacts between the surfaces (e.g. tripartite rather than bipartite contacts between charged residues),⁴⁶ making it intrinsically robust to perturbation by single-site mutagenesis.

Finally, to compare the organization of the 5925 ^CDD–5924 ^NDD docking complex to that of the prototypical type 2 interaction between CurG ^CDD and CurH ^NDD,¹⁶ we determined the RMSD between the backbones of various portions of the structures, by specifically including or excluding helix α_2 of 5925 ^CDD. Comparing residues T2620–L2637 of 5925 ^CDD (i.e. the region encompassing both α -helices) and K11–L38 of 5924 ^NDD, to those of K4–I21 of CurG ^CDD and I49–L76 of CurH ^NDD, gave a backbone RMSD of 0.99 Å. However, the same calculation with only the residues encompassing α_1 of 5925 ^CDD (5925 ^CDD T2620–L2628 and 5924 ^NDD K11–38 vs. K4–S13 of CurG ^CDD and I49–L76 of CurH ^NDD) yielded an improved RMSD of 0.60 Å. Thus, overall, the topology of the first ^CDD α -helix and the ^NDD is identical for the two complexes, while that of the second ^CDD α -helix differs (**Figure 3F**). The 5925 ^CDD–5924 ^NDD complex architecture is therefore novel, but nonetheless ‘2-like’.

Docking at the 5924 ^CDD/5923 ^NDD split module interface. Unlike the other PKS interfaces within the enacyloxin system which are formed between C-terminal ACP and N-terminal KS domains, this intersubunit junction occurs in the middle of a module – nominally between an upstream KS and a downstream KR domain, although there are some 320+ amino acids at the end of subunit 5924 which could encompass an additional domain of unknown function (**Figure S1**). It was thus unclear whether the N-terminus of subunit 5923 must associate with the extreme C-terminus of subunit 5924, or rather directly with the

upstream KS domain (**Figure 1**). On the other hand, within the extreme C-terminal 53 amino acids of subunit 5924 there is only a single predicted α -helix, and so we focused on this region as a potential ^CDD, and aimed to determine whether it interacted with 5923 ^NDD.

Unfortunately, we were not able to produce either the ^CDD or the ^NDD as recombinant proteins from *E. coli*, and so these were obtained by chemical synthesis (peptides of 22 and 34 aa, respectively) (**Table 1** and **Table S2**); we also purchased an ^NDD missing the first 5 residues, to correspond with an alternative start site published during this work (this peptide was only investigated for its interaction with the ^CDD, *vide infra*). Both DDs were predicted to be α -helical (and the 5923 ^NDD weakly predicted to form a coiled-coil), while neither was predicted to be an IDR (**Table 1** and **Figures S2**).

Although CD analysis of the 5923 ^NDD confirmed its α -helical character and ability to fold independently (**Figure S5**), the 5924 ^CDD was completely unstructured, a result consistent with a putative IDR (**Figure S5**). The 5923 ^NDD was further shown to be homodimeric by SEC-SLS, but showed a loss of α -helical structuration (21.1% decrease) by CD when diluted from 100 to 10 μ M (**Figure S5**). As observed previously, dilution analysis by ITC of the 5924 ^CDD peptide yielded virtually no signal consistent with it being monomeric, but a weak heat signal was observed when the 5923 ^NDD was titrated on its own, in agreement with its homodimeric nature (**Figure S5**). Indeed, homodimerization of 5923 ^NDD is potentially important for the downstream KR, as the module lacks a dimerization motif which is often present upstream of KR domains in *cis*-AT PKSs,⁴⁷ and by SEC-SLS, a construct incorporating both the 5923 ^NDD and the downstream KR (5923 ^NDD-KR) was shown to be homodimeric, as anticipated (**Figure S5**).

By ITC and CD, we were unable to detect an interaction between the ^CDD and the two versions of the ^NDD under standard conditions (i.e. no induced structuration occurred) (**Figure S5**), despite the fact that they again exhibit highly divergent pIs suggestive of strong electrostatic attraction (5924 ^CDD = 4.6, 5923 ^NDD = 10.6) (**Table 1**). These results are consistent with the idea that this junction involves one or more up- and/or downstream domains within subunits 5924 and 5923 (**Figure 1**). To address this possibility, we attempted to express the 5924 ^CDD along with its upstream KS domain, and in addition to the 5923 ^NDD-KR construct described earlier, also attempted to obtain 5923 ^NDD-KR-ACP. Unfortunately, neither the 5924 ^CDD-KS nor the 5923 ^NDD-KR-ACP proteins proved soluble, while 5923 ^NDD-KR was unstable during analysis by ITC, precluding investigation of its interaction with the 5924 ^CDD.

Docking at the 5923 ^CDD/5922 ^NDD interface. Analysis of the potential 5923 ^CDD region predicted two or even three α -helices within the 54 residues (**Figure S2**), while the 32-residue 5922 ^NDD was predicted to contain a single α -helix and form a coiled-coil over the second half of the sequence (**Figure S2**). Given that docking by some ^CDDs involves two α -helices¹⁶ and others one,^{14,15} we expressed and purified a full-length version of the 5923 ^CDD encompassing all putative α -helices, and a short version including only the last α -helix (**Table S2**) (only the full-length gave an interaction by ITC (*vide infra*)). As observed for the other interfaces, analysis by CD showed the 5922 ^NDD to be independently well-structured and α -helical, although it underwent destructure upon dilution from 100 to 10 μ M (13.6% decrease in α -helix) (**Figure S6**). This, coupled

with evidence for monomerization by dilution ITC (**Figure S6**), suggests that the homodimer interface is not as stable as that of other ^NDDs within the same system. The 5923 ^CDD short and long versions also behaved similarly to the other ^CDDs, with no evidence from CD or ITC for structuration in the absence of ^NDD partner, again consistent with IDRs (**Figure S6**). Combining 5923 ^CDD long together with 5922 ^NDD lead to increased structuration as measured by CD (10% increase when measured at pH 6.5; 1% at pH 8.0), with a binding affinity for the interaction of 26 μ M and stoichiometry of 1, as determined by ITC (at pH 8.0) (**Figure S6**). In contrast, no induced structuration or interaction were detected when the short version of the ^CDD was used, indicating that multiple ^CDD α -helices are involved in docking. Interestingly, the electrostatic drivers for association of the domains appear reduced relative to the previously-discussed interfaces, as the pIs are much more closely matched (5.0 and 5.4, respectively, for 5923 ^CDD long and 5922 ^NDD) (**Table 1**).

Docking at the 5922 ^CDD/5921 ^NDD interface. The potential 5922 ^CDD region downstream of the C-terminal ACP is unusually long (80 residues (**Table S1**)); the first and the last portions were predicted to be α -helical, and essentially the whole region to behave as an IDR (**Figure S2**). We therefore expressed full-length 5922 ^CDD, but additionally a short version incorporating only the C-terminal-most 21 residues (**Table S2**). The 31-residue 5921 ^NDD which we expressed in its entirety (**Table S2**), was predicted to be α -helical and to (weakly) form a coiled-coil encompassing the full region (**Figure S2**). Neither of the expressed 5922 ^CDD versions (**Figure S3**) showed evidence by CD for independent structuration, nor for monomerization upon dilution by ITC (**Figure S7**), consistent with the IDR prediction. The 5921 ^NDD exhibited α -helical character, though less pronounced than the other ^NDDs, and was found to be monomeric by SEC-SLS (**Figure S7**). Indeed, we observed that diluting the 5921 ^NDD (from 100 to 10 and 1 μ M) produced little change in the CD spectrum, nor did the domain undergo obvious dissociation by ITC. We thus hypothesize that the 5921 ^NDD may require the presence of the downstream homodimeric KS in order to stably self-associate into a coiled-coil.

Nonetheless, by ITC, the 5921 ^NDD was found to interact with both 5922 ^CDD versions with low affinity. The measured K_d was approximately two-fold higher for the long 5922 ^CDD relative to the short ($K_d = 57 \mu\text{M}$ vs. $110 \mu\text{M}$; $N = 1$) (**Figure S7**), indicating a potential role for the extended ^CDD, and consistent with its IDR character and strong predicted interaction propensity (**Figure S2**). Nonetheless, combining the short 5922 ^CDD and the 5921 ^NDD produced a detectable increase in overall structuration by CD (10% increase in α -helix) (**Figure S7**), showing that full-length 5922 ^CDD is not obligatory for an interaction. Complementary electrostatic interactions are likely implicated in associating the DDs, as the pIs are widely different (4.1/3.9 and 9.7, respectively, for 5922 ^CDD long/short and 5921 ^NDD) (**Table 1**).

While we observed the 5922 ^CDDs and 5921 ^NDD to interact, the unexpected monomeric nature of the 5921 ^NDD and the overall weak docking affinity prompted us to consider whether the flanking domains were also implicated at this junction. To address this question, we aimed to express the full 5922 ^CDD along with the upstream ACP domain, and the 5921 ^NDD with the downstream KS. We successfully obtained recombinant 5922 ACP-^CDD, and evaluated binding to 5921 ^NDD by ITC

(**Figure S7**). Although we were unable to saturate binding due to the tendency of 5921 ^NDD to aggregate and the poor availability of 5922 ACP-^CDD, the measured affinity of $19 \pm 4 \mu\text{M}$ showed that the presence of the ACP increased the interaction strength by at least 3-fold. Unfortunately, the 5921 ^NDD-KS construct proved unstable, and so we were unable to judge the additional effects of potentially homodimerizing the ^NDD and the presence of the KS domain, on the interaction with the 5922 ^CDD.

Docking at the 5921 ^CDD/5920 ^NDD interface. As with the previous junction, α -helices were predicted for both the beginning and the end of the 5921 ^CDD, as well as strong IDR character (**Figure S2**). We therefore expressed and purified from *E. coli* a long (40 residues) and a short (20 residues) version of the 5921 ^CDD (**Figure S3** and **Table S2**). In the case of the ^NDD, we obtained two constructs, the first beginning from the originally-published start site (30 residues),²⁸ and the second from a revised start-site proposed while this work was in progress (44 residues) (https://www.ncbi.nlm.nih.gov/nuccore/NC_008392.1) (**Table S2**). Neither of the 5921 ^CDDs was structured by CD, nor did they exhibit evidence for monomerization by ITC dilution analysis (**Figure S8**), again consistent with already monomeric IDRs. The short version of the 5920 ^NDD was α -helical by CD (**Figure S8**). Unfortunately, we were unable to analyze the 5920 ^NDD long by this method, as the presence in the sequence of Cys residues required reducing conditions that were incompatible with CD. 5920 ^NDD short is likely homodimeric, as it shows evidence for monomerization by ITC (**Figure S8**). Consistent with this, 5920 ^NDD short is strongly predicted to form a coiled-coil over the full length of the domain (**Figure S3**). However, the homodimer is somewhat unstable, as destructureation was noted by CD on diluting the domain from 100 to 10 μM (8.0% decrease in α -helix) (**Figure S8**).

We next tested for interactions between both versions of the 5921 ^CDD and both versions of the 5920 ^NDD (long ^CDD/long ^NDD, long ^CDD/short ^NDD, and short ^CDD/short ^NDD) (**Figure S8**). As using the longer versions routinely yielded lower affinities, we could conclude that the two shorter versions are physiologically relevant (measured $K_d = 8.5 \pm 0.5 \mu\text{M}$; $N = 1.1$). As with the 5923 ^CDD/5922 ^NDD interface, the pI difference between the two partners is less pronounced than for other junctions (5921 ^CDD = 3.6, 5920 ^NDD = 6.4) (**Table 1**).

Docking at the *cis*-AT/*trans*-AT PKS interface, 5920 ^CDD/5919 ^NDD. The last junction in the PKS is potentially the most interesting, as it represents a functional interface between a *cis*-AT and a *trans*-AT PKS subunit. The 40-residue C-terminus of 5920 was predicted to contain one α -helix at the C-terminal end, with the whole region showing fairly strong IDR character (**Figure S2**). The 29-amino acid 5919 ^NDD was predicted to incorporate a single α -helix, and to form a coiled-coil encompassing the domain (**Figure S2**). This was the first surprising result, as the 4 α -helix bundle type ^NDDs described from *trans*-AT PKSs are monomeric and IDRs.^{21,22} Nonetheless, analysis of the *trans*-AT/*cis*-AT PKS junction in the kirromycin PKS (between subunits KirAV and KirAVI)⁴⁸ reveals putative docking domains with similar characteristics (**Figure S2**). While we were able to obtain the 5919 ^NDD as a recombinant protein from *E. coli* (**Figure S3** and **Table S2**), neither the full-length 5920 C-terminal region or a short version encompassing

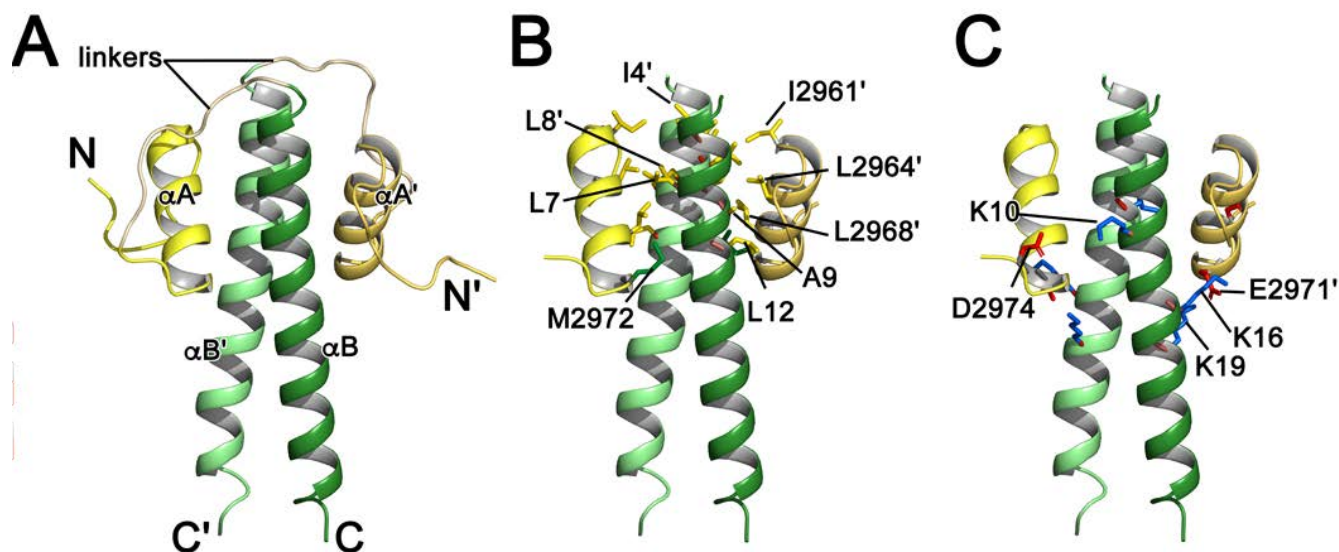


Figure 4. NMR structure of the 5920 ^{13}CDD –5919 ^{15}NDD complex. (A) Cartoon representation of the 5920 ^{13}CDD –5919 ^{15}NDD fusion construct, with 5920 ^{13}CDD shown in yellow and 5919 ^{15}NDD in green. (B) Hydrophobic interactions between 5920 ^{13}CDD and 5919 ^{15}NDD . Residues are shown as sticks. (C) Polar interactions between 5920 ^{13}CDD and 5919 ^{15}NDD . Residues are shown as sticks. In (B) and (C), aliphatic residues are colored yellow, methionine green, basic blue and acidic red. For clarity of presentation, linkers in panels (B) and (C) have been removed.

just the last α -helix, was well-expressed. We therefore purchased a synthetic peptide corresponding to the short version of 5920 ^{13}CDD . To test for the effect of the missing region of the 5920 ^{13}CDD , we also expressed and purified a construct which included the full 5920 ^{13}CDD and the upstream ACP (**Figure S3** and **Table S2**).

Analysis by CD of the 5920 ^{13}CDD peptide showed it to be unstructured, again in agreement with the IDR prediction (**Figure S9**). Indeed, analysis by SAXS yielded a Kratky plot³⁶ consistent with a disordered region (**Figure S9**, **Table S3**). The 5919 ^{15}NDD was independently well-folded and α -helical by CD, while dilution analysis by ITC and CD (from 100 to 10 μM , 42.2 % decrease in α -helix), supported its homodimeric character (**Figure S9**). Combining the two DDs yielded an increase in secondary structure by CD (17.3% increase in α -helix), evidence that the presence of 5919 ^{15}NDD induces 5920 ^{13}CDD folding, while ITC yielded a K_d of $43 \pm 3 \mu\text{M}$ and a stoichiometry of 0.8 (with exceptionally the 5920 ^{13}CDD in the syringe, as the 5919 ^{15}NDD aggregated at higher concentrations) (**Figure S9**). Repeating the experiment with 5919 ^{15}NDD and 5920 ACP- ^{13}CDD did not substantially change the affinity (estimated $K_d = 50 \mu\text{M}$, as we were not able to saturate the complex, **Figure S9**), confirming that only the last α -helix of the 5920 ^{13}CDD is important for docking.

To investigate this interaction further, we analyzed ^{15}N -labeled 5919 ^{15}NDD in the presence of unlabeled 5920 ^{13}CDD by NMR, demonstrating that the complex could be saturated with a large excess of 5920 ^{13}CDD (**Figure S9**). This result motivated us to solve the structure of the docking complex by NMR. For this, we created a construct in which the two DDs were fused using a $(\text{G}_3\text{S})_2$ linker,¹⁶ which yielded a stable protein. Analysis by SAXS confirmed that, as expected based on the behavior of 5919 ^{15}NDD , the fusion construct was homodimeric (**Figure S9**). Full structure elucidation of the complex was then carried out by heteronuclear multi-dimensional NMR (**Figure 4A** and **Table S4**), as described earlier. Overall, the 5920 ^{13}CDD –5919 ^{15}NDD

complex structure resembles those of the type 1 DD pairs DEBS 2 ^{13}CDD /DEBS 3 ^{15}NDD (type 1a)¹⁴ and PikAIII ^{13}CDD /PikAIV ^{15}NDD (type 1b)¹⁵, except that in contrast to the DEBS and Pik ^{13}CDD s which incorporate three successive α -helices (the first two of which constitute a dimerization motif, and the last of which mediates docking), the 5920 ^{13}CDD only contains a single α -helix.

Within the docking complex, 5919 ^{15}NDD forms a coiled-coil comprising residues Q2–A28, via hydrophobic contacts involving M1, I4, L8, I15, L18 and L25 of each monomer. The bound form of 5920 ^{13}CDD is structured, with residues I2961–E2971 forming a single α -helix. The docking interface comprises a series of hydrophobic interactions: I2961, L2964, L2968 and M2972 of 5920 ^{13}CDD contact I4, L8, A9 and L12 of one 5919 ^{15}NDD monomer and L7 of the other (**Figure 4B**). In addition, polar interactions are observed, with E2971 of one 5920 ^{13}CDD monomer forming salt bridges with K16 and K19 of one 5919 ^{15}NDD monomer, while D2974 of the same 5920 ^{13}CDD monomer potentially contacts K10 of the second 5919 ^{15}NDD monomer (**Figure 4C**). Based on this analysis, while the 5920 ^{13}CDD –5919 ^{15}NDD complex is clearly type 1, it is not possible to conclusively designate it as being 1a or 1b. For example, despite the presence of a single docking α -helix (type 1a-like), it is of intermediate length between those described for 1a and 1b structures (**Figure 1**).^{14,15} Furthermore, certain polar contacts at the interface (**Figure S9**) and comparative sequence analysis with other type 1a/1b pairs (**Figures S10** and **S11**) are consistent with both type 1a and type 1b interactions. We thus favor the attribution ‘1-like’ for this DD interface.

Analysis of interface cross-talk in the enacyloxin system. One critical role of docking domains is to ensure that the communication between the various subunits is specific. To directly test this aspect of the enacyloxin PKS, we evaluated whether there was any cross-talk between 5925 ^{13}CDD (two interaction α -helices) and the ^{15}NDD s of all of its noncognate subunits (5923–5919), as well as that between 5921 ^{13}CDD (one interac-

tion α -helix) and the N DDs of 5924–5921 and 5919, by ITC under the same conditions for which we detected binding between the native C DD/ N DD pairs (**Figure S12**). With one exception (that between 5921 C DD and 5924 N DD which yielded a very weak affinity ($>100 \mu\text{M}$)), no evidence for any non-cognate interaction was observed.

■ DISCUSSION

Since the discovery of the underlying modular genetic architecture of the PKSs, attempts have been made to reconfigure these systems by genetic engineering.¹¹ The limited success of these experiments despite several decades of effort may be due, at least in part, to the fact that the catalytic and carrier protein domains which comprise the modules exhibit an unexpectedly high degree of functional and structural interdependence,^{12,13} and thus are not truly ‘modular’. This observation suggests that a more productive strategy will be to combine intact modules and subunits to generate novel assembly lines. Although fusing these elements together is one potential strategy, it is arguably preferable to engineer non-covalent interfaces, as this avoids creating even larger proteins. However, the success of this strategy relies on obtaining a collection of multiple pairs of docking domains (DDs) to mediate such interactions, which: i) have known boundaries; ii) are truly portable into novel contexts, or for which the constraints in terms of flanking domains are defined; iii) are compatible with the type of PKS to be modified (i.e. *cis*-AT and/or *trans*-AT and mixed PKS/NRPS systems); and iv) are orthogonal to each other, affording specific intersubunit contacts, and thus a defined subunit order in the engineered system.

To further this goal, we have comprehensively characterized docking within the enacyloxin PKS, as it represents a particularly stringent test case for DD-facilitated interactions. By structure elucidation of fused DD complexes, we could conclusively assign the 5925/5924 and 5920/5919 interfaces to types 2- and 1-like, respectively. For the remaining junctions (5924/5923, 5923/5922, 5922/5921 and 5921/5920) reasonable assignments have been made based on the accumulated biophysical data in combination with comparative sequence analysis between the identified interaction regions and representative type 1a, 1b, 2 and 4-helix C DDs and N DDs (**Figures S10 and S11**; type 3 (SLiM- β H β D) could be excluded because the N DDs are insufficiently long). Taken together (**Figure 1** and **Table 1**), our results show that nature has deployed several strategies to achieve specificity in this multi-subunit system.

Firstly, docking is mediated by at least three distinct types of docking domain which are intrinsically non-interacting: 1-like (at the 5924/5923, 5922/5921, 5921/5920 and 5920/5919 junctions; one definitive experimental characteristic of these DDs permitting this assignment is their length (≤ 22 residues; **Table 1**)), 2-like (5925/5924 and 5923/5922) and 3 (5917 (PCP domain)/5915 (C domain)¹⁸). Furthermore, interfaces which involve the same DD class are not equivalent. Specifically, although the 5921/5920 and 5920/5919 junctions likely both involve type 1-like DDs, and the C DDs of both pairs lack the typical four α -helix dimerization motif of type 1 complexes,¹⁴ there is a pronounced divergence in the ΔpI s across these junctions (5921/5920 = 2.8 and 5920/5919 = 6.5 (**Table 1**)), suggesting a critical role for differential electrostatics in steering the correct associations.⁴⁵ In terms of the other two type 1-like junctions

(5924/5923 and 5922/5921), although the C DDs both incorporate a predicted single, extreme C-terminal interaction α -helix, the 5922 C DD utilizes the upstream IDR region to interact with the N DD partner. Both interfaces also appear to include contacts between the flanking domains, though these differ between the junctions (KS/KR for 5924/5923 and ACP/KS for 5922/5921), and in the case of the 5924/5923 interface, no binding is actually observed between the excised DDs in the absence of these neighbors. Several explanations for this latter result are possible, including that the DDs only function when linked to their respective up- and downstream domains, that they do not contact each other directly but instead the flanking domains (as with the DHDs²⁴), or indeed that DDs do not contribute to re-constituting a functional module at this point in the assembly line.

Finally, in the case of the two interfaces mediated by type 2 domains, divergent electrostatics appears to be the main contributor to partner selection, as the ΔpI across the 5925/5924 junction is 5.9, while that at the 5923/5922 interface is only 0.4. Indeed, the solved structure of the 5925/5924 docking complex (**Figure 3**) reveals at least three key polar interactions which likely contribute to specificity, although not in binary fashion.

Another notable finding of our work is that all of the C DDs show strong characteristics of IDRs. Although IDRs were previously identified in *trans*-AT PKSs,²¹ this is a novel finding for *cis*-AT PKSs. It is nonetheless possible that the docking α -helices of the previously-reported 1a¹⁴ and 1b¹⁵ type C DDs are in fact IDRs, but this behavior was missed as the C DDs were studied in the context of fused complexes with their partner N DDs, and thus were stably folded. In any case, it is now abundantly clear that nature favors the use of IDRs in docking as a means to achieve both relatively low affinity and highly specific binding.⁴⁹ Indeed, whereas classical IDRs often engage with multiple partners which has been argued to intrinsically reduce their ability to assure high interaction specificity,⁴⁹ C DDs in PKS systems need only bind a single N DD. We have also provided evidence by NMR and site-directed mutagenesis that, at least for the 5925 C DD/5924 N DD interaction, docking is intrinsically robust to single amino acid exchanges, a useful adaptation that would in principle prevent alterations in docking preference as a result of punctual mutations.

Even with high-resolution structural information in hand, the overall low sequence similarity between the DDs within the same structural class, means that we were not able to discern an obvious amino acid ‘code’ which correlates with docking specificity and which could be manipulated to influence partner choice (**Figures S10 and S11**). In any case, it is clear that docking preference depends on more than just a limited number of polar contacts that could be targeted for site-directed mutagenesis. In addition, such efforts would be complicated by the fact that a substantial number of both C DDs and N DDs are now known to be IDRs, for which the effects of point mutations on induced structuration/docking are difficult to predict. Taken together, these results argue that the most productive strategy for deploying our growing toolbox of DDs in PKS engineering will be to use pairs of matched domains, including those identified here.¹¹

In the case of type 1(-like) DDs, these could encompass the upstream dimerization motif as demonstrated previously,⁵⁰ or include only the third docking α -helix, which is functionally in-

dependent.¹⁴ Specificity can be achieved by employing multiple, intrinsically orthogonal structural types, but a single class can also be used at several interfaces if, for example, there are large discrepancies in the ΔpI s of the DDs sets or if their dependencies on flanking domains are different. We have also shown that type 1-like DDs can mediate interaction between an upstream *cis*-AT PKS and a downstream *trans*-AT PKS, while sequence analysis of KirAV^{CDD} and KirAVI^{NDD} (a *trans*-AT/*cis*-AT junction) suggests that they are also type 1(a) DDs (Figure S13). Thus, this type of DD should be a generally suitable element for creating additional such hybrid systems as a means of boosting polyketide diversity.

■ MATERIALS AND METHODS

Sequence Analysis. The sequences of the type 1a^{CDD}s and^{NDD}s were obtained from ref.¹⁴, the type 1b domains from ref.¹⁵, the type 2 DDs from ref.¹⁶, and additional 4 α -helix type domains from ref.²². For analysis of the putative enacyloxin^{CDD}s, the C-terminal boundaries of the upstream domains (ACP or KS) were determined by multiple sequence alignment of the subunit ends using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), and comparison with domain boundaries identified in the solved structures of homologous domains.^{51–54} Similarly, the start sites of the domains downstream of the N-terminal docking domains (KS or KR (refs.^{55,56})) were assigned by alignment of the beginning of the subunits and comparison with the sequences of the solved structures of homologous domains. Sequence alignments shown in figures were generated using Clustal Omega and displayed using BoxShade (http://www.ch.embnet.org/software/BOX_form.html), and sequence logos were created with WebLogo 3 (weblogo.threeplusone.com/create.cgi). Secondary structure prediction was carried out with PSIPRED,³² analysis for disorder and interaction propensity with IUPred2A (<https://iupred2a.elte.hu/>),³³ the presence of coiled-coils predicted with the NPS@ web server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_lupas.html),³⁴ and pIs calculated using ExPASy (https://web.expasy.org/compute_pi/).³⁵

Materials and DNA Manipulation and Sequencing. *Burkholderia ambifaria* AMMD was purchased from BCCM/LMG (Ghent University, Belgium) and genomic DNA was extracted using Wizard Genomic DNA Purification Kit (Promega). 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). Isotopically-labeled chemicals were sourced from Eurisotop. The enzymes for genetic manipulation were purchased from ThermoFisher Scientific. DNA isolation and manipulation in *Escherichia coli* was carried out according to standard methods.⁵⁷ PCR amplifications were performed 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). Synthetic peptides were obtained from GeneCust (Boynes, France).

Cloning, Expression and Purification. Twenty-three constructs (5925^{CDD}, 5924^{NDD}, 5924^{CDD}, 5924^{KS-CDD}, 5923^{NDD}, 5923^{NDD-KR}, 5923^{NDD-KR-ACP}, 5923^{CDD} long and short versions, 5922^{NDD}, 5922^{CDD} long and short

versions, 5922^{ACP-CDD}, 5921^{NDD}, 5921^{NDD-KS}, 5921^{CDD} long and short versions, 5920^{NDD} long and short versions, 5920^{CDD} long and short versions, 5920^{ACP-CDD} and 5919^{NDD}) were amplified directly from *B. ambifaria* genomic DNA using forward primers incorporating a *Bam*HI restriction site (or in select cases, *Sac*I or *Sal*I restriction sites) and a reverse primer including a *Hind*III restriction site (Table S1), and were ligated into the equivalent sites of vector pBG102 (or pLM302 for 5922^{ACP-CDD} only). Vector pBG102 codes for a His₆-SUMO tag and pLM-302 codes for a His₆-maltose binding protein (MBP) tag (Center for Structural Biology, Vanderbilt University). Following cleavage of the tags, the proteins incorporated a non-native N-terminal Gly-Pro-Gly-Ser sequence when *Bam*HI was used, a Gly-Pro-Gly-Ser-Pro-Asn-Ser-Ser sequence for *Sac*I or a Gly-Pro-Gly-Ser-Pro-Asn-Ser-Ser-Ser-Val-Glu sequence for *Sal*I. Nonetheless, these additional residues were not expected to impact on docking, because it has been shown repeatedly that it is possible to covalently fuse^{CDD} and^{NDD} pairs using variable length linkers, without altering the native interactions.^{14–16,21} From selected vectors and *B. ambifaria* genomic DNA, covalent fusions were constructed (5925^{CDD}–5924^{NDD} and 5920^{CDD}–5919^{NDD}) using overlap PCR. Specifically, fragments encoding 5925^{CDD}, 5924^{NDD}, 5920^{CDD} and 5919^{NDD} were amplified, and then used as templates in a second PCR reaction to obtain the fusion constructs. The resulting fragments were ligated into vector pBG102. Site-directed mutagenesis of 5924^{NDD} was carried out using Phusion high-fidelity DNA polymerase and vector pBG102 containing the wild-type sequence.

The vectors were used to transform *E. coli* BL21 or Rosetta 2 (DE3) (Novagen), and constructs were expressed as His₆-SUMO-tagged or His₆-MBP-tagged fusions by growth in LB medium at 37 °C to an A₆₀₀ of 0.7, followed by induction with 300 μ M IPTG and incubation at 20 °C for a further 12–18 h. The *E. coli* cells were collected by centrifugation (3500 g) and resuspended in buffer 1 (50 mM Tris (pH 8.0), 300 mM NaCl). Cells were lysed by sonication, and cell debris was removed by centrifugation and filtration (0.45 μ m). When soluble protein was obtained (*vide infra*), 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 50 mM or 20 mM imidazole, and His-tagged proteins were eluted using 500 mM imidazole.

All His₆-tagged constructs 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 10 mM or 20 mM imidazole. Following this step, size-exclusion chromatography on a Superdex 75 26/60 column (GE) in 100 mM sodium phosphate (pH 6.5), or 20 mM Tris and 100 mM NaCl (pH 8.0) for 5923^{CDD} long version and 5922^{NDD}, or on a Superdex 200 26/60 column (GE) in the case of 5923^{NDD-KR}, resulted in a homogeneous preparation of each protein (Figure S1).

Production of ¹⁵N- and ¹³C,¹⁵N-enriched 5925^{CDD}, 5924^{NDD}, 5919^{NDD}, 5925^{CDD}–5924^{NDD} (covalent fusion) and 5920^{CDD}–5919^{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

$^{13}\text{C}_6$ -glucose (2.0 g/L) as the sole sources of nitrogen and carbon. The isotopically-labeled constructs were purified to homogeneity as described above (Figure S1).

Size-exclusion Chromatography Coupled to Static Light Scattering. SEC-SLS were performed using two different setups (SEC-RALS/LALS or SEC-MALS): (i) AKTA Explorer HPLC system (GE Healthcare) coupled to both RALS/LALS light scattering and refractive index detectors (Viscotek TDA302, Malvern Panalytical). The SEC columns (Shodex KW-802.5 or KW804, 7 μm , 8 mm ID $\times$ 300 mm, void volume $\sim$ 6 mL, total volume $\sim$ 12.5 mL) were equilibrated with NaPi buffer (100 mM, pH 6.5). The flow rate ranged from 0.2–0.5 mL/min, and the temperature was 30 $^\circ\text{C}$. Data were processed with the Omnisec software (v5.12, Malvern Panalytical) using RALS/LALS detectors for 5923 $^{\text{NDD}}$ -KR, and only the RALS detector for the other proteins; and (ii) (for the 5923 $^{\text{NDD}}$) AKTA pure (GE Healthcare) coupled with a MALS detector (miniDAWN TREOS II, Wyatt Technology) and a refractometer (Optilab T-rEx, Wyatt Technology). The SEC column (Superdex 75, 10 mm ID $\times$ 300 mm, void volume $\sim$ 8 mL, total volume $\sim$ 24 mL) was equilibrated with 20 mM Tris.HCl (pH 7.5), 500 mM NaCl. The flow rate was 0.5 mL/min and the temperature was 25 $^\circ\text{C}$. Data were processed with the Astra software (v. 7.14, Wyatt Technology). The refractometer was used as the concentration detector and the refractive index increment value (dn/dc) used to determine the molecular weight was 0.185 mL/g.

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 185/190/195–260 nm at 20 $^\circ\text{C}$, using a temperature-controlled chamber. For most measurements, a 0.01 cm cuvette containing 30 μL of protein sample at 100 μM was used. In the case of 5922 $^{\text{CDD}}$, 5921 $^{\text{NDD}}$, 5920 $^{\text{CDD}}$ and 5919 $^{\text{NDD}}$, measurements were made on 300 μM samples. For analysis of 10 μM protein, 200 μL of sample were used in a 0.1 cm cuvette, and for 1 μM , 1 mL of sample in a 1 cm cuvette. Measurements were performed 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 two docking domains were combined,²¹ we first acquired two sets of spectra corresponding to the docking domains alone, either by placing the $^{\text{CDD}}$ into one cuvette and the partner $^{\text{NDD}}$ into another (30 μL each at 100 μM or 300 μM for 5922 $^{\text{CDD}}$, 5921 $^{\text{NDD}}$, 5920 $^{\text{CDD}}$ and 5919 $^{\text{NDD}}$) and analyzing the two cuvettes simultaneously, or by averaging the spectra of the $^{\text{CDD}}$ and $^{\text{NDD}}$ acquired separately (these protocols consistently yielded the same result). We then combined the two docking domains, and placed the mixture into both cuvettes (30 μL at 100 μM or 300 μM for 5922 $^{\text{CDD}}$, 5921 $^{\text{NDD}}$, 5920 $^{\text{CDD}}$ and 5919 $^{\text{NDD}}$), again analyzing the two simultaneously. Spectrum deconvolution was carried out using the CDNN software.³⁷

Analysis by Isothermal Titration Calorimetry. ITC was performed using a MicroCal ITC 200 (GE) at 20 $^\circ\text{C}$. A 300 μL aliquot of $^{\text{CDD}}$ (wild type) or $^{\text{NDD}}$ (wild type or mutant) at 60 μM , 120 μM or 200 μM was placed in the calorimeter cell, and $^{\text{NDD}}$ (wild type or mutant) or wild type $^{\text{CDD}}$, at 600 μM , 1.8 mM or 2 mM, was added as follows: 0.4 μL over 0.8 s for the first injection, followed by 19 injections of 2 μL over 4 s at 120

s intervals. The heat of reaction per injection (microcalories/second) was determined by integration of the peak areas using either the Origin (MicroCal) or Ntripic software.⁵⁸ 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 the DDs alone (initial concentrations of 600 μM , 1.8 mM or 2 mM), and were subtracted from the corresponding experiments before curve fitting.

NMR Data Acquisition. To minimize the amount of protein needed, 300 μL of protein solution were loaded into a 4 mm NMR tube. All NMR data were recorded at 25 $^\circ\text{C}$ on a Bruker DRX600 spectrometer equipped with a cryogenic probe (Unité Mixte de Service (UMS) 2008 Ingénierie-Biologie-Santé en Lorraine (IBSLor)). Backbone and sequential resonance assignments were obtained by the combined use of 2D ^{15}N - ^1H and ^{13}C - ^1H HSQC spectra and 3D HNCA, HNCACB, CBCA(CO)NH, HNHA, HBHA(CO)NH, HN(CA)CO, and HNCO experiments. Assignments of aliphatic side chain resonance were based on 2D aromatic ^{13}C - ^1H HSQC, 3D (H)CC(CO)NH, H(CC)(CO)NH, HNCO-ECOSY, CCH-TOCSY, and HCCH-TOCSY experiments (reviewed in ref. ⁵⁹). To collect NOE-based distance restraints for the 5924 $^{\text{NDD}}$, 5925 $^{\text{CDD}}$ -5924 $^{\text{NDD}}$ and 5920 $^{\text{CDD}}$ -5919 $^{\text{NDD}}$ structure calculations, 3D ^{15}N NOESY-HSQC and ^{13}C NOESY-HSQC were recorded on uniformly ^{13}C , ^{15}N enriched 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.⁴⁰ 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.⁶⁰ In each cycle, 100 structures starting from random torsion angle values were calculated with 15 000 steps of torsion angle dynamics-driven simulated annealing. A total of 1824 NOE-based distances and 116 backbone angle restraints were used for structure calculation of the 5924 $^{\text{NDD}}$ domain, whereas 2875 and 2504 NOE-based distances associated with 136 and 148 backbone angle restraints were used for the 5925 $^{\text{CDD}}$ -5924 $^{\text{NDD}}$ and 5920 $^{\text{CDD}}$ -5919 $^{\text{NDD}}$ complexes, respectively. To ensure that the overall structures of 5925 $^{\text{CDD}}$ -5924 $^{\text{NDD}}$ and 5920 $^{\text{CDD}}$ -5919 $^{\text{NDD}}$ were homodimeric, weak restraints (9 $\text{\AA}$) were applied to the two monomers of the $^{\text{NDD}}$ regions during the first two cycles of CYANA calculation. The angle restraints were obtained from $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, $^{13}\text{C}'$, ^{15}N , ^1HN , and $^1\text{H}\alpha$ chemical shifts using TALOS-N⁶¹ with an assigned minimum range of $\pm 20^\circ$. 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. ⁴¹ and ⁴²) 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 representative ensembles correspond to the 25 conformers from each calculation with the lowest restraint energy terms.

Small angle X-ray Data Collection. SAXS data were collected on the SWING beamline at the Synchrotron SOLEIL (France). Frames were recorded using an Eiger 4M detector at an energy of 12 keV. The distance between the sample and the detector was set to 2000 mm, leading to scattering vectors q ranging from 0.0005 to 0.5 Å⁻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 Å (4.6 mm ID × 300 mm), Agilent) 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.⁶² After equilibrating the column in the protein buffer (20 mM Tris.HCl (pH 7.5), 200 mM NaCl, 5% glycerol (as radioprotectant)), 100 successive frames of 0.75 s were recorded of the buffer background. A 50 µL aliquot of the protein sample (at 10 or 30 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).⁶² Data reduction to absolute units, frame averaging, and solvent 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.

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.⁶³

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,⁶⁴ which depends on integration of the distance distribution function (because this is known to be proportional to molecular weight), and size exclusion chromatography-static light scattering (SEC-SLS), for which the intensity of light scattered off protein molecules in solution is proportional to their molecular weight.⁶⁵

For MW determination by HPLC-MS, samples were equilibrated to 0.5% TFA, 2% ACN and diluted, if required, to 1 pM, and 5 µl were fractionated through nano-HPLC on an Ultimate 3000 system equipped with a pepMap 100 C18 desalting column and a 15 cm, pepMap RSLC C18 fractionation column (all from ThermoFischer) and eluted using a 2–45 % ACN gradient over 30 min at 300 nL/min. Eluting material was detected based on its absorbance at 214 nm to verify the relative purity of compounds. Fractions (170, 9 s each) were collected on a Proteiner FcII and elutions were mixed on MTP-1536 TF target spots with α -cyano-4-hydroxycinnamic acid (all from Bruker). Spots corresponding to eluting peaks based on absorbance were investigated by MALDI-TOF and MALDI-TOF/TOF (in LIFT mode) mass spectrometry on an Autoflex IV (Bruker), using near neighbor external calibration.

***Ab initio* Shape Determination.** *Ab initio* protein shapes were calculated from the experimental SAXS data using the bead-modeling program DAMMIN.⁶⁶ At least 20 to 30 independent fits were carried out without imposition of symmetry for shape determination without restraint, and with a 2-fold symmetry restraint when the dimeric state had been established by SAXS MoW and/or SEC-SLS. The results of several DAMMIN models were then compared using the program SUPCOMB⁶⁷ in order to determine a consensus model. Low-resolution models were then averaged using the DAMAVER and DAMFILT packages.^{68,69} The quality of the 3D modeling was determined using the discrepancy χ^2 , defined according to Konarev *et al.*⁶⁸ 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.⁷⁰ 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.⁷⁰ High-resolution structures were then superimposed into the *ab initio* protein shapes using SUPCOMB.⁶⁷ Figures displaying the protein structures were generated using PYMOL (Schrödinger, LLC), and the low-resolution bead models from DAMMIN were converted into mesh.

Accession numbers. Chemical shifts for 5924 ¹⁵NDD, 5925 ¹³DD–5924 ¹⁵NDD and for 5920 ¹³DD–5919 ¹⁵NDD have been deposited in the Biological Magnetic Resonance Bank with acces-

sion codes 34446, 34448 and 34447, respectively. PDB coordinates for 5924 ^NDD, 5925 ^CDD–5924 ^NDD and for 5920 ^CDD–5919 ^NDD have been deposited under accession codes 6TDD, 6TDN and 6TDM.

ASSOCIATED CONTENT

Supplementary results including: supplementary methods, 13 supplementary figures, 4 supplementary tables and supplementary references. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Authors

*arnaud.gruez@univ-lorraine.fr

*benjamin.chagot@univ-lorraine.fr

*kira.weissman@univ-lorraine.fr

Present Addresses

[†]Department of Chemistry, University of Basel, Mattenstrasse 24a, CH-4058 Basel, Switzerland; Department of Biosystems Science and Engineering, ETH-Zürich, Mattenstrasse 26, CH-4058 Basel, Switzerland

[‡]Université de Lorraine, Laboratoire d'Ingénierie des Biomolécules (LIBio), 2 Avenue de la Forêt de Haye, TSA 40602, 54518, Vandœuvre-lès-Nancy, France

Author Contributions

This manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported by the Agence Nationale de la Recherche (ANR-16-CE92-0006-01 PKS STRUCTURE to K.J.W.), the Centre National de la Recherche Scientifique (CNRS), and the 'IMPACT Biomolécules' project of the Lorraine Université d'Excellence (Investissements d'avenir – ANR 15-004).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge Javier Perez and Aurelien Thureau (Soleil Synchrotron, Swing) for help in data acquisition. The NMR data were recorded on the NMR spectrometer of the Plateforme de Biophysique et Biologie Structurale (B2S) of the UMS2008 IBSLor (Université de Lorraine-CNRS-INSERM – <http://umsibslor.univ-lorraine.fr>), and we would also like to thank Jean-Baptiste Vincourt from the Proteomics Core Facility, while ITC and SLS data were acquired on the Plateforme Approches Fonctionnelles et Structurales des InterActions Cellulaires (ASIA), of the Faculté des Sciences et Technologies. Appreciation also goes to Alexandre Kriznik for assistance in analyzing the ITC and CD data.

REFERENCES

- (1) Davison, E. K.; Brimble, M. A. Natural Product Derived Privileged Scaffolds in Drug Discovery. *Curr Opin Chem Biol* **2019**, *52*, 1–8. <https://doi.org/10.1016/j.cbpa.2018.12.007>.
- (2) Staunton, J.; Weissman, K. J. Polyketide Biosynthesis: A Millennium Review. *Nat Prod Rep* **2001**, *18* (4), 380–416.
- (3) Weissman, K. J. Introduction to Polyketide Biosynthesis. *Meth. Enzymol.* **2009**, *459*, 3–16. [https://doi.org/10.1016/S0076-6879\(09\)04601-1](https://doi.org/10.1016/S0076-6879(09)04601-1).
- (4) Jenke-Kodama, H.; Sandmann, A.; Müller, R.; Dittmann, E. Evolutionary Implications of Bacterial Polyketide Synthases. *Mol. Biol. Evol.* **2005**, *22* (10), 2027–2039. <https://doi.org/10.1093/molbev/msi193>.
- (5) Jenke-Kodama, H.; Dittmann, E. Evolution of Metabolic Diversity: Insights from Microbial Polyketide Synthases. *Phytochemistry* **2009**, *70* (15–16), 1858–1866. <https://doi.org/10.1016/j.phytochem.2009.05.021>.
- (6) 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.* **2008**, *26* (2), 225–233. <https://doi.org/10.1038/nbt1379>.
- (7) Helfrich, E. J. N.; Piel, J. Biosynthesis of Polyketides by Trans-AT Polyketide Synthases. *Nat Prod Rep* **2016**, *33* (2), 231–316. <https://doi.org/10.1039/c5np00125k>.
- (8) O'Brien, R. V.; Davis, R. W.; Khosla, C.; Hillenmeyer, M. E. Computational Identification and Analysis of Orphan Assembly-Line Polyketide Synthases. *J. Antibiot.* **2014**, *67* (1), 89–97. <https://doi.org/10.1038/ja.2013.125>.
- (9) 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* **2010**, *27* (4), 571–616.
- (10) Sundaram, S.; Hertweck, C. On-Line Enzymatic Tailoring of Polyketides and Peptides in Thiotemplate Systems. *Curr Opin Chem Biol* **2016**, *31*, 82–94. <https://doi.org/10.1016/j.cbpa.2016.01.012>.

- (11) Weissman, K. J. Genetic Engineering of Modular PKSs: From Combinatorial Biosynthesis to Synthetic Biology. *Nat Prod Rep* **2015**. <https://doi.org/10.1039/c5np00109a>.
- (12) 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.; et al. Structure of a Modular Polyketide Synthase. *Nature* **2014**, *510* (7506), 512–517. <https://doi.org/10.1038/nature13423>.
- (13) 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.; et al. Structural Rearrangements of a Polyketide Synthase Module during Its Catalytic Cycle. *Nature* **2014**, *510* (7506), 560–564. <https://doi.org/10.1038/nature13409>.
- (14) 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.* **2003**, *10* (8), 723–731.
- (15) Buchholz, T. J.; Geders, T. W.; Bartley, 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.* **2009**, *4* (1), 41–52. <https://doi.org/10.1021/cb8002607>.
- (16) 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.* **2013**, *20* (11), 1340–1351. <https://doi.org/10.1016/j.chem-biol.2013.09.015>.
- (17) Richter, C. D.; Nietlispach, D.; Broadhurst, R. W.; Weissman, K. J. Multienzyme Docking in Hybrid Megasyntetases. *Nat. Chem. Biol.* **2008**, *4* (1), 75–81. <https://doi.org/10.1038/nchembio.2007.61>.
- (18) Kosol, S.; Gallo, A.; Griffiths, D.; Valentini, T. R.; Masschelein, J.; Jenner, M.; de Los Santos, E. L. C.; Manzi, L.; Sydor, P. K.; Rea, D.; et al. Structural Basis for Chain Release from the Enacyloxin Polyketide Synthase.
- (19) Hacker, C.; Cai, X.; Kegler, C.; Zhao, L.; Weickhmann, A. K.; Wurm, J. P.; Bode, H. B.; Wöhnert, J. Structure-Based Redesign of Docking Domain Interactions Modulates the Product Spectrum of a Rhabdopeptide-Synthesizing NRPS. *Nat Commun* **2018**, *9* (1), 4366. <https://doi.org/10.1038/s41467-018-06712-1>.
- (20) Weissman, K. J. The Structural Basis for Docking in Modular Polyketide Biosynthesis. *Chembiochem* **2006**, *7* (3), 485–494. <https://doi.org/10.1002/cbic.200500435>.
- (21) Dorival, J.; Annaval, T.; Risser, F.; Collin, S.; Roblin, P.; Jacob, C.; Gruez, A.; Chagot, B.; Weissman, K. J. Characterization of Intersubunit Communication in the Virginiamycin Trans-Acyl Transferase Polyketide Synthase. *J. Am. Chem. Soc.* **2016**. <https://doi.org/10.1021/jacs.5b13372>.
- (22) 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.* **2016**, *11* (9), 2466–2474. <https://doi.org/10.1021/acscchem-bio.6b00345>.
- (23) van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; et al. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114* (13), 6589–6631. <https://doi.org/10.1021/cr400525m>.
- (24) Jenner, M.; Kosol, S.; Griffiths, D.; Prasongpholchai, P.; Manzi, L.; Barrow, A. S.; Moses, J. E.; Oldham, N. J.; Lewandowski, J. R.; Challis, G. L. Mechanism of Intersubunit Ketosynthase-Dehydratase Interaction in Polyketide Synthases. *Nat. Chem. Biol.* **2018**, *14* (3), 270–275. <https://doi.org/10.1038/nchembio.2549>.

- (25) Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. Mechanism and Specificity of the Terminal Thioesterase Domain from the Erythromycin Polyketide Synthase. *Chem. Biol.* **1999**, *6* (2), 117–125. [https://doi.org/10.1016/S1074-5521\(99\)80008-8](https://doi.org/10.1016/S1074-5521(99)80008-8).
- (26) Acuner Ozbabacan, S. E.; Engin, H. B.; Gursoy, A.; Keskin, O. Transient Protein-Protein Interactions. *Protein Eng. Des. Sel.* **2011**, *24* (9), 635–648. <https://doi.org/10.1093/protein/gzr025>.
- (27) Nooren, I. M. A.; Thornton, J. M. Structural Characterisation and Functional Significance of Transient Protein-Protein Interactions. *J. Mol. Biol.* **2003**, *325* (5), 991–1018.
- (28) 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.* **2011**, *18* (5), 665–677. <https://doi.org/10.1016/j.chembiol.2011.01.020>.
- (29) <http://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed>.
- (30) Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.; et al. Minimum Information about a Biosynthetic Gene Cluster. *Nat. Chem. Biol.* **2015**, *11* (9), 625–631. <https://doi.org/10.1038/nchembio.1890>.
- (31) Masschelein, J.; Sydor, P. K.; Hobson, C.; Howe, R.; Jones, C.; Roberts, D. M.; Yap, Z. L.; Parkhill, J.; Mahenthiralingam, E.; G. L. Challis. A Dual Transacylation Mechanism for Polyketide Synthase Chain Release in Enacyloxin Antibiotic Biosynthesis.
- (32) Buchan, D. W. A.; Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz297>.
- (33) Mészáros, B.; Erdos, G.; Dosztányi, Z. IU-Pred2A: Context-Dependent Prediction of Protein Disorder as a Function of Redox State and Protein Binding. *Nucleic Acids Res.* **2018**, *46* (W1), W329–W337. <https://doi.org/10.1093/nar/gky384>.
- (34) Lupas, A.; Van Dyke, M.; Stock, J. Predicting Coiled Coils from Protein Sequences. *Science* **1991**, *252* (5009), 1162–1164. <https://doi.org/10.1126/science.252.5009.1162>.
- (35) Protein Identification and Analysis Tools in the ExPASy Server https://web.expasy.org/compute_pi/pi_tool-doc.html (accessed May 9, 2019).
- (36) Receveur-Brechot, V.; Durand, D. How Random Are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. *Curr. Protein Pept. Sci.* **2012**, *13* (1), 55–75.
- (37) Böhm, G.; Muhr, R.; Jaenicke, R. Quantitative Analysis of Protein Far UV Circular Dichroism Spectra by Neural Networks. *Protein Eng.* **1992**, *5* (3), 191–195.
- (38) Uversky, V. N. A Decade and a Half of Protein Intrinsic Disorder: Biology Still Waits for Physics. *Protein Sci.* **2013**, *22* (6), 693–724. <https://doi.org/10.1002/pro.2261>.
- (39) Sahu, D.; Bastidas, M.; Lawrence, C. W.; Noid, W. G.; Showalter, S. A. Assessing Coupled Protein Folding and Binding Through Temperature-Dependent Isothermal Titration Calorimetry. *Meth. Enzymol.* **2016**, *567*, 23–45. <https://doi.org/10.1016/bs.mie.2015.07.032>.
- (40) Güntert, P. Automated NMR Structure Calculation with CYANA. *Methods Mol. Biol.* **2004**, *278*, 353–378. <https://doi.org/10.1385/1-59259-809-9:353>.
- (41) (9) AMBER 14, University of California, San Francisco https://www.researchgate.net/publication/313508727_AMBER_14_University_of_California_San_Francisco (accessed Jun 22, 2018).

- (42) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J Comput Chem* **2005**, *26* (16), 1668–1688. <https://doi.org/10.1002/jcc.20290>.
- (43) Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR* **1996**, *8* (4), 477–486.
- (44) 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. D Biol. Crystallogr.* **2010**, *66* (Pt 1), 12–21. <https://doi.org/10.1107/S0907444909042073>.
- (45) Shammass, S. L.; Crabtree, M. D.; Dahal, L.; Wicky, B. I. M.; Clarke, J. Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. *J. Biol. Chem.* **2016**, *291* (13), 6689–6695. <https://doi.org/10.1074/jbc.R115.692715>.
- (46) Longhi, S.; Receveur-Bréchet, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.; Yeo, R.; Finet, S.; Canard, B. The C-Terminal Domain of the Measles Virus Nucleoprotein Is Intrinsically Disordered and Folds upon Binding to the C-Terminal Moiety of the Phosphoprotein. *J. Biol. Chem.* **2003**, *278* (20), 18638–18648. <https://doi.org/10.1074/jbc.M300518200>.
- (47) Zheng, J.; Fage, C. D.; Demeler, B.; Hoffman, D. W.; Keatinge-Clay, A. T. The Missing Linker: A Dimerization Motif Located within Polyketide Synthase Modules. *ACS Chem. Biol.* **2013**, *8* (6), 1263–1270. <https://doi.org/10.1021/cb400047s>.
- (48) Weber, T.; Laiple, K. J.; Pross, E. K.; Textor, A.; Grond, S.; Welzel, K.; Pelzer, S.; Vente, A.; Wohlleben, W. Molecular Analysis of the Kirromycin Biosynthetic Gene Cluster Revealed Beta-Alanine as Precursor of the Pyridone Moiety. *Chem. Biol.* **2008**, *15* (2), 175–188. <https://doi.org/10.1016/j.chem-biol.2007.12.009>.
- (49) Mollica, L.; Bessa, L. M.; Hanouille, X.; Jensen, M. R.; Blackledge, M.; Schneider, R. Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Front. Mol. Biosci.* **2016**, *3*. <https://doi.org/10.3389/fmolb.2016.00052>.
- (50) Yan, J.; Gupta, S.; Sherman, D. H.; Reynolds, K. A. Functional Dissection of a Multimodular Polypeptide of the Pikromycin Polyketide Synthase into Monomodules by Using a Matched Pair of Heterologous Docking Domains. *Chembiochem* **2009**, *10* (9), 1537–1543. <https://doi.org/10.1002/cbic.200900098>.
- (51) Alekseyev, V. Y.; Liu, C. W.; Cane, D. E.; Puglisi, J. D.; Khosla, C. Solution Structure and Proposed Domain Domain Recognition Interface of an Acyl Carrier Protein Domain from a Modular Polyketide Synthase. *Protein Sci.* **2007**, *16* (10), 2093–2107. <https://doi.org/10.1110/ps.073011407>.
- (52) 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.* **2014**, *5* (8), 3081–3095. <https://doi.org/10.1039/C3SC53511H>.
- (53) 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.* **2006**, *103* (30), 11124–11129. <https://doi.org/10.1073/pnas.0601924103>.
- (54) Gay, D. C.; Gay, G.; Axelrod, A. J.; Jenner, M.; Kohlhaas, C.; Kampa, A.; Oldham, N. J.; Piel, J.; Keatinge-Clay, A. T. A Close Look at a Ketosynthase from a Trans-Acyltransferase Modular Polyketide Synthase. *Structure* **2014**, *22* (3), 444–451. <https://doi.org/10.1016/j.str.2013.12.016>.

- (55) 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* **2006**, *14* (4), 737–748. <https://doi.org/10.1016/j.str.2006.01.009>.
- (56) Keatinge-Clay, A. T. A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol.* **2007**, *14* (8), 898–908. <https://doi.org/10.1016/j.chembiol.2007.07.009>.
- (57) Sambrook, J.; Fritsch, E. F.; Maniatis, T. Molecular Cloning: A Laboratory Manual. *Molecular cloning: a laboratory manual*. **1989**, No. Ed. 2.
- (58) Keller, S.; Vargas, C.; Zhao, H.; Piszczek, G.; Brautigam, C. A.; Schuck, P. High-Precision Isothermal Titration Calorimetry with Automated Peak-Shape Analysis. *Anal. Chem.* **2012**, *84* (11), 5066–5073. <https://doi.org/10.1021/ac3007522>.
- (59) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Rance, M.; Skelton, N. J. PREFACE. In *Protein NMR Spectroscopy (Second Edition)*; Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M., Skelton, N. J., Eds.; Academic Press: Burlington, 2007; pp v–vi. <https://doi.org/10.1016/B978-012164491-8/50000-2>.
- (60) Herrmann, T.; Güntert, P.; Wüthrich, K. Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. *J Mol Biol* **2002**, *319* (1), 209–227. [https://doi.org/10.1016/S0022-2836\(02\)00241-3](https://doi.org/10.1016/S0022-2836(02)00241-3).
- (61) Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J. Biomol. NMR* **2013**, *56* (3), 227–241. <https://doi.org/10.1007/s10858-013-9741-y>.
- (62) 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 Cryst* **2009**, *42* (5), 892–900. <https://doi.org/10.1107/S0021889809029288>.
- (63) Svergun, D. I. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J Appl Cryst, J Appl Crystallogr* **1992**, *25* (4), 495–503. <https://doi.org/10.1107/S0021889892001663>.
- (64) Fischer, H.; de Oliveira Neto, M.; Napolitano, H. B.; Polikarpov, I.; Craievich, A. F. Determination of the Molecular Weight of Proteins in Solution from a Single Small-Angle X-Ray Scattering Measurement on a Relative Scale. *J. Appl. Cryst.* **2010**, *43*, 101–109.
- (65) Knobloch, J. E.; Shaklee, P. N. Absolute Molecular Weight Distribution of Low-Molecular-Weight Heparins by Size-Exclusion Chromatography with Multiangle Laser Light Scattering Detection. *Anal. Biochem.* **1997**, *245* (2), 231–241. <https://doi.org/10.1006/abio.1996.9984>.
- (66) Svergun, D. I. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophysical Journal* **1999**, *76* (6), 2879–2886. [https://doi.org/10.1016/S0006-3495\(99\)77443-6](https://doi.org/10.1016/S0006-3495(99)77443-6).
- (67) Kozin, M. B.; Svergun, D. I. Automated Matching of High- and Low-Resolution Structural Models. *J Appl Cryst, J Appl Crystallogr* **2001**, *34* (1), 33–41. <https://doi.org/10.1107/S0021889800014126>.
- (68) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. PRIMUS: A Windows PC-Based System for Small-Angle Scattering Data Analysis. *J Appl Cryst, J Appl Crystallogr* **2003**, *36* (5), 1277–1282. <https://doi.org/10.1107/S0021889803012779>.

- (69) Volkov, V. V.; Svergun, D. I.; IUCr. Uniqueness of ab initio shape determination in small-angle scattering <https://scripts.iucr.org/cgi-bin/paper?ks7125> (accessed May 9, 2019). <https://doi.org/10.1107/S0021889803000268>.
- (70) 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, 2013.

SUPPLEMENTARY INFORMATION

Understanding Intersubunit Interactions in the Enacyloxin Mixed *cis-/trans*-Acyltransferase Polyketide Synthase

Fanny Risser,[†] Sabrina Collin, Raphael Dos Santos-Morais,[‡] Arnaud Gruez,^{*} Benjamin Chagot,^{*} and Kira J. Weissman^{*}

UMR 7365, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA), CNRS-Université de Lorraine, Biopôle de l'Université de Lorraine, Campus Biologie Santé, 9 Avenue de la Forêt de Haye, BP 20199, 54505 Vandœuvre-lès-Nancy, France

[†]Present address: Department of Chemistry, University of Basel, Mattenstrasse 24a, CH-4058 Basel, Switzerland; Department of Biosystems Science and Engineering, ETH-Zürich, Mattenstrasse 26, CH-4058 Basel, Switzerland

[‡]Present address: Université de Lorraine, Laboratoire d'Ingénierie des Biomolécules (LIBio), 2 Avenue de la Forêt de Haye, TSA 40602, 54518, Vandœuvre-lès-Nancy, France

CONTENTS

SUPPLEMENTARY METHODS

Table S1. Primers used in this study.

SUPPLEMENTARY RESULTS

Figure S1. Sequence analysis of the enacyloxin termini for putative docking domain regions.

Figure S2. Bioinformatics analysis of the consecutive interfaces within the enacyloxin PKS, as well as potential DDs at the KirAV/KirAVI interface within the kirromycin PKS.

Table S2. Sequences of constructs investigated in this study.

Figure S3. Evidence for the identity of proteins used in this study.

Figure S4. Characterization of the 5925/5924 interface.

Table S3. Summary of the SAXS data obtained on the DD constructs.

Table S4. NMR and refinement statistics.

Figure S5. Characterization of the 5924/5923 interface.

Figure S6. Characterization of the 5923/5922 interface.

Figure S7. Characterization of the 5922/5921 interface.

Figure S8. Characterization of the 5921/5920 interface.

Figure S9. Characterization of the 5920/5919 interface.

Figure S10. Comparative sequence analysis of the enacyloxin C-termini shown to participate in docking, with ^CDD classes 1a/1b, 2 and 4 α -helix bundle.

Figure S11. Comparative sequence analysis of the enacyloxin N-termini shown to participate in docking, with ^NDD classes 1a/1b, 2 and 4 α -helix bundle.

Figure S12. ITC experiments to detect cross-talk between mis-matched docking domain pairs.

Figure S13. Comparative sequence analysis of the putative kirromycin DDs operating at the KirAV/KirAVI junction.

SUPPLEMENTARY REFERENCES

SUPPLEMENTARY METHODS

Table S1. Primers used in this study.

Construct	Primer type	Oligonucleotides (5'-3') ^a
5925 ^C DD short version	f	TTTCCGGGATCCTACGCCCCGCTCGATACCGAG
	r	TTTCCGAAGCTTTTATTACGGATGAACTCCTTGCC
5925 ^C DD long version	f	TTTCCGGGATCCTACTCCGCCGTGGCGGAG
	r	TTTCCGAAGCTTTTATTACGGATGAACTCCTTGCC
5924 ^N DD	f	TTTCCGGGATCCATGAATAAGCCACCTCGTC
	r	TTTCCGAAGCTTTTCAACCCTGCTCGAGTTG
5924 KS- ^C DD	f	TTTCCGGAGCTCTATCGCCGTGATCGGCATGGC
	r	TTTCCGAAGCTTTTATTTCGGCCTGCAAATCCTCCAGCAGG
5923 ^N DD	f	TTTCCGGGATCCTACATGGGGAGGACCGTGGG
	r	TTTCCGAAGCTTTTCAACGCAGTTGCAGGATCTCC
5923 ^N DD-KR	f	TTTCCGGGATCCATGGGGAGGACCGTGGCAC
	r	TTTCCGAAGCTTTTCAAGCGTCGAGCGGCGG
5923 ^N DD-KR-ACP	f	TTTCCGGGATCCATGGGGAGGACCGTGGCAC
	r	TTTCCGAAGCTTTTCAAGCGTCGCGCCAGGC
5923 ^C DD short version	f	TTTCCGGTTCGACTACGACCACAGTGCCGCCGAACCTCG
	r	TTTCCGAAGCTTTTCAAAGTGCTCCGCGTGATTTCGAGAC
5923 ^C DD long version	f	TTTCCGGTTCGACTACGCGCGCGGTGCGCTGG
	r	TTTCCGAAGCTTTTCAAAGTGCTCCGCGTGATTTCGAGAC
5922 ^N DD	f	TTTCCGGGATCCTACATGACGGACATGGACAAGGACCTGC
	r	TTTCCGAAGCTTTTCAATGGCCTTGCTCGGCCTGC
5922 ^C DD short version	f	TTTCCGGGATCCTACGATGCCGTGACCGATGAAATC
	r	TTTCCGAAGCTTTTCAACGGTCCTCCAGCAGC
5922 ^C DD long version	f	TTTCCGGGATCCTACGGCAGCACCGGTAACGCT
	r	TTTCCGAAGCTTTTCAACGGTCCTCCAGCAGC
5922 ACP- ^C DD	f	TTTCCGGGATCCCGCGAGTCGATGCTCGAC
	r	TTTCCGAAGCTTTTCAACGGTCCTCCAGCAGC
5921 ^N DD	f	TTTCCGGGATCCTACATGGACTCGAAAGACCGCG
	r	TTTCCGAAGCTTTTCACTGGTCCAGCCGCGC
5921 ^N DD-KS	f	TTTCCGGAGCTCGATGGACTCGAAAGACCGCGAGCTGCTG
	r	TTTCCGAAGCTTTTCAAGCCGCGTCGGGCGCC
5921 ^C DD short version	f	TTTCCGGGATCCTACGCCGACGCGCTCGATC
	r	TTTCGCAAGCTTTTACAGCAATTTGTGAGAACGTCCG
5921 ^C DD long version	f	TTTCCGGGATCCTACGCGGGGCCGCGCAG
	r	TTTCCGAAGCTTTTACAGCAATTTGTGAGAACGTCCG
5920 ^N DD short version	f	TTTCCGGGATCCATGAACGACTACAAGGCATTGCTCAAG
	r	TTTCCGAAGCTTTTCACTCGTCGACGCCGGACTC
5920 ^N DD long version	f	TTTCCGGGATCCATGCCGCGCCTGCCGACG
	r	TTTCCGAAGCTTTTCACTCGTCGACGCCGGACTCGAGTTC
5920 ^C DD short version	f	TTTCCGGGATCCTACGATGCCGCCTTGCCGATC
	r	TTTCCGAAGCTTTTATGTCGTCGCCATCTCTTGC
5920 ^C DD long version	f	TTTCCGGGATCCTACGCGGCACAGGCGCGG
	r	TTTCCGAAGCTTTTATGTCGTCGCCATCTCTTGC

5920 ACP- ^C DD	f	TTTCCGGGATCCGCCAGCTGATGCCCGAC
	r	TTTCCGAAGCTTTTAGTCGTCGCCCATCTCTTGC
5919 ^N DD	f	TTTCCGGGATCCT TAC ATGCAAGACATTACAGCAGCTCCTCG
	r	TTTCCGAAGCTTTCACTCGCGGGGGCCTG
5925 ^C DD–5924 ^N DD (G ₃ S) ₃ covalent fusion	f	TTTCCGGGATCCT TAC GGCCCGCTCGATAACCGAG
	r	TTTCCGAAGCTTTACCCCTGCTCGAGTTG
	Overlap 1	GGTGGAGGCAGTGGTGGAGGCAGTGGTGGAGGCAGTATGAATAAGCC CACCTCGTC
	Overlap 2	ACTGCCTCCACCACTGCCTCCACCACTGCCTCCACCTTCACGGATGA ACTCCTTGCC
5925 ^C DD–5924 ^N DD (G ₃ S) ₂ covalent fusion	f	TTTCCGGGATCCT TAC GGCCCGCTCGATAACCGAG
	r	TTTCCGAAGCTTTACCCCTGCTCGAGTTG
	Overlap 1	GGAGGTAGTGGTGGAGGCAGTATGAATAAGCCACCTCGTC
	Overlap 2	GCCTCCACCACTACCTCCCCCTTCACGGATGAACTCCTTGCC
5925 ^C DD–5924 ^N DD direct covalent fusion	f	TTTCCGGGATCCT TAC GGCCCGCTCGATAACCGAG
	r	TTTCCGAAGCTTTACCCCTGCTCGAGTTG
	Overlap 1	TTCATCCGTGAAATGAATAAGCCACCTCGTCCG
	Overlap 2	GGTGGGCTTATTCAATTCACGGATGAACTCCTTGCC
5920 ^C DD–5919 ^N DD (G ₃ S) ₂ covalent fusion	f	TTTCCGGGATCCT TAC GATGCCGCCTTGCCGATC
	r	TTTCCGAAGCTTTCACTCGCGGGGGCCTG
	Overlap 1	GGAGGTAGTGGTGGAGGCAGTATGCAAGACATTACAGCAGCTCC
	Overlap 2	GCCTCCACCACTACCTCCCCCGTCGTCGCCCATCTCTTGC
5924 ^N DD R17A	f	AAGGACGACTACCTGAGCGGCTGTCGCGG
	r	CTTCGACAGCCGCGACAGCGCGCTCAGGTA
5924 ^N DD R17E	f	AAGGACGACTACCTGAGCGAGCTGTCGCGG
	r	CTTCGACAGCCGCGACAGCTCGCTCAGGTA
5924 ^N DD K32A	f	CTGATGGCGCTCGCGCTCGCGCTCAAGCAG
	r	GAGTTGCTGCTGCTTGAGCGCGAGCGCGAG
5924 ^N DD K32E	f	CTGATGGCGCTCGCGCTCGAGCTCAAGCAG
	r	GAGTTGCTGCTGCTTGAGCTCGAGCGCGAG
5924 ^N DD Q36A	f	GCGCTCAAGCTCAAGCAGGCGCAACTCGAG
	r	TCACCCCTGCTCGAGTTGCGCCTGCTTGAG
5924 ^N DD Q36E	f	GCGCTCAAGCTCAAGCAGGAGCAACTCGAG
	r	TCACCCCTGCTCGAGTTGCTCCTGCTTGAG

^aIntroduced restriction sites are underlined, while the base pairs coding for an additional Tyr are indicated in bold.

SUPPLEMENTARY RESULTS

A

```

5922_Cter 1 -----APLPRAGITLD-IDWARFRPIYRHLDALFAELGTP-----ADGCA
5925_Cter 1 -----VVHWRLD-----ALARVPDPALE-----RDLVTATSVATPA-----T-AQA
5920_Cter 1 LRRVWLAGWPAGRRMRCLV-----RHLPEAGAAGELVYDLALIDEQERFAIT HARFRRAALLVEDEAVPAAGQD-----E-AARA
VirA ACP5b 1 -----
VirA ACP5a 1 -----
5923_Cter 1 -----LASGAGQEAACVFDIDWRRHAATYAP-----PLWAG-----LLGD
5921_Cter 1 -----NWPGWGGVGMARNARGEPPGSLRRLAPERALGELGRALAGGAQVADIDWSVFGRDWRAS-----AAAVPALVEDWFAAHQAAPRQALAA

                                     End ACP↓

5922_Cter 44 GAKAAGDAAAFRRAYA--AAGYRESIDELLALREVLGSGNFAAYAGTGFDLGLDSLITLSAEKLGARVGPVSSVDVFDNANPARLRGWLAIRL
5925_Cter 39 GP-----ASYRQRLADALPAERAAARRLVAEQIALTGT-AAATTEPAAPLSILGDSLKSALSDALAHCLGAAASATLFDHPTLDALHLALAN
5920_Cter 78 AAPAEPALPDEFAQLMPDAQDVARLVREILVRELKIDAAASDERPFELGDSLSALEPSDELGACALDHVDITFDYPSVASLSAYLLERL
VirA ACP5b 1 -----PVARALREELARTLYC--PDDDEASNTLGDLSLGVFVAFVNOTGDEKAGLYDHPSLAALSRLHARLAR
VirA ACP5a 1 -----EIAEEARLAGLY--PDRDPEETLTLGDLSLGVFVAAVNAAPVGKATALYDHPAAAFARHAESE
5923_Cter 38 GGAPAEPPSFAERLAEVPPERRRARRARLEAAACGRARAAATDTDGAEAGDSLHTVLHRQLERECAAATTAFDHTVAADCLARGA
5921_Cter 91 SAVPAEAAREL--AAAVDVDALEARRHLVGVRRIMALDAARPAAQNKSCHLGLDSLHATELKRALQEGARAPATMTFDYFDLSLAHWLAAPA

5922_Cter 142 KALYA-AAPAA-A-STGNAGATGATSAFSAADATHDPDATDAPPPAGPLAATVSPTASDAAGDAVTDERELEQTMQALLEDNR---
5925_Cter 132 APAGSAVAEASVPAVEA-T-----EPAAAPLDTELSE-----IEGLQDDLAALLGKEFIRE-----
5920_Cter 177 AAQAQAOAPP--EPAGQAAP-----ADAALPID-----ELSA LRQEMGDD-----
VirA ACP5b -----
VirA ACP5a 75 -----
5923_Cter 136 LAELF--APAIVAIPAQLANA-----AADA-----S-----LGDHSAEELARL LAHELGGLESRGAL
5921_Cter 188 QAA-R--APASASITPRA-----APAD-----A-----LDQLDEGEADLDKLL-----

```

B

```

DEBS_KS5_Nter 1 -----PAAIVGHACRFPGDVDPSTMEFTSGGCDTIAAPADRG-----EPD
5921_Nter 1 -----DSKDELLARQDARQGIQQAISA--RLDQPAIVGIGCRVPGD-SPALMELLRDGREATAEPPGRMDDAYVDAT
5920_Nter 1 -----MNDYKALLKSKIRKHQEQDRNRELESQV--DEPVAITGSCRFPGP-DA-AFMRAENGADITTTMTQRMEMEAHTDA
5919_Nter 1 -----MQDIQQLAKSITELRLKAANDALEQNR--REPAAIVGAICRPPGGIGSLQLMTALEAGRGTRTVGGERPQORLTDD
5924_Nter 1 MNKPTSSDGWKDDYLSRSRLSKNQDMA-----LALKLQQQLEQG-PAAEPAAIVGVCRLPGGVAGPDYALLRSAGSGIVENQDQRWNAAYEDAD
5922_Nter 1 -----TMDKDLLOSQTIRELKTAAQAEGQHHEAVAAIVGSLRFPGGVTDLSWALLREERSGITIEEPKWSNRQVDPD

DEBS_KS5_Nter 45 P-----DARLGGIAAAGDFDAFFGISPREALMDPQQRLLLEISWEALERAGHPVSLRGSATGVFTGVTVDYGRP-----DEAPTEVLGVY
5921_Nter 81 P-----GTPYTYARRAGVLEVDHDFARFFGISPREAQMDPQORLLLEVSHRALEDALPYTALREQPVGVFVGISSGYYVTFDKA--RSASQDAIS
5920_Nter 80 ASAEAEEAGITYTRRFGLLEIDGFEPCAFGISEEAPYDPOHLLLEQAMFCEHAGLDAKTVKSGDYGVFVGOMNDYARLII-----RSLDLPNVY
5919_Nter 81 P-----HRPGGIYDAMGLLEAIDGFDHFGIRHDEATHDPOHLLLEVAAEFEDAGYADAFSGSRTGVVGMNDYDGLQ-----GPLEAASLYI
5924_Nter 95 P-----EAGGIHTRSLGLVDEDRFDADFSTISPREAESMDPQORLLLEVAAEFERSGHACALDGRQVGVFVGMNDYHLNAPDITTEARHSPYY
5922_Nter 84 Y-----AAGKLVTPYGLLEHIYDFDAEFFGSALEAENMDPQORLLLEQSMALADAGYDGRLRGSDTGVVVGIGSQDYGMA-----LLDPAHANPVY

DEBS_KS5_Nter 131 GIGASSVAGRAYCLGLEPATTDTACSSGLTIALHANEESLRDECGALAGGVITSSPGAFTEFRSQGLAADGRCPFSKAADGGLTEGAGVY
5921_Nter 175 IIGSMNSAAGRLAYHYGFNGPALATDTACSSSLVALHQAARSLNNECHTALAGGVNCLPEPITALAQNVLSAGGRCSPFSAEADGVVRAEGCGVL
5920_Nter 175 GAGAPSAAGRLSYVFGKGPSTIDTACSSSLVAVHLASQSLRUGCCALAGGVNLLSPETAACACVARMLSARGCNTFGLEADGVVRAEGCGVL
5919_Nter 172 GSGIAKSCAAGRLAYITFGLGPIALATDTACSSSLVAVHLAQALRRGECDAALAGGVNLLSPQGTVACRSQMLSPAGRCRTFASADGVVRAEGCGVL
5924_Nter 191 ASGEAFSSAAGRLAYITFGLGPIALATDTACSSSLVAVHLACRSLECECALAGGTSLLSPEASTVSSNARMSPGGQWCFDHRADGVVRAEGCGVL
5922_Nter 176 ASGNLSLAAGRLSYVDFSGPLTIDTACSSSLVAVHACRRLQLGECGLALAGVNAHLPHAGINFSRARMSTERDCTFDARAGVYVRAEGCGVL

DEBS_KS5_Nter 231 VLKRLSARRRGRLAVIRGSVAVNODGASGLTAPSGPAQQRVIRALENAGVAGVVDYVEAHGTGTRLGDPTEVHAALSTYGAERIPDDPLMIGSVK
5921_Nter 275 VLKRLDALAQGRILAVIRGSVAVNODGASSGLTPNGYAQAALATALLRARAPAGVYVEAHGTGTRLGDPTEVHAALQALGAREARPLVIGAVK
5920_Nter 275 LKTLSSRARADGDTLAVIRGSVAVNODGRSHGLAPNGPAQGVIRDALARARIDPAEVGYETHGTGTLGDPTEVHAALQALGAREARPLVIGAVK
5919_Nter 272 LKRLSDAERDGRILAVIRGSVAVNODGRSHGLTAPSGPAQQRVIRALENAGVAGVVDYVEAHGTGTRLGDPTEVHAALQALGAREARPLVIGAVK
5924_Nter 291 VLKRLSRLADGDTLAVIRGSVAVNODGRSHGLTAPNTAQAALREALGAKIDARIRYVEAHGTGTRLGDPTEVHAALQALGAREARPLVIGAVK
5922_Nter 276 VLKRLADAQADGRILAVIRGSVAVNODGSSGLTPNGSAQRAVIRALENAGVAGVVDYVEAHGTGTRLGDPTEVHAALQALGAREARPLVIGAVK

DEBS_KS5_Nter 331 SNIGHTCAAGVAGVMAKVALRHGEIPRTIHFDEPSQITWDLGAVSVVS
5921_Nter 375 AHIGHLEAAGVAGIKTVLALRHRLPAQNLGTPTHVDMSSG-----
5920_Nter 375 ANIGHESAAGTAGLIKLVQLRHDSLPVVALDALNPHFGLSDQ-----
5919_Nter 372 SNIGHESAAGTAGLIKLVQVRRRRPRNHFITLNPQIRVDLERLRI--
5924_Nter 391 TIGHTEACAGVAGLIKLVCAHDRVPO-----
5922_Nter 376 ANIGHLEAAGTAGLIKVALRHGEAPPQPGFETLNPATGWDV-----

```

Figure S1. Sequence analysis of the enacyloxin termini for putative docking domain regions. (A) Analysis of the C-termini (C-ter) of subunits 5925 and 5923–5920 which contain acyl carrier protein (ACP) domains, by comparison to two ACPs from the virginiamycin *trans*-AT PKS (ACPs 5a and 5b of subunit VirA) for which NMR structures are available.¹ This analysis identifies a conserved L/V at the end of the ACPs, and so the downstream sequence was analyzed for potential DD regions. (B) Analysis of the N-termini (N-ter) of subunits 5924 and 5922–5919 which incorporate ketosynthase (KS) domains, by comparison with the sequence of KS₅ from the erythromycin (DEBS) system, for which a crystal structure is available.² The highly-conserved KS start site is readily identifiable, and thus the putative DD-containing regions of the subunits.

C

5924_module 3	1	DIAIVGMACRPPGANDVGAFWQDISGDMIRPFD-GTRWDVPRFYTPGSTEDCKMVANDGGQIADVHGFDNRFPGIGDREAEMDPQQRITALEVAMET
DEBSS_KS5_2HG4	1	PIAIVGMACRPPGQVDSPESFWEFVSGGGDAAEAPADRGWEP-----DPD-----ARLGGMLAAAGDFDAGFFGISPREALAMPQQRITILEISWEA
DEBSS_KS3_2Q03	1	PIAIVSMACRPPGGVNIQRLWELLREGGETLSGFPTDRGWDLARLHHPDPNPCTSYVDKGGFLDDAAGFDAEFFGNSPREAAAMPDQQRILLETSWEL
5924_module 3	100	LESANYTPEQLAD-BAQVFPGCPSPDEADLSORHAGALVGLIGPCHHVSAPRTAHLFDWQGPCVAIDTACSSSLVAHVAAQHLRERECRVALAGGVN
DEBSS_KS5_2HG4	89	LERAGHDVSLRGATGVETGVGTVDYGPDPDEADVIGYGTGTASSVASGRITAYCGLGEPAMIVDTACSSGLTALHLAESLRDECGVALAGGVN
DEBSS_KS3_2Q03	101	VENAGIDPHSLRGATGVETGVKFGYGDIA-AAEDIVEGVSYTGMAPAVASGRISYTHGLEGPSTSVDTACSSSLVALHLAESLRDEGSSVALVGGAA
5924_module 3	199	VILSPANNVLSKAGMLSPAGRCITFDVGADGVRSIGCGVLLKRLDIALAGDAILEVIRGSAVNHNGRGGLTAPSSRQQRITAEALARACTLPSE
DEBSS_KS5_2HG4	189	VMSSPGAFTEFRSQGLAADGRCKPFSKADGFGLAEGAGVLYLQRLSARREGPVLAIVRGSANQDGASNGLTAPSGPAQQRVIRRALENAGVRAGD
DEBSS_KS3_2Q03	200	VMAIPGVFVDFSQRLAADGRSKAFGAGADGFGFSEGVTVLLERLSARRNGLEVLAVRGSANQDGASNGLSAPSGPAQQRVIRRALENAGVRAGD
5924_module 3	299	IRYVEAHGTGTPLGDPTEAALKATYGAHRDAADPLVYGAVKSAIGHTESAAGVAGLTKVLLMRRHMPPTLHNTLNPHLEIDPRTIRPTAPQPLLA
DEBSS_KS5_2HG4	289	VDYVEAHGTGTRLDGPIEHALLSTYGAERDDPLWIGSVKSNIGHTQAAAGVAGVMKAVLALRHGENPRTLHFEPSQTEWDLGAVSVQARSMPA
DEBSS_KS3_2Q03	300	VDAVEAHGTGTRLDGPIEANALLDTYCRDRDAPLWIGSVKSNIGHTQAAAGVTGLLKVVLALRNGELPATLHVEEPIPHVDWSSGVALAGNQPRRR
End conserved KS↓		
5924_module 3	399	REIGTLSCAVSSFCFSGTNAHLIVAAPPGKPARPLARSGRGLFAVSARSRLPALARLCERHAVHLARAGTAEPDLADLCSTLLGRRRFEHVLCLYPDSHAE
DEBSS_KS5_2HG4	389	-GERPRRAIVSSFGTSGTNAHVIVEEA-----
DEBSS_KS3_2Q03	400	-GERTRRAAVSAFGISGTNAHVIVEE-----
5924_module 3	499	LTIAQLRASAARLRQAPAPAAPAIDTLALRLVAGAALPAATLAGWHDEPRFAAALAARDALSAQAQEPAGEGAPASLDAAWFCVLHALSHCMAAFVGE
DEBSS_KS5_2HG4		-----
DEBSS_KS3_2Q03		-----
5924_module 3	599	PDRIDYRGRLWLAAPIHQAGSLDEAARRFLAADPAPAPKRLGGIALRPADQCEGDEGEGGFLMLRDASGRATRHLDAAGLDEAWRRAFGALWEGGR
DEBSS_KS5_2HG4		-----
DEBSS_KS3_2Q03		-----
5924_module 3	699	VDWLAGFAGSAYRRVALPAYPFEHRDCSRPARLPAGERSLELLLEDLQAE
DEBSS_KS5_2HG4		-----
DEBSS_KS3_2Q03		-----

D

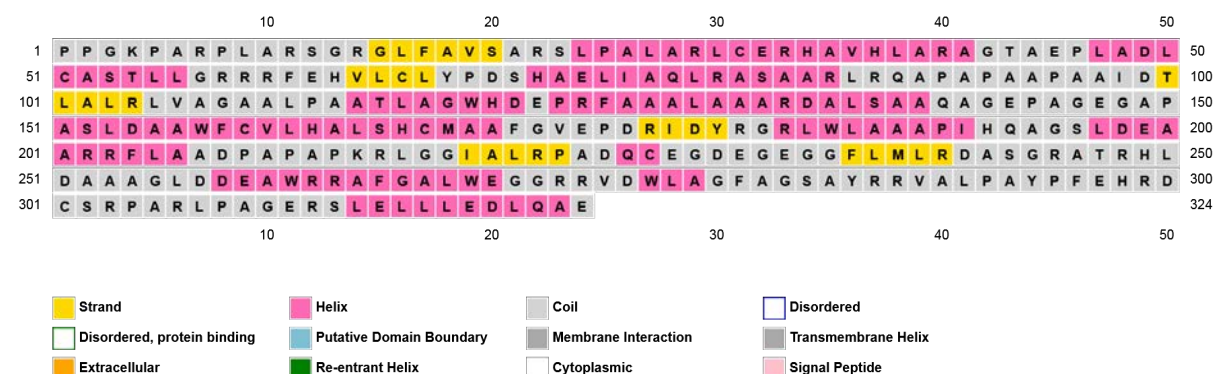


Figure S1, cont. (C) Comparison of the 5924 module 3 sequence (in which a KS had previously been identified³), and the sequences of representative *cis*-AT PKS KSs (DEBSSs KS₅ (ref. ²) and KS₃ (ref. ⁴) (the respective PDB entries are indicated). This analysis shows that some 320 residues (highlighted in green) lie downstream of the well-conserved KS domain. We therefore focused on this region for secondary structure prediction, to determine if we could discern a putative ^cDD-containing region. (D) Analysis using PSIPRED⁵ of the region shown in green in (C). This analysis identifies a large gap between the secondary structure elements of an as-yet unidentified putative domain (ending in ...WLA) and a last α -helix. We therefore investigated this final α -helix as a potential ^cDD.

E

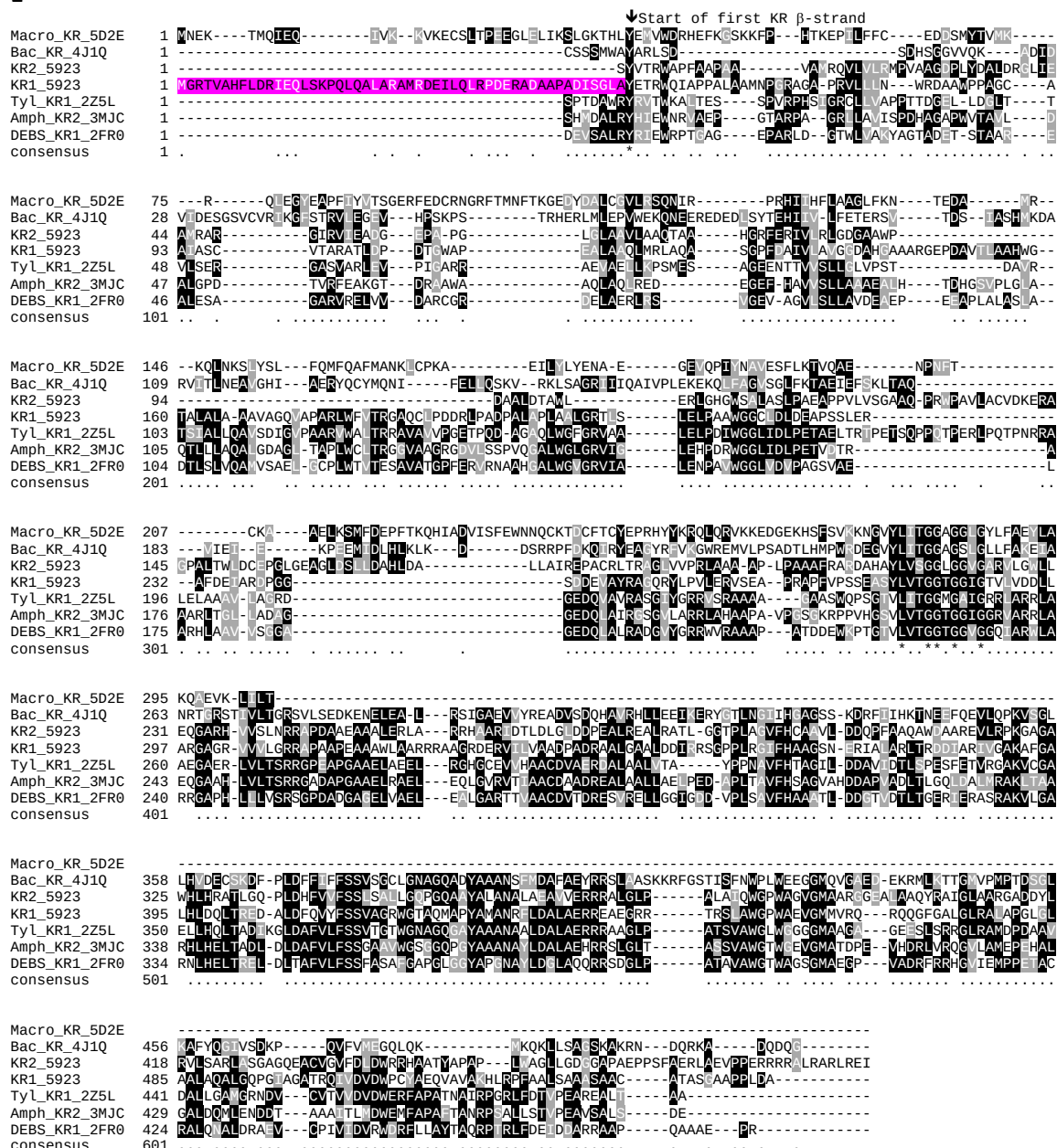
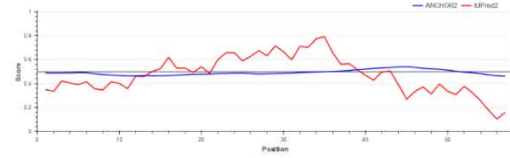


Figure S1, cont. (E) Alignment of the N-terminus of subunit 5923 and the second ketoreductase (KR) from the same subunit (from module 4, **Figure 1**), with the sequences of multiple PKS KR domains from both *cis*-AT and *trans*-AT PKSs whose crystal structures are available. Key: Macro, macrolactin; Bac, bacillaene;⁶ Tyl, tylosin;⁷ Amph, amphotericin;⁸ DEBS, erythromycin⁹ (the corresponding PDB files used are indicated). This comparative analysis clearly identifies the conserved start of the first KR secondary structure element, and therefore the region containing a putative ^NDD in subunit 5923 (highlighted in pink).

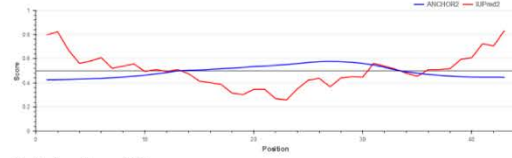
Secondary structure prediction (PSIPRED)

Conf: 977687615788764203100104877787530134430788004899861654519
Pred: CCCCCCCC#####CCHCCCCCCCCCCCC#####C#####C
AA: LAANAPAGSAVAEAAASVAPAVEATEPAA¹⁰APLD³⁰TELSE⁴⁰TEI⁵⁰GLQDD⁶⁰DLAALL⁷⁰KE⁸⁰FIRE

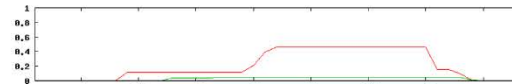
5924 ^NDD

Secondary structure prediction

Conf: 9888888714789888865328789999999988976536789
Pred: CCCCCCCCXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCCC
AA: NKPTSSDGWKDDYLSRLSRLSKNQLMALALKLKQQQLQGPA
10 20 30 40

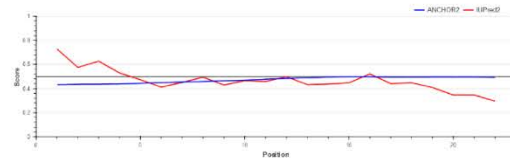


Coiled-coil prediction

5924 ^cDD

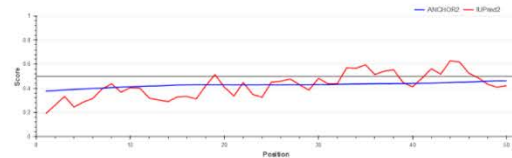
Secondary structure prediction

Conf: 8854541140499999887409
Pred: CCCCCCCC HHHHHHHHHHHH CC
AA: RPARLPAGERSLELLLEDLQAE
10 20

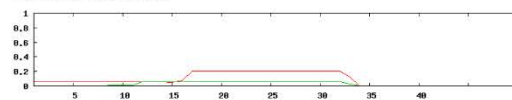
5923 ^NDD

Secondary structure prediction

Conf: 515788899887419699999999998975320988
Pred: CHHHHHHHHHHHHHHHCHHHHHHHHHHHHHHHHHCCCC
AA: GRTVAHF^LDRIEQLSKPQLQALARAMRDEILQLRPD
10 20 30

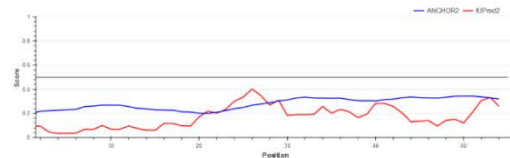


Coiled-coil prediction

5923 ^cDD

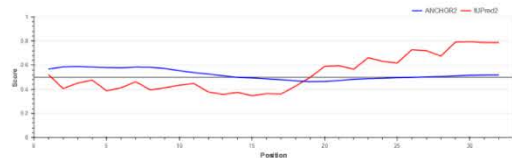
Secondary structure prediction

Conf: 91467777677776421788767764024865299999999984683113689
Pred: CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHCCCCCCCCC
AA: ARGALAEI.FAPATVAAPAQLANAAADASI.GDHSAAELARILAHETGGLESRGAL

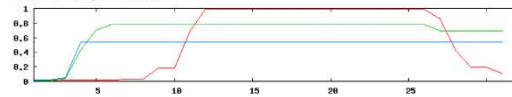
5922 ^NDD

Secondary structure prediction

Conf: 60259999999999999999999897311149
Pred: CCHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCC
AA: TDMDKDLLLSIQITRELKTRLAQAEQGHHEA
 10 20 30



Coiled-coil prediction

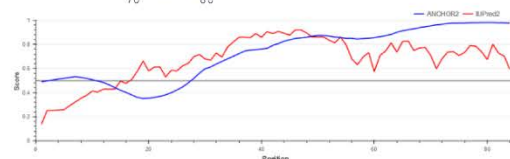
5922 ^cDD

Secondary structure prediction

```

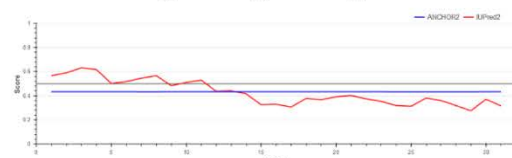
Conf: 9567789973432447778888765221012346799998999999842010788433
Pred: C HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: ARLKALYAAAAPAAAGTGNGATGTAATAACATHDPATDAPPFGLVLTAFSTASD
                20      40      60
Conf: 1356103999999999999875229
Pred: CCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: AGAGDVTDTEIRELQTMALLERD
                70      80

```



5921 NDD

Secondary structure prediction

[illegible]

Coiled-coil prediction

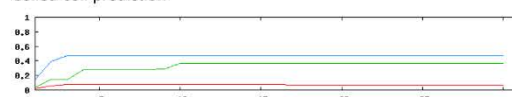


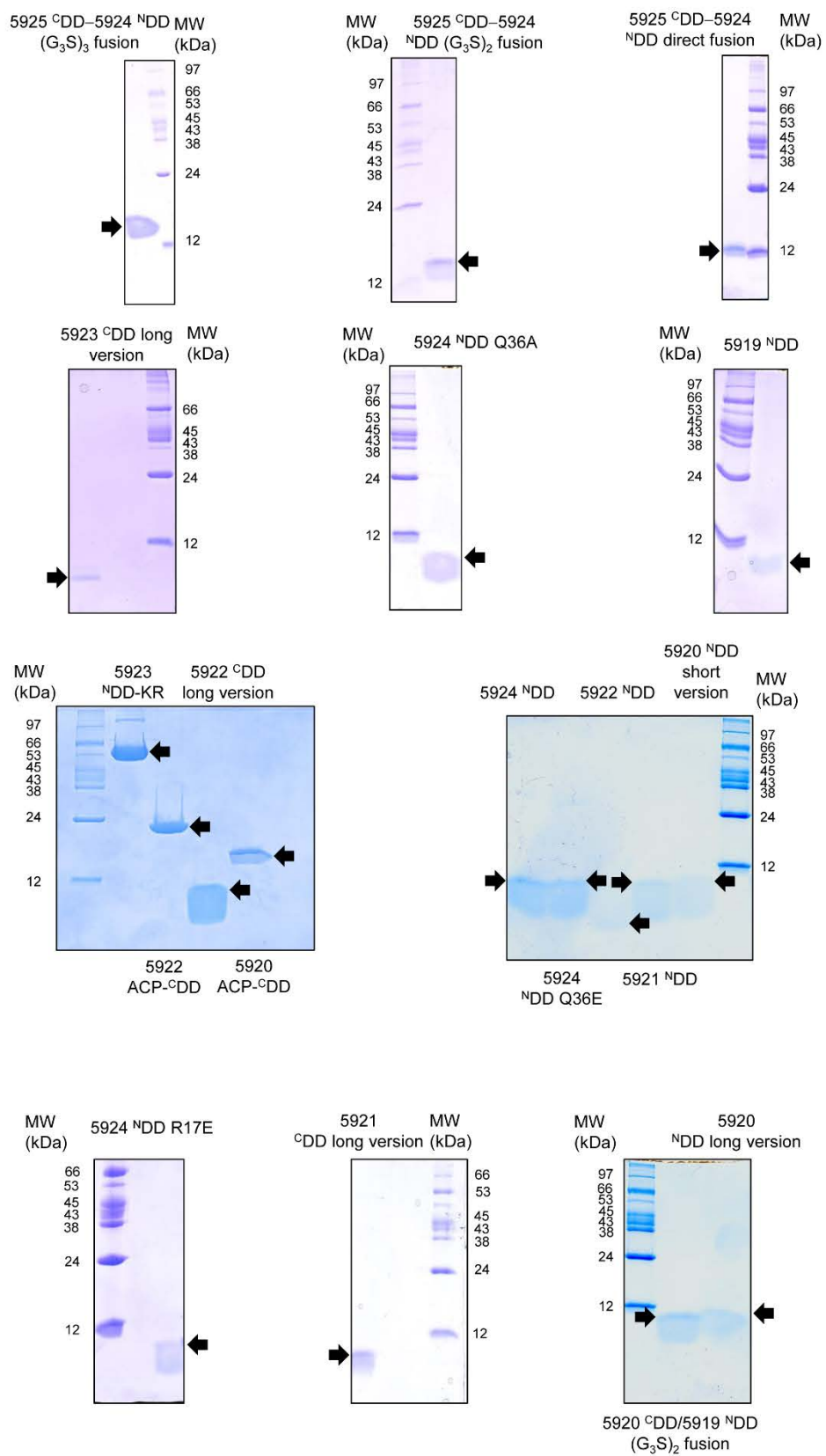
Table S2. Sequences of constructs investigated in this study.

Construct	Sequence ^a
5925 ^C DD short version	GP G S Y APLDTELSEIEGLQDDDLAALLGKEFIRE
5925 ^C DD long version	GP G S Y SAVAEAAASVAPAVEATEPAAAPLDTELSEIEGLQDDDLAALLGKEFIRE
5924 ^N DD	GP G S M N K P T S S D G W K D D Y L S R L S R L S K N Q L M A L K L K Q Q L E Q G
5924 ^C DD	GP G S Y RPARLPAGERSLELLLEDLQAE
5924 KS- ^C DD	GP G S P N S S I A V I G M A C R M P A G A N D V G A F W D Q L I S G T M V R P D G T R W D V P R F Y T P G S T E D G K M V A N D G G Q I A D V H G F D N R F F G I G D R E A E M D P Q R I A L E V A W E T L E S A A Y T P E Q L A D G A G V F I G P G P S D F A D L S Q R H A G A L V G L M G P G H V S A I P G R I A H L F D W Q G P C M A I D T A C S S L V A V H V A A Q H L R E R E C R V A L A G G V N V I L S P A N N I V L S K A G M L S P A G R C R T F D V G A D G Y V R S D G C G M V L K R L D D A L A D G D A I L G V I R G S A V N H N G R G Q L T A P S S R Q Q A R L I E A A L A R A G T L P S E I R Y E A H G T G T P L G D P I E M A A L K A T Y G A H R D A A D P L Y V G A V K S A I G H T E S A A G V A G L I K V L L M M R H R M I P T L H L N T L N P H L E I D P R T I R I P T A P Q L L A R E D G T L S C A V S S F G F S G T N A H L I V A A P P G K P A R P L A R S G R G L F A V S A R S L P A L A R L C E R H A V H L A R A G T A E L A D L C A S T L L G R R R F E H V L C L Y P D S H A E L I A Q L R A S A A R L Q A P A P A A I D T L A L R L V A G A L P A A T L A G W H D E P R F A A A L A A R D A L S A A Q A G E P A G E G A P A S L D A A W F C V L H A L S H C M A A F G V E P D R I D Y R G R L W L A A A P I H Q A G S L D E A A R R F L A A D P A P A K R L G G I A L R P A D Q C E G D E G E G F L M R D A S G R A T R H L D A A A G L D D E A W R R A F G A L W E G G R R V D W L A G F A G S A Y R R V A L P A Y P F E H R D C S R P A R L P A G E R S L E L L L E D L Q A E
5923 ^N DD	GP G S Y MGRTVAHFLDRIEQLSKPQLQALARAMRDEILQLR
5923 ^N DD-KR	GP G S M G R T V A H F L D R I E Q L S K P Q L Q A L A R A M R D E I L Q L R P D E R A D A A P A D I S G L A Y E T R W Q I A P P A L A A M N P G R A G A P R V L L L N W R D A A W P P A G C A A I A S C V T A R A T L D P D T G W A P E A L A A Q L M R L A Q A S G P F D A I V L A V G G D A H G A A A R G E P D A V T L A A H W G T A L A L A A V A G Q V A P A R L W F V T R G A Q C L P D D R L P A D P A L A P L A A L G R T L S L E L P A A W G C L D L D E A P S S L E R A F D E I A R D P G G S D D E V A Y R A G Q R Y L P V L E R V S E A P R A P F V P S S E A S Y L V T G G T G G I G T V L V D D L L A R G A G R V V L G R R A P A A P E A A A W L A A R R R A A G R D E R V I L V A A D P A D R A A L G A L D D I R R S G P L R G I F H A A G S N E R I A L A R L T R D D I A R I V G A K A F G A L H D Q L T R E D A L D F Q Y F S S V A G R W G T A Q M A P Y A M A N R F L D A L A E R R E A E G R R T R S L A W G P W A E V G M M V R Q R Q G F G A L G L R A L A P G L G L A A L A Q A L G Q P G I A G A T R Q I V D V D W P C Y A E Q V A V A K H L R P F A A L S A A A S A C A T A S G A A P L D A
5923 ^N DD-KR-ACP	GP G S M G R T V A H F L D R I E Q L S K P Q L Q A L A R A M R D E I L Q L R P D E R A D A A P A D I S G L A Y E T R W Q I A P P A L A A M N P G R A G A P R V L L L N W R D A A W P P A G C A A I A S C V T A R A T L D P D T G W A P E A L A A Q L M R L A Q A S G P F D A I V L A V G G D A H G A A A R G E P D A V T L A A H W G T A L A L A A V A G Q V A P A R L W F V T R G A Q C L P D D R L P A D P A L A P L A A L G R T L S L E L P A A W G C L D L D E A P S S L E R A F D E I A R D P G G S D D E V A Y R A G Q R Y L P V L E R V S E A P R A P F V P S S E A S Y L V T G G T G G I G T V L V D D L L A R G A G R V V L G R R A P A A P E A A A W L A A R R R A A G R D E R V I L V A A D P A D R A A L G A L D D I R R S G P L R G I F H A A G S N E R I A L A R L T R D D I A R I V G A K A F G A L H D Q L T R E D A L D F Q Y F S S V A G R W G T A Q M A P Y A M A N R F L D A L A E R R E A E G R R T R S L A W G P W A E V G M M V R Q R Q G F G A L G L R A L A P G L G L A A L A Q A L G Q P G I A G A T R Q I V D V D W P C Y A E Q V A V A K H L R P F A A L S A A A S A C A T A S G A A P L D A V P S A A E D F G Q A G A T L A L R E L V A E L T G A A L P E R G E A R P M Q E L G L T S L L S I E L S Q K L R Q R L G V P C R P T V V F D H A N L R A L A E S L A Q A W A H A
5923 ^C DD short version	GP G S P N S S S V D Y DHSAELARILAHELGGLESRGAL
5923 ^C DD long version	GP G S P N S S S V D Y ARGALAELEFAPAIVAAPQLANAAADASLGDHSAELARILAHELGGLESRGAL
5922 ^N DD	GP G S Y MTDMDKDLLLSIQITIRELKTRLAQAEQGHHEA
5922 ^C DD short version	GP G S Y DAVTDEIERELQTMQALLEDR
5922 ^C DD long version	GP G S Y GSTGNAGATGATSAFSAADATHPDATDAPPPAGPLAATVSPTASDAAGDAVTDEIERELQTMQALLEDR
5922 ACP- ^C DD	GP G S R E S M L D E L L H A L L R E V L G L S G N F A A Y A G T G F H D L G M D S L L T L S F A E K L G A R V G L P V S S V D V F D N A N P A R L R G W L A A R L K A L Y A A A P A A A G S T G N A G A T G A T S A F S A A D A T H P D A T D A P P P A G P L A A T V S P T A S D A A G D A V T D E I E R E L Q T M Q A L L E D R
5921 ^N DD	GP G S Y MDSKRELLARQRDAIRQGIQIQALSARLDQ

5921 ^N DD-KS	GP GSP NS S MDSKDRELLARQRDAIRQGIQKIQALSARLDQPVAIVGIGCRVPGADSPEALWELL RDGREALAEVPPGRWDLDAYYDATPGTPYKYARRAGYLDEVHFDARFFGISPREAQRMDPQQ RLLLEVSHRALEDAELPVTALREQPVGVFVGISSGEYAVMTFDKARSDSQDAWSITGTSMNSAA GRLAYHYGFNGPALAIDTACSSSLVAIHQAVRSLNNEECHTALAGGVNCLLTPEPSIALAQNKV LSASGRCSPPFSAEADGLVRGEGCGMLVLKRLDDALAQQCRILAVIRGSHVNQDGASSGLTVPNG YAQQALIATALKRARLAPGAIGYVEAHGTGTALGDPIEIKALQQALGAGREAGRPVLIGALKAH IGHLEAASGVAGVIKTVLALRHRLPAQINLGTPTPHVDWSSGGVAVVSESTPIAYGPDAPFYA GVSSFGFSGTNAHLILQDPVSAVPAEAAAGAAATQAVARFVGLSAKDPAAALRELLARCHAYFR EVPDWAAACEALNAGRAHYAHRAGFVARDRETLLLEQLAAAAAAEALEADAPAPDAA
5921 ^C DD short version	GP G S Y ADALDQLDEGELADVLDKLL
5921 ^C DD long version	GP G S Y GPAQAARAPASASAPTPRAAPADALDQLDEGELADVLDKLL
5920 ^N DD short version	GP G S L NDYKALLKSSIRKIQEQDRRIRELESGVDE
5920 ^N DD long version	GP G S M PRLPTRSINWTKANWPTFSTNCCEPELNDYKALLKSSIRKIQEQDRRIRELESGVDE
5920 ^C DD short version	GP G S Y DAALPIDELSALLRQEMGDD
5920 ^C DD long version	GP G S Y AAQARQAPPGPAGQAAPADAALPIDELSALLRQEMGDD
5920 ACP- ^C DD	GP G S A QLMPDAKQDLVARLVRELLVRFKIDAGAVSDERPFELGMDSVSALEFSDELGACFAL DLHVDITFDYPSVASLSAYLLERLAAAQARQAPPGPAGQAAPADAALPIDELSALLRQEMGDD
5919 ^N DD	GP G S Y MQDIQQLLAKSLTEIKRLKAANQALEQARRE
5925 ^C DD–5924 ^N DD (G ₃ S) ₃ covalent fusion	GP G S Y APLDTELSEIEGLQDDDLAALLGKEFIRE GGGSGGGSGGG SMNKPTSSDGWKDDYLSRL SRLSKNQLMALALKLKQQQLEQG
5925 ^C DD–5924 ^N DD (G ₃ S) ₂ covalent fusion	GP G S Y APLDTELSEIEGLQDDDLAALLGKEFIRE GGGSGGG SMNKPTSSDGWKDDYLSRLSRLS KNQLMALALKLKQQQLEQG
5925 ^C DD–5924 ^N DD covalent fusion	GP G S Y APLDTELSEIEGLQDDDLAALLGKEFIREMNKPTSSDGWKDDYLSRLSRLSKNQLMALA LKLKQQQLEQG
5920 ^C DD–5919 ^N DD (G ₃ S) ₂ covalent fusion	GP G S Y DAALPIDELSALLRQEMGDD GGGSGGG SMQDIQQLLAKSLTEIKRLKAANQALEQARRE
5924 ^N DD R17A	GP G S M NKPTSSDGWKDDYLS A LSRLSKNQLMALALKLKQQQLEQG
5924 ^N DD R17E	GP G S M NKPTSSDGWKDDYLS E LSRLSKNQLMALALKLKQQQLEQG
5924 ^N DD K32A	GP G S M NKPTSSDGWKDDYLS E LSRLSKNQLMALAL A LKQQQLEQG
5924 ^N DD K32E	GP G S M NKPTSSDGWKDDYLS E LSRLSKNQLMALAL E LKQQQLEQG
5924 ^N DD Q36A	GP G S M NKPTSSDGWKDDYLS E LSRLSKNQLMALALEL KQ AQLEQG
5924 ^N DD Q36E	GP G S M NKPTSSDGWKDDYLS E LSRLSKNQLMALALEL KQ E QLEQG

^a**Red** = sequence introduced from the plasmid; **green** = added Tyr; **blue** = linker used to fuse the docking domains; **purple** = position of introduced mutation.

A



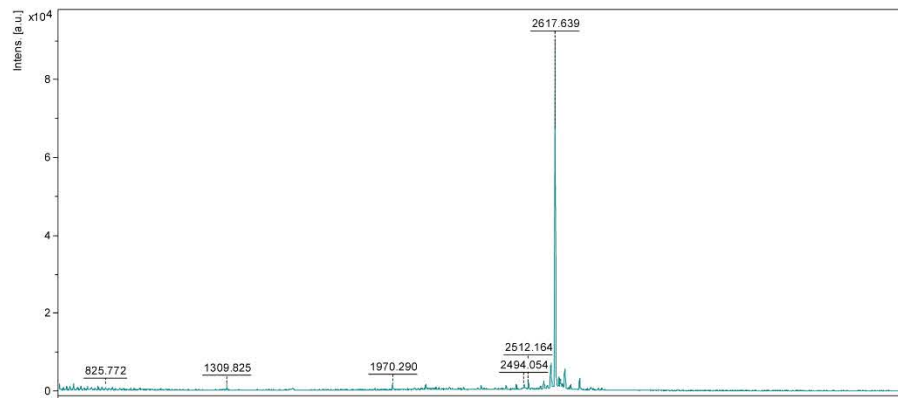
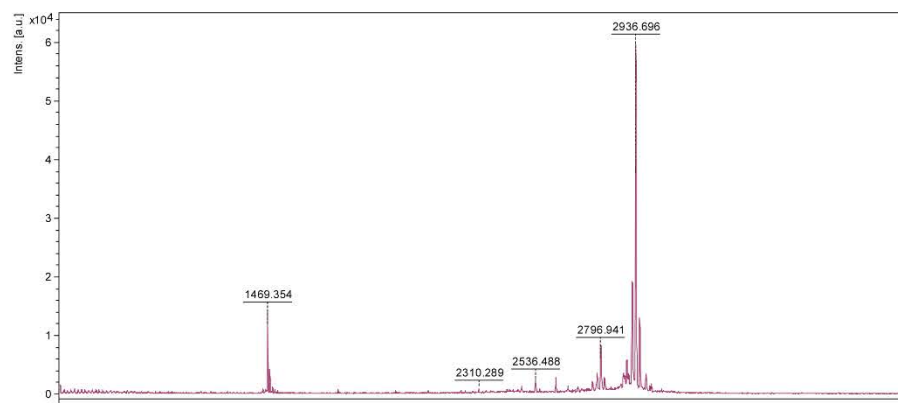
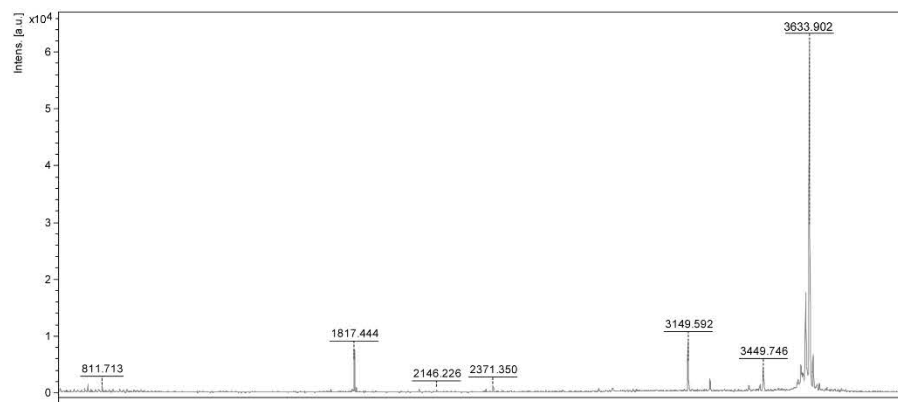
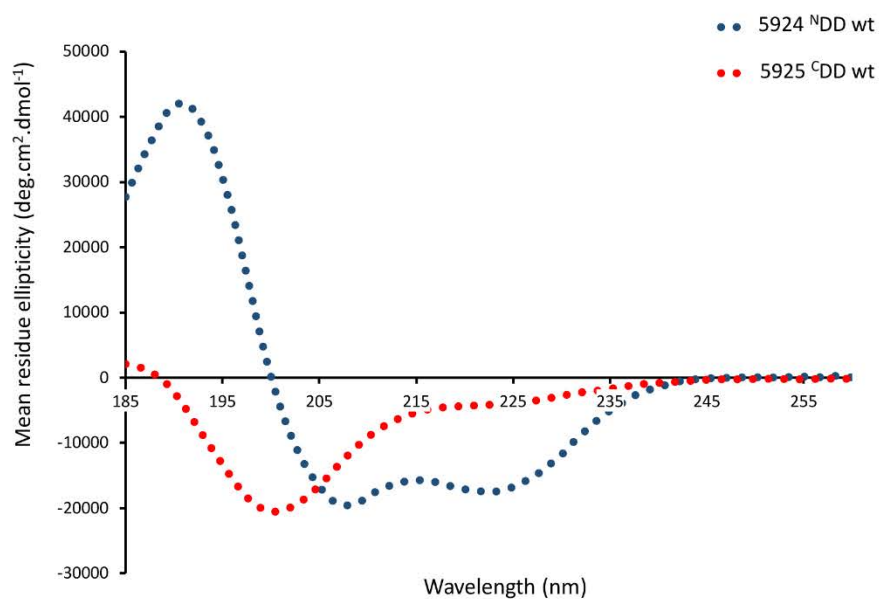
B**C****D**

Figure S3. Evidence for the identity and purity of proteins used in this study. (A) SDS-PAGE analysis of constructs used in this study (arrows indicate the corresponding bands): 5925 ^CDD–5924 ^NDD fusion (G_3S_3) (calc'd: 9.14 kDa); 5925 ^CDD–5924 ^NDD fusion (G_3S_2) (calc'd: 8.88 kDa); 5925 ^CDD–5924 ^NDD direct fusion (calc'd: 8.37 kDa); 5924 ^NDD (calc'd: 5.02 kDa); 5924 ^NDD R17E (calc'd: 4.99 kDa); 5924 ^NDD Q36A (calc'd: 4.94 kDa); 5924 ^NDD Q36E (calc'd: 4.99 kDa); 5923 ^CDD long version (calc'd: 6.41 kDa); 5923 ^NDD-KR (calc'd: 57.4 kDa); 5922 ^NDD (calc'd: 4.28 kDa); 5922 ^CDD long version (calc'd: 7.14 kDa); 5922 ACP-^CDD (calc'd: 16.3 kDa); 5921 ^CDD long version (calc'd: 4.52 kDa); 5921 ^NDD (calc'd: 4.18 kDa); 5920 ^NDD long version (calc'd: 7.29 kDa); 5920 ^NDD short version (calc'd: 4.00 kDa); 5920 ACP-^CDD (calc'd: 13.5 kDa); 5919 ^NDD (4.06 kDa); 5920 ^CDD–5919 ^NDD fusion (G_3S_2) (calc'd: 6.73 kDa). The molecular weights of the markers are indicated. (B) MS spectrum of 5921 ^CDD short version (calc'd: 2.62 kDa). (C) MS spectrum of 5922 ^CDD short version (calc'd: 2.94 kDa). (D) MS spectrum of 5923 ^CDD short version (calc'd: 3.63 kDa). (The MW for 5925 ^CDD short version was verified by SAXS MoW (see **Table S3**))

A



B

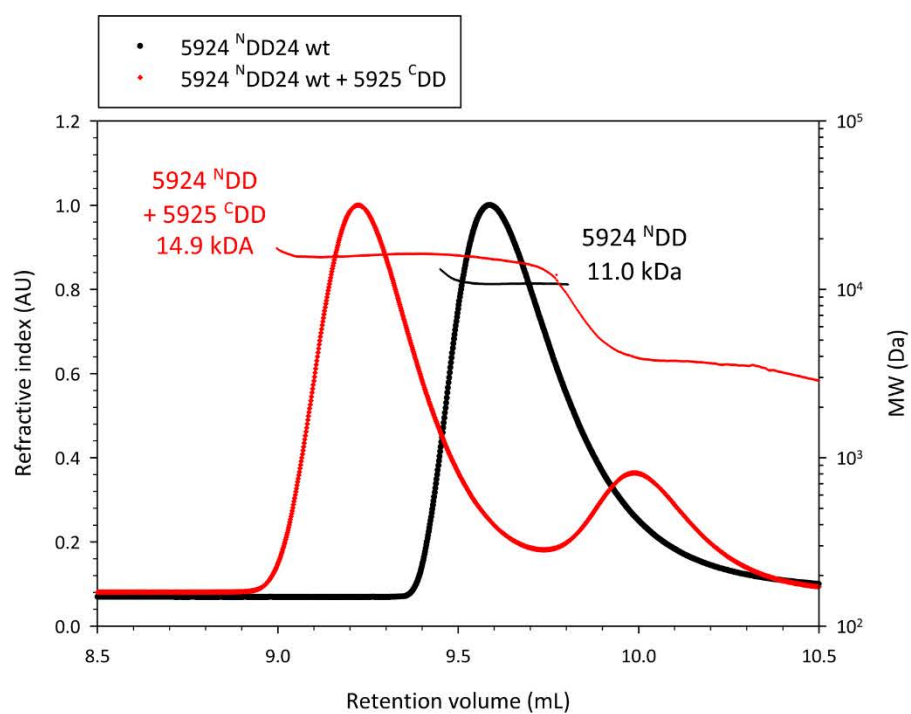


Figure S4. Characterization of the 5925/5924 interface. (A) CD analysis of 5925^{CDD} and 5924^{NDD} at 100 μ M. (B) Analysis by SLS of complex stoichiometry in the presence of excess 5925^{CDD}. The molecular weight of the complex corresponds to one copy of 5925^{CDD} interacting with two copies of 5924^{NDD}. (Calc'd MW 5925^{CDD} (monomer): 3.6 kDa, ^{NDD} (monomer): 5.0 kDa).

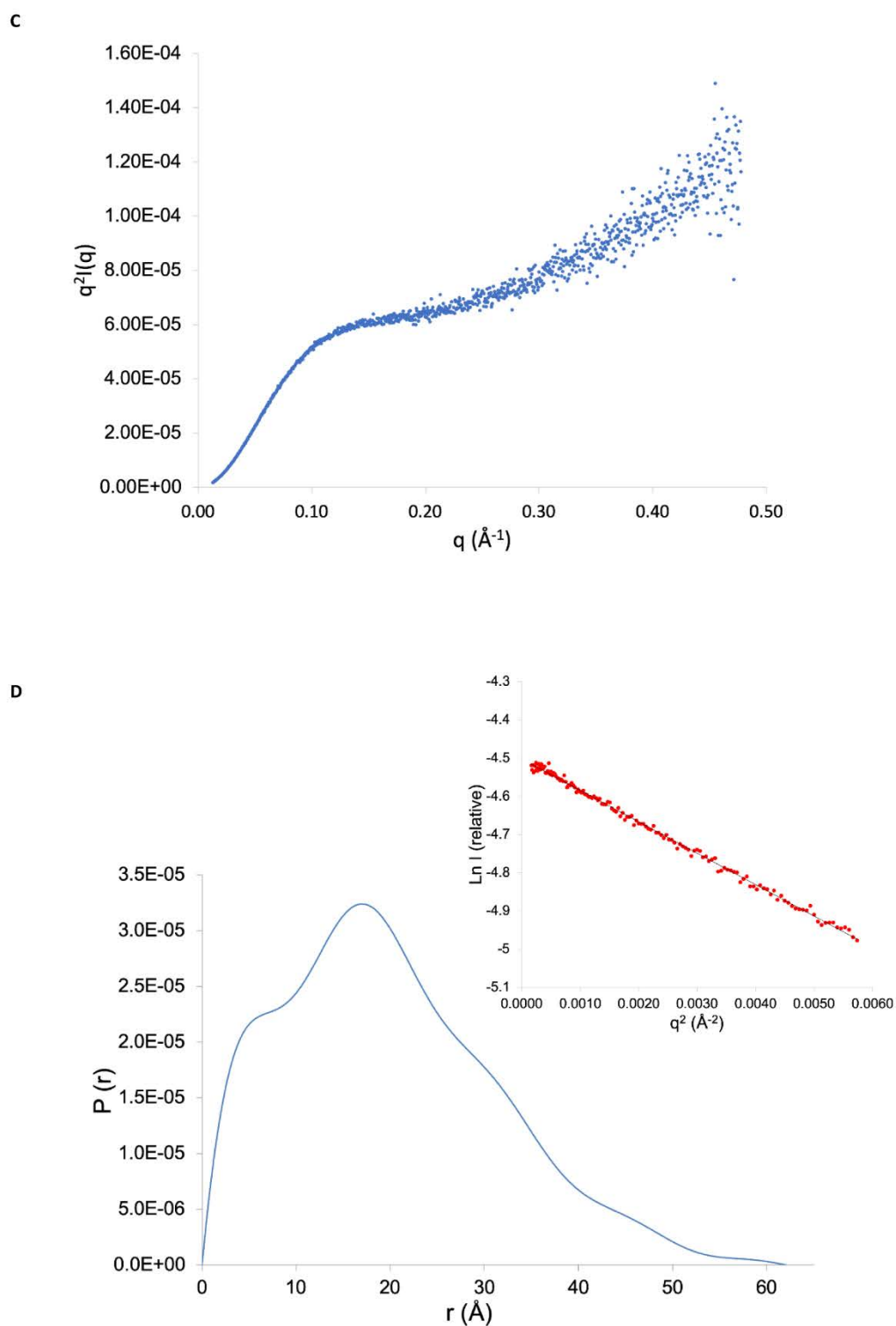


Figure S4, cont. (C) The Kratky plot (blue dots) for 5925 $^{\text{C}}$ DD. (D) The distance distribution function for 5925 $^{\text{C}}$ DD calculated with GNOM.¹² The Guinier plot is inset.

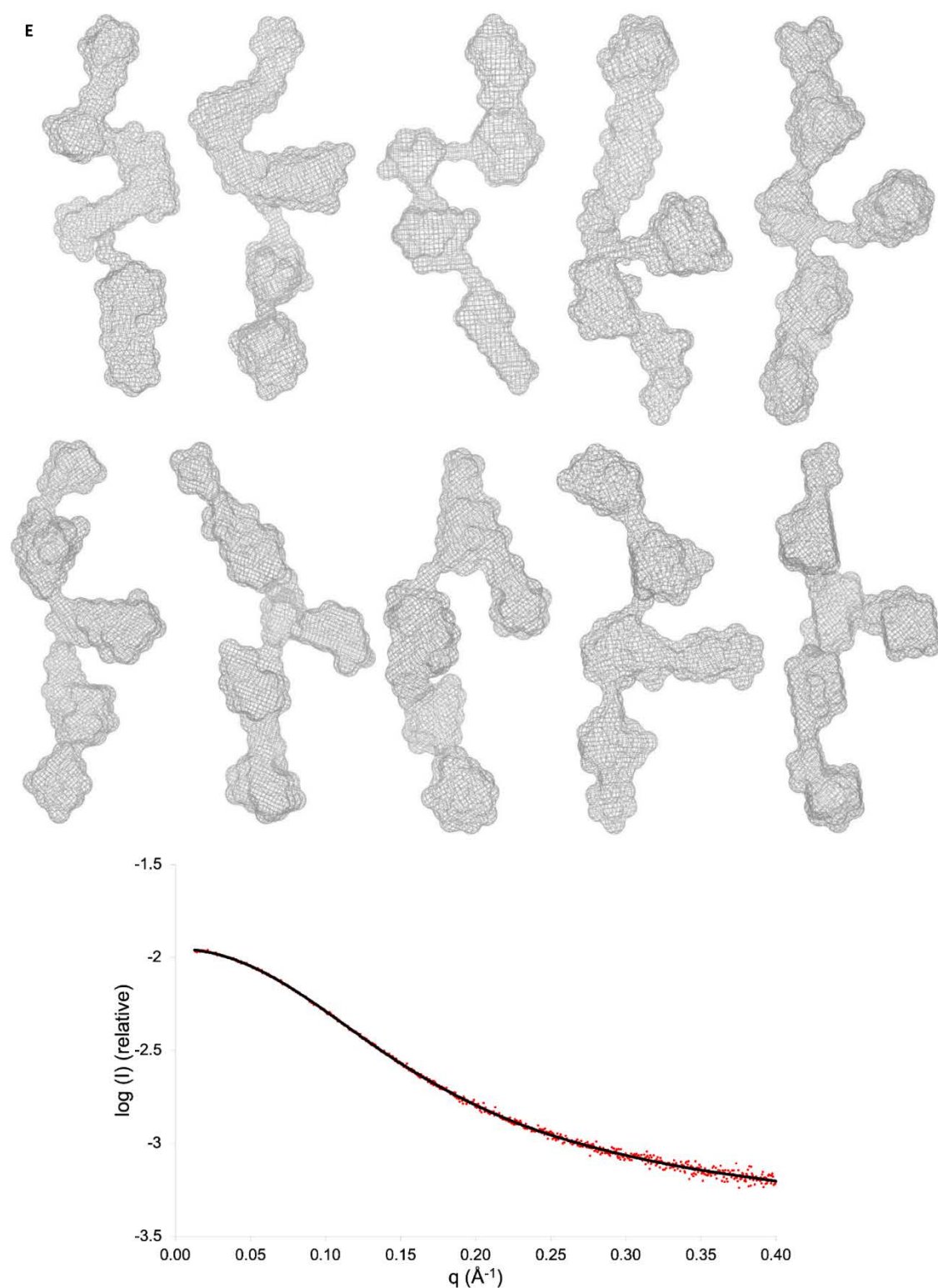


Figure S4, cont. (E) Panel of 10 representative *ab initio* form calculations for 5925 ^6DD obtained with DAMMIN.¹³ The fit between the SAXS data and the 10 form calculations is shown ($\chi^2 = 1.133\text{--}1.223$).

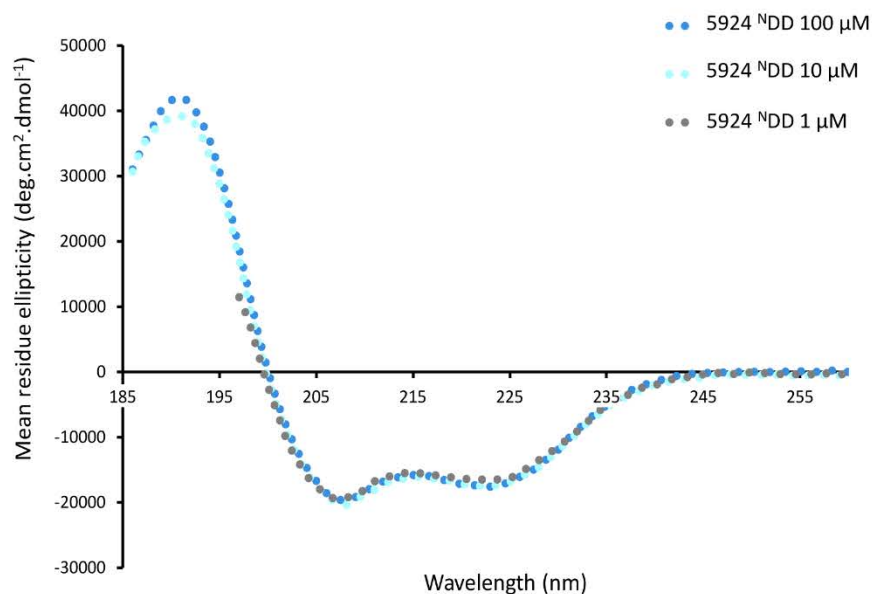
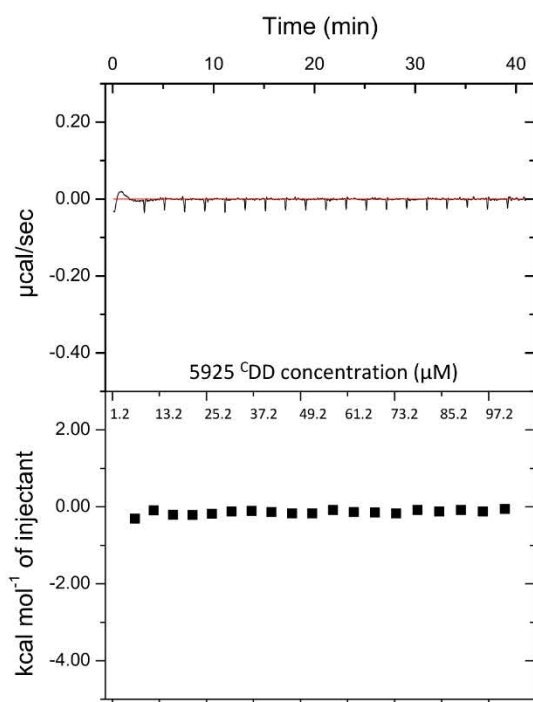
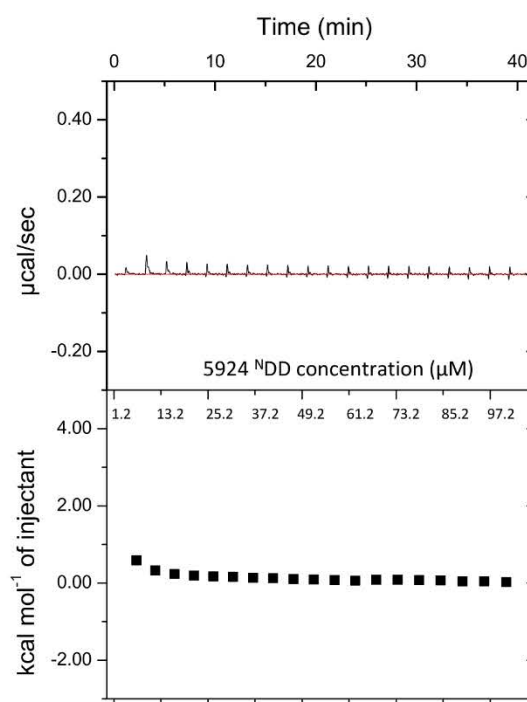
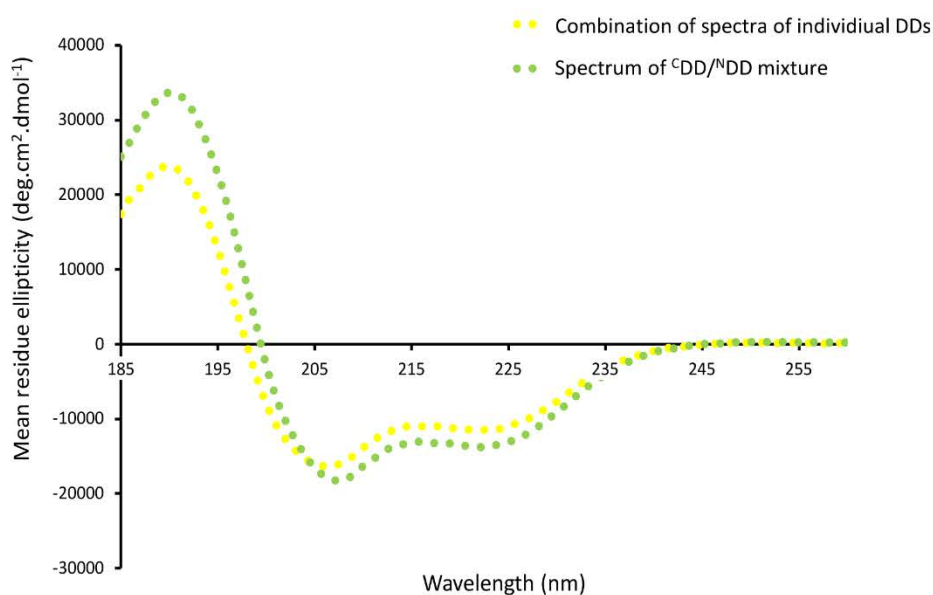
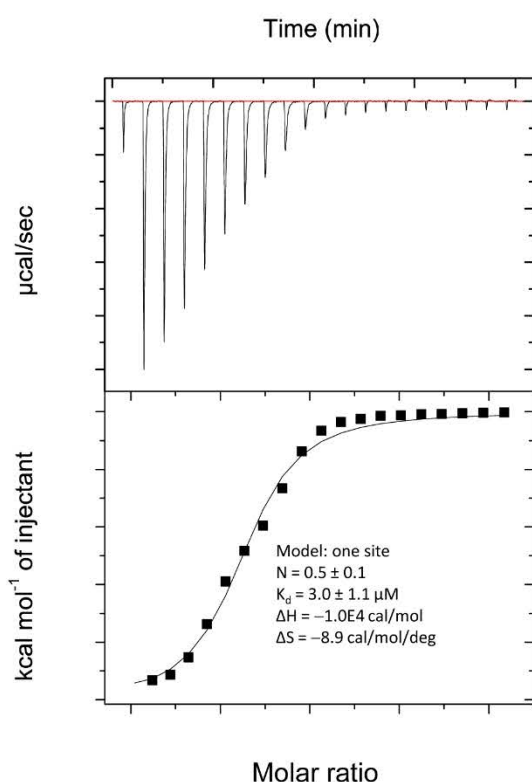
F**G****H**

Figure S4, cont. (F) CD analysis of 5924 ^NDD at 100 μ M, 10 μ M and 1 μ M, showing that no significant loss of α -helix signal is observed. (G) Dilution analysis by ITC of 5925 ^CDD, which gave no change in the observed heat signal. (H) Dilution analysis by ITC of 5924 ^NDD, which gave no significant change in the observed heat signal.

I



J



K

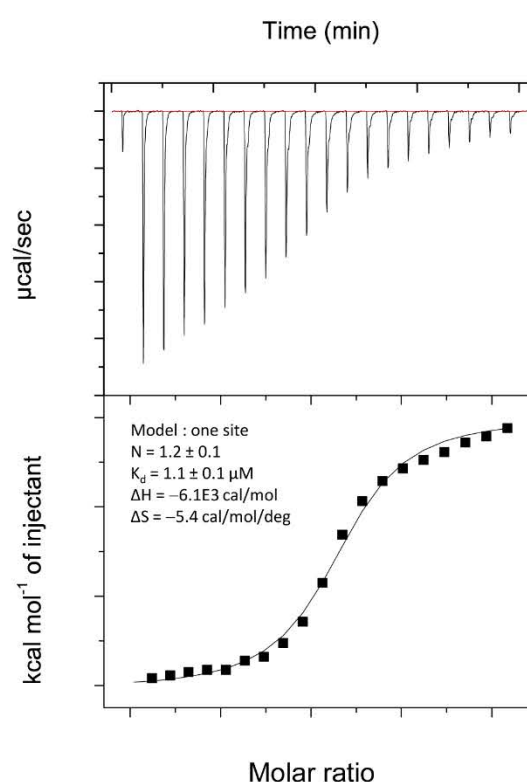
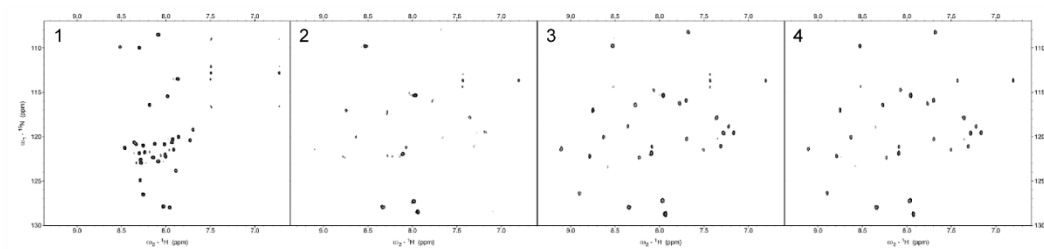


Figure S4, cont. (I) CD spectra showing net structuration upon combining 5925 ¹³DD and 5924 ¹⁵DD. A combined spectrum was obtained by simultaneously analyzing the two isolated DDs (both at 100 μM). Comparison with the spectrum of a mixture of the two docking domains, again at an overall concentration of 100 μM, provides evidence for increased structuration. (J) Analysis by ITC of binding between 5925 ¹³DD in the syringe and 5924 ¹⁵DD in the cell. (K) Analysis by ITC of binding between 5924 ¹⁵DD in the syringe and 5925 ¹³DD in the cell. Panels (J) and (K) show representative data of the triplicate experiments that were carried out, while the measured N and K_d values and their errors are based on the triplicate measurements.

L



M

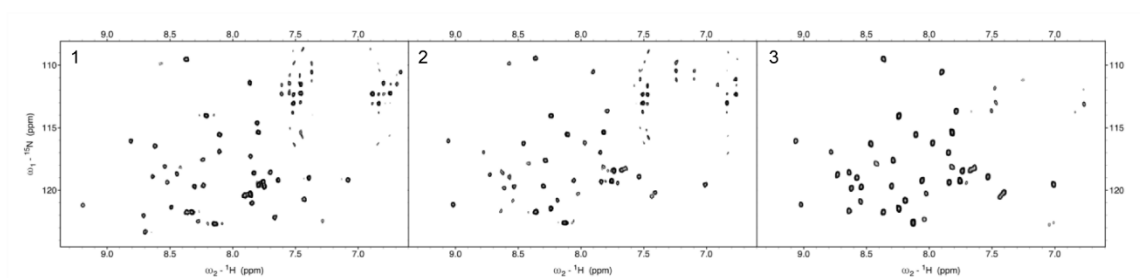


Figure S4, cont. (L) Saturation experiment with ^{15}N -labeled 5925 $^{\text{C}}\text{DD}$ in the presence of unlabeled 5924 $^{\text{N}}\text{DD}$ (concentration ratios ($^{\text{C}}\text{DD}:\text{NDD}$): (1) 400 μM :0 μM ; (2) 200 μM :240 μM ; (3) 100 μM :140 μM ; and, (4) 60 μM :600 μM). (M) Saturation experiment with ^{15}N -labeled 5924 $^{\text{N}}\text{DD}$ in the presence of unlabeled 5925 $^{\text{C}}\text{DD}$ (concentration ratios ($^{\text{N}}\text{DD}:\text{CDD}$): (1) 400 μM :0 μM ; (2) 300 μM :350 μM ; and, (3) 76 μM :600 μM).

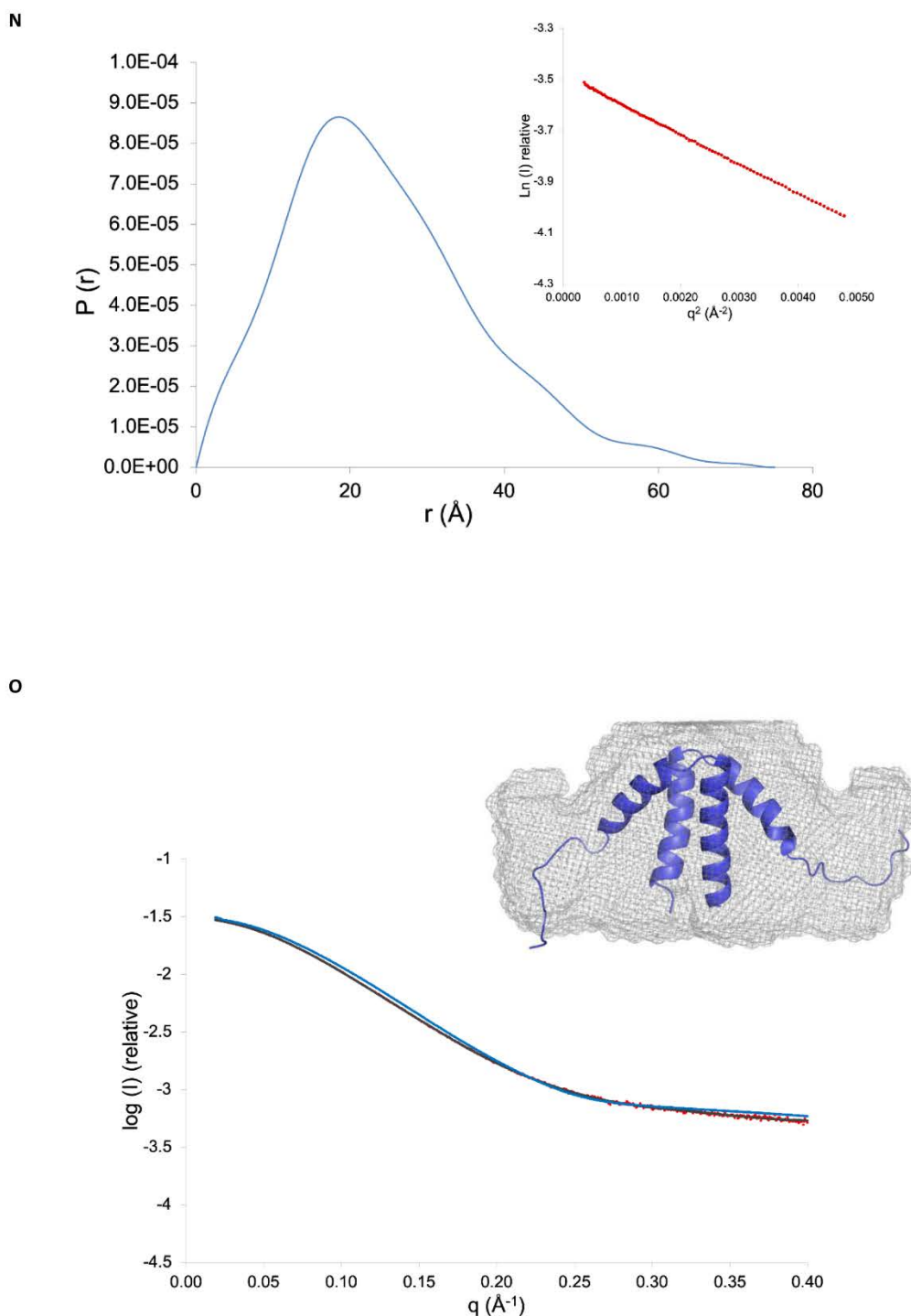
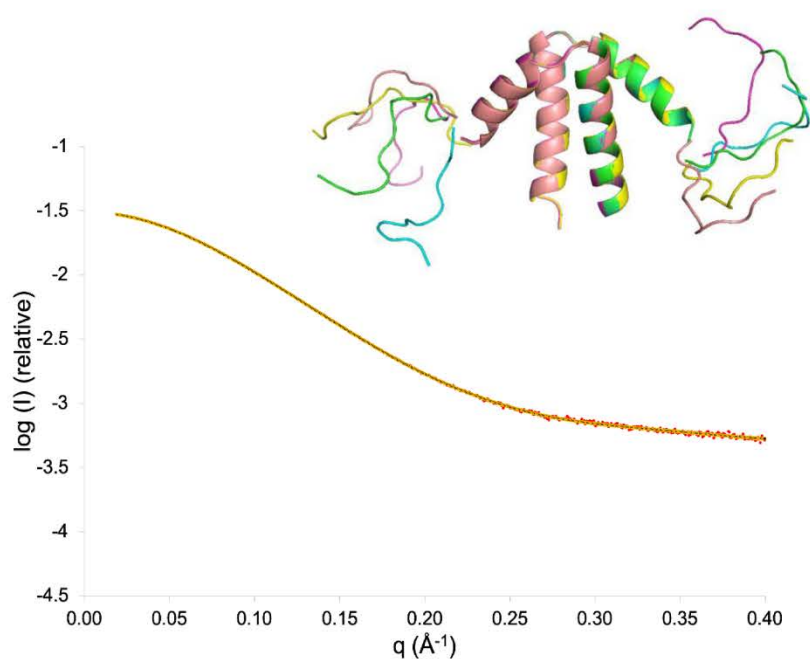


Figure S4, cont. (N) The distance distribution function for 5924 ^NDD calculated with GNOM. The Guinier plot is inset. (O) Analysis of 5924 ^NDD by SAXS showing the difference between the experimental scattering curve (red dots) and theoretical scattering curve predicted from its NMR structure (CRY SOL) ($\chi^2 = 7.2$) (blue line). The fit between the *ab initio* model computed with DAMMIN (dark grey line) and the experimental SAXS data ($\chi^2 = 1.085$) is also shown. The NMR structure of 5924 ^NDD was placed into the average molecular envelope computed by DAMAVER^{14,15} (in grey) using SUPCOMB.¹⁶ This average molecular envelope was obtained using 4 different envelopes computed with DAMMIN (χ^2 ranging from 1.085–1.110).

P



Q

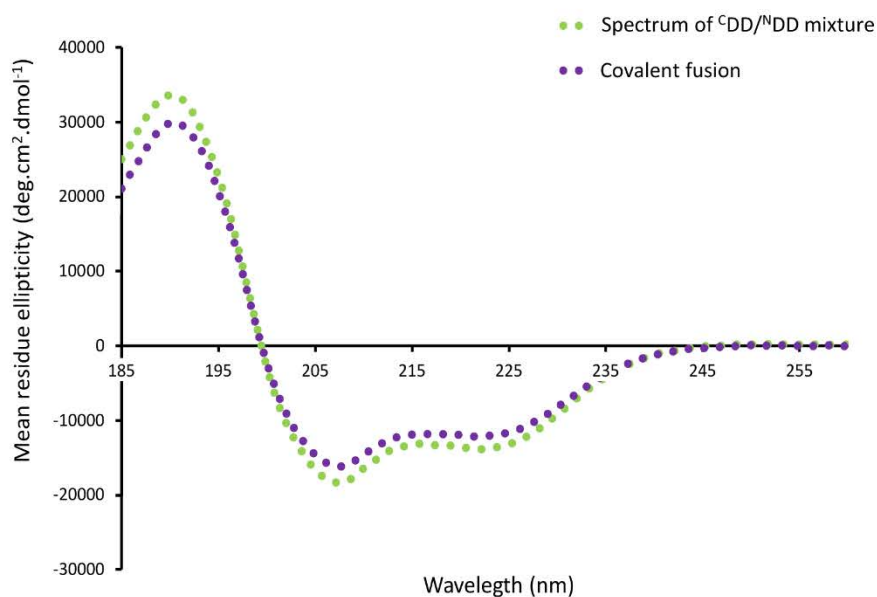


Figure S4, cont. (P) The NMR structure of 5924 ^NDD was divided into 3 different pdb files corresponding to the central 4 α-helix bundle (residues 13–56 of each monomer) and the two disordered regions in N-terminus of each monomer (residues 1–12). By rigid body simulated annealing (SASREF), these segments were positioned in order to fit the experimental scattering curve (orange line) ($\chi^2 = 1.1$). Five possible structures of 5924 ^NDD calculated by SASREF¹⁷ are shown. (Q) CD analysis of a mixture of 5925 ^CDD and 5924 ^NDD (both at 100 μM) and the 5925 ^CDD–5924 ^NDD covalent fusion at 100 μM.

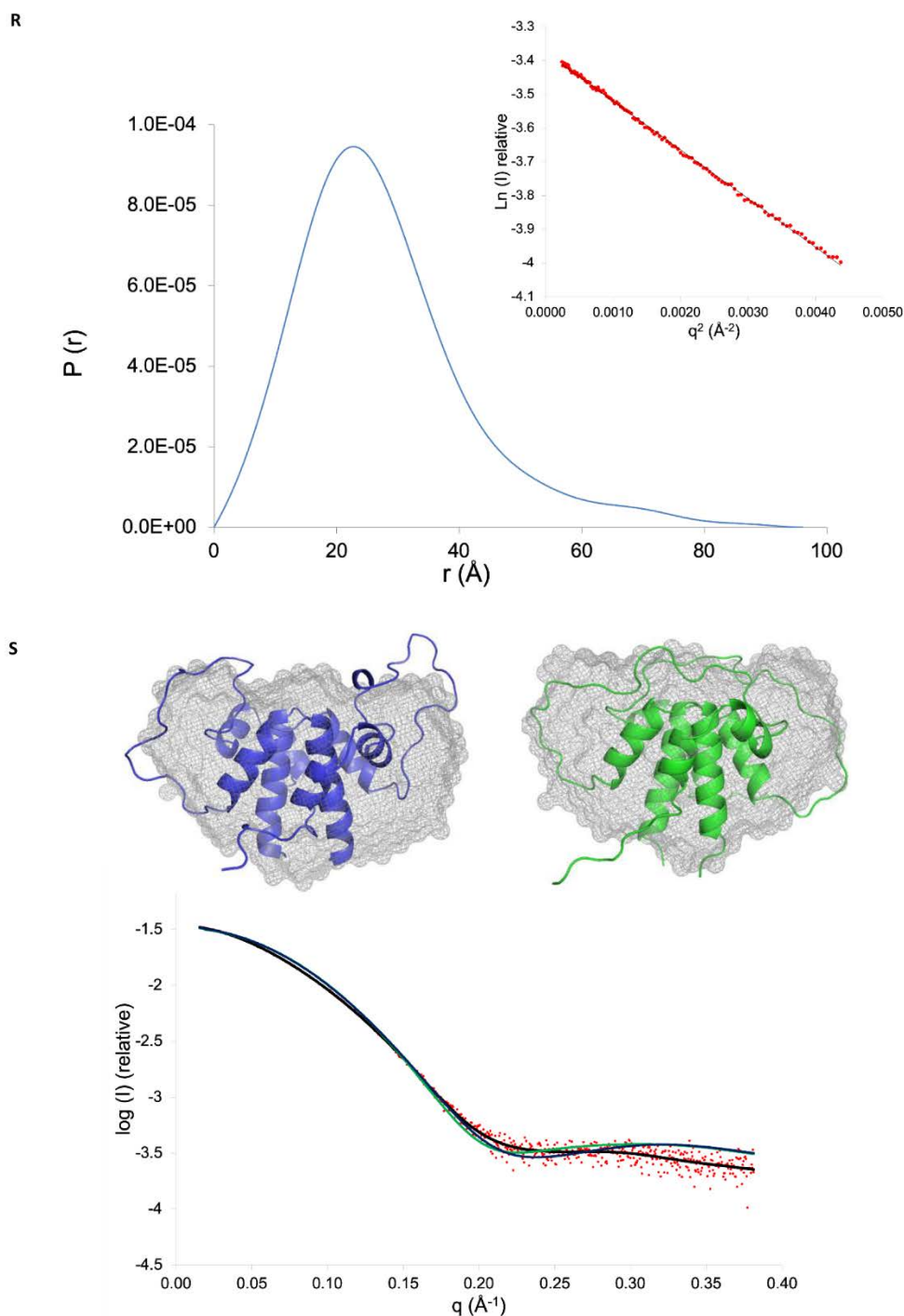


Figure S4, cont. (R) The distance distribution function for the 5925 ^{13}CDD –5924 ^{15}NDD covalent fusion calculated with GNOM. The Guinier plot is inset. (S) Analysis of the 5925 ^{13}CDD –5924 ^{15}NDD covalent fusion by SAXS showing the difference between the experimental scattering curve (red dots) and theoretical scattering curves predicted from the NMR structure in which the artificial $(\text{G}_3\text{S})_3$ linker results in an intermolecular ('inter') interaction between the DDs (blue line) ($\chi^2 = 9.4$), and the NMR structure in which the artificial $(\text{G}_3\text{S})_3$ linker enforces an intramolecular ('intra') DD interaction (green line) ($\chi^2 = 14.4$). The fit between the *ab initio* model computed with DAMMIN (dark grey line) and the experimental SAXS data ($\chi^2 = 1.147$) is also shown. These two NMR structures ('inter' in blue and 'intra' in green) were placed into the average molecular envelope computed by DAMFILT (in grey) using SUPCOMB. This average molecular envelope was obtained using 5 different envelopes computed by DAMMIN (χ^2 ranging from 1.147–1.173).

T

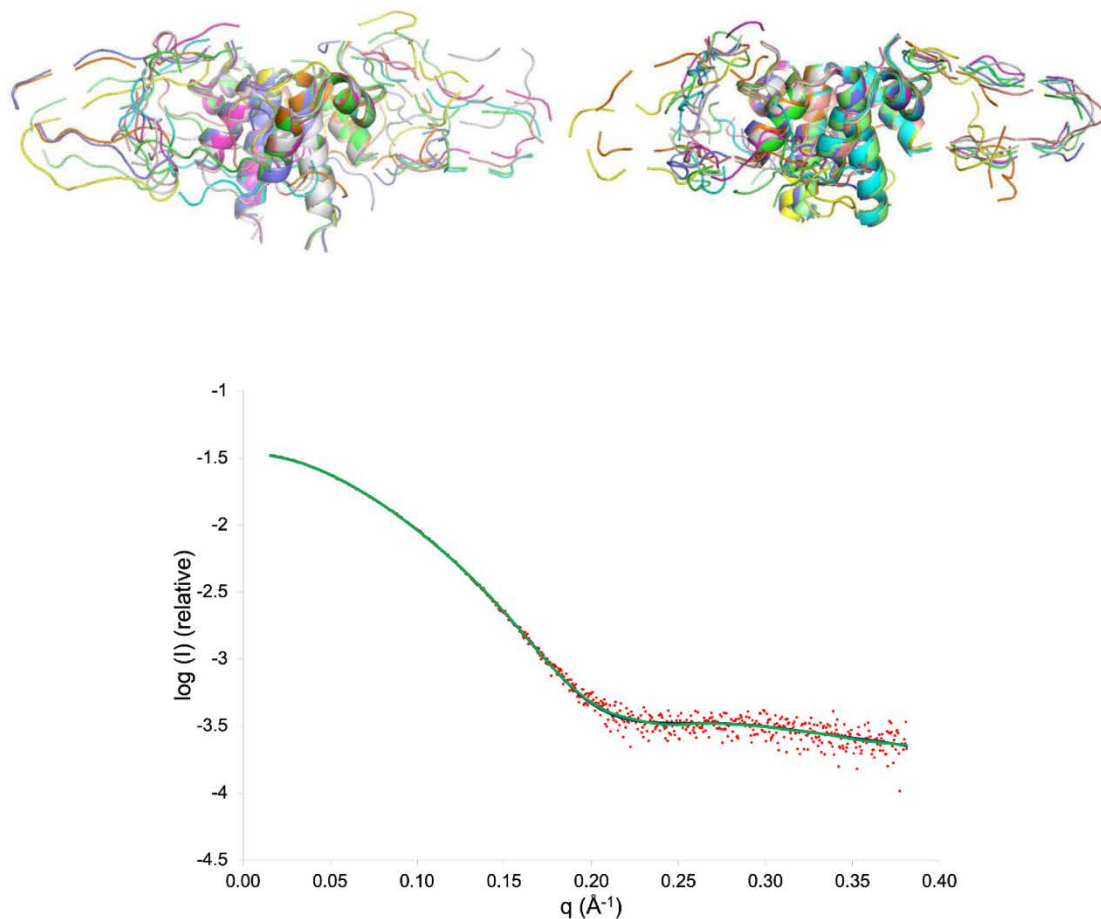
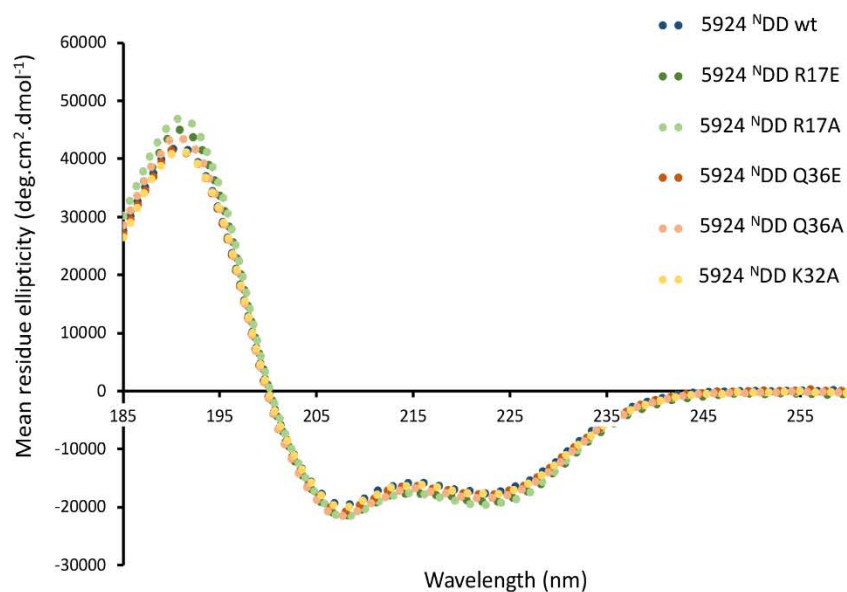


Figure S4, cont. (T) The NMR structures of the 5925 ¹³DD–5924 ¹⁵DD covalent complex where divided into 13 different pdb files corresponding to the two N-terminal disordered regions (residues 1–10), the 8 α -helix bundle of the complex (residues 11–33, 55–87, 88–120 and 142–174), and the two artificial linkers divided into 5 different segments of equal length. By rigid body simulated annealing (SASREF) the linker region was forced to allow ‘inter’ communication between the two monomers (nine of the resulting possible structures are shown in the left-hand panel). The fit between the experimental data curve (red dots) and the theoretical curves computed by SASREF is also shown (blue, NMR structure in which the linker is ‘inter’, $\chi^2 = 1.2$; green, NMR structure in which the linker is ‘intra’, $\chi^2 = 1.2$).

U



V

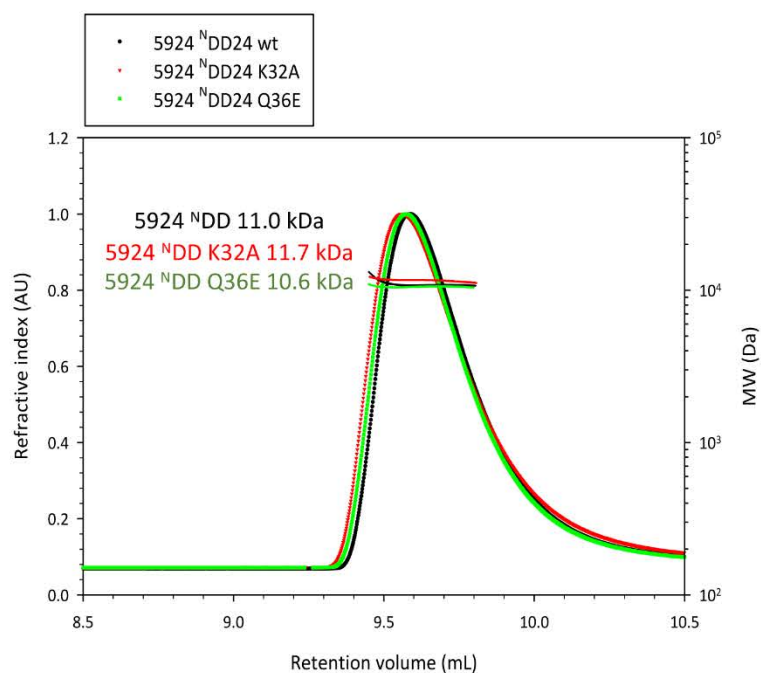
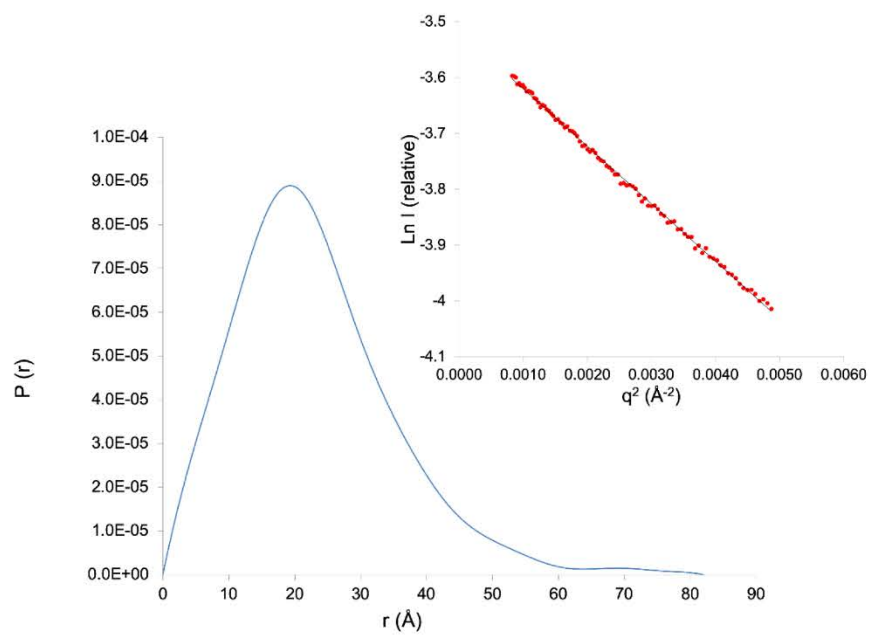


Figure S4, cont. (U) CD analysis of 5924^{NDD} and its mutants. (V) Molecular weight determination of 5924^{NDD} wt, 5924^{NDD} K32A and 5924^{NDD} Q36E by SEC-SLS. (Calc'd MW 5924^{NDD} wt (monomer): 5.0 kDa; 5924^{NDD} K32A (monomer): 4.9 kDa; 5924^{NDD} Q36E (monomer): 5.0 kDa).

W



X

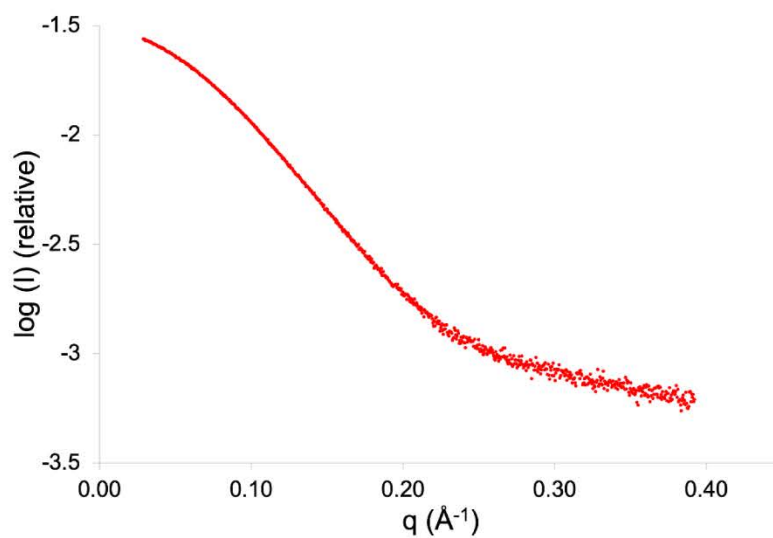
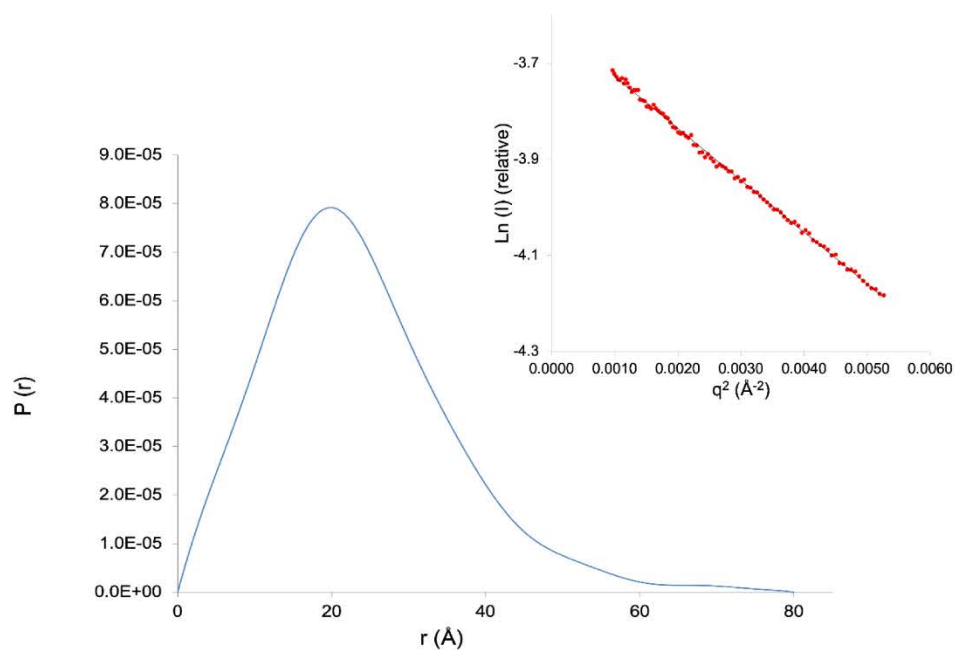


Figure S4, cont. (W) The distance distribution function for 5924 ^NDD R17A calculated with GNOM. The Guinier plot is inset. (X) Experimental scattering curve obtained for 5924 ^NDD R17A.

Y



Z

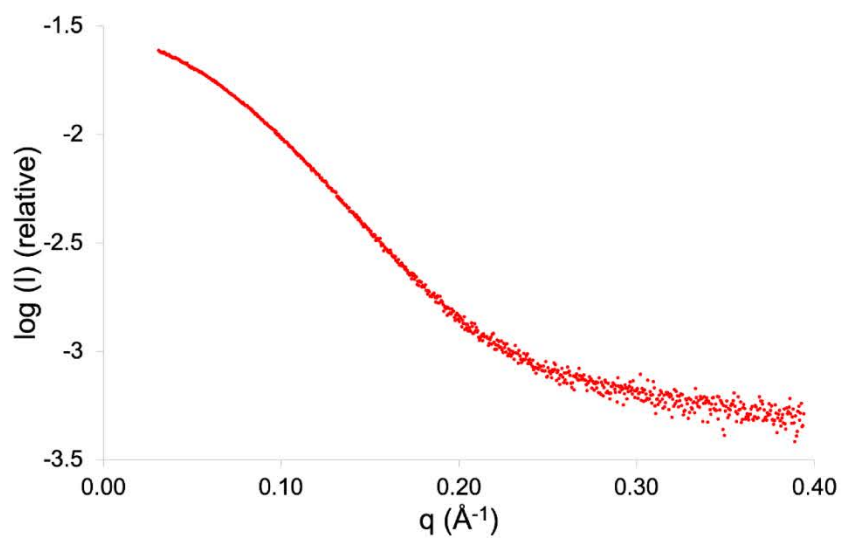


Figure S4, cont. (Y) The distance distribution function for 5924 N DD K32A calculated with GNOM. The Guinier plot is inset. (Z) Experimental scattering curve obtained for 5924 N DD K32A.

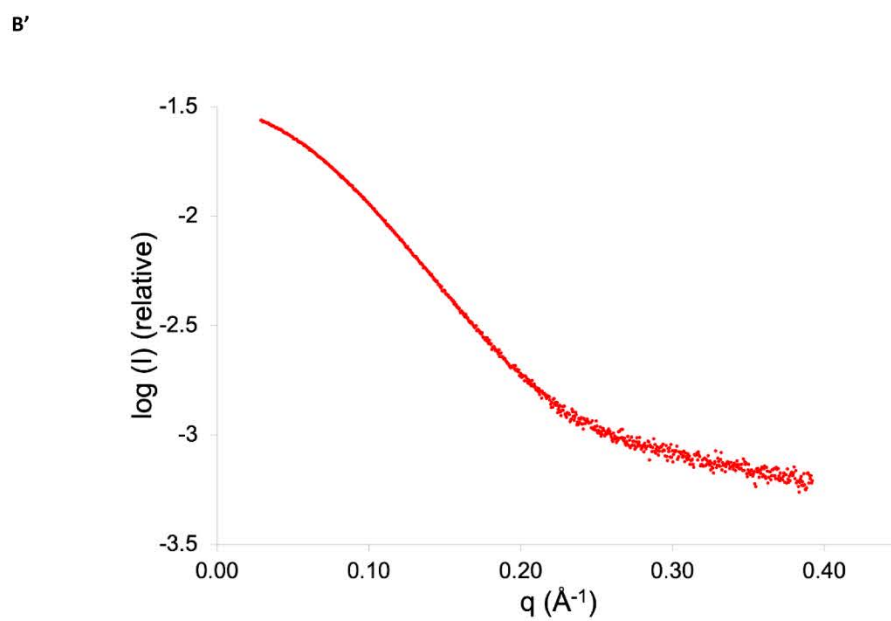
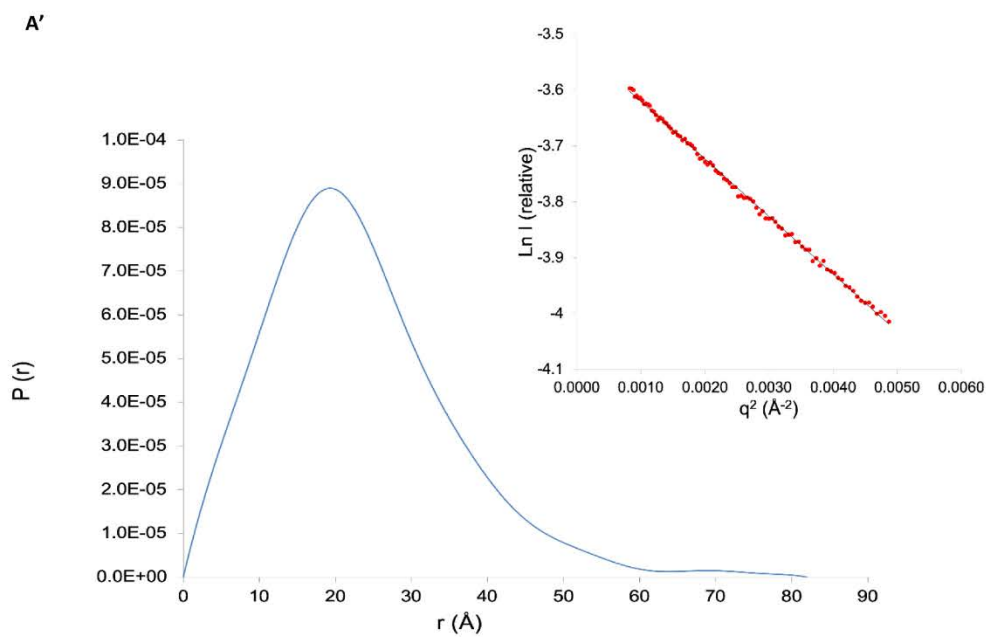


Figure S4, cont. (A') The distance distribution function for 5924 N DD Q36E calculated with GNOM. The Guinier plot is inset. (B') Experimental scattering curve obtained for 5924 N DD Q36E.

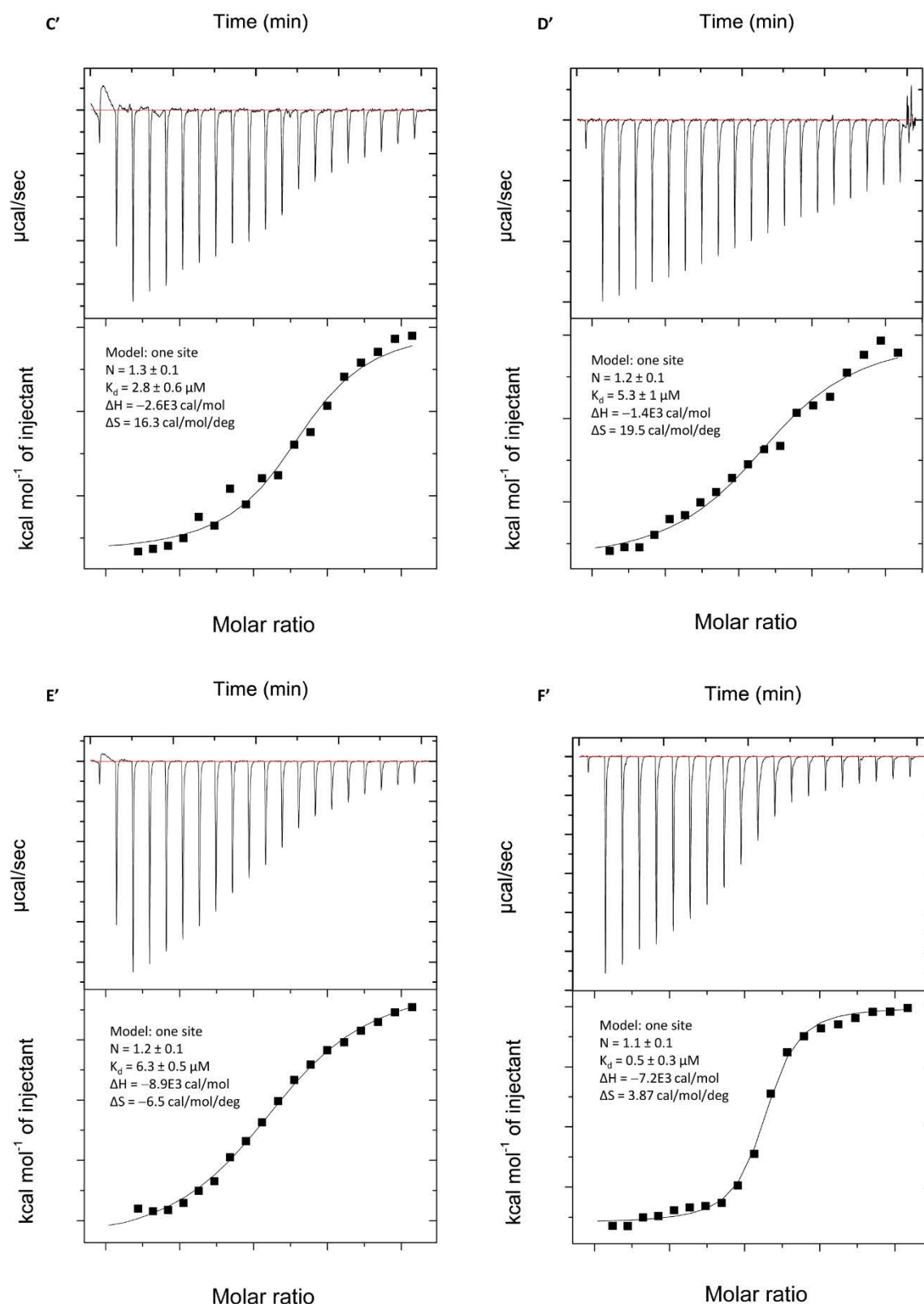


Figure S4, cont. (C') Analysis by ITC of binding between 5924 ^NDD R17A in the syringe and 5925 ^CDD in the cell. (D') Analysis by ITC of binding between 5924 ^NDD R17E in the syringe and 5925 ^CDD in the cell. (E') Analysis by ITC of binding between 5924 ^NDD K32A in the syringe and 5925 ^CDD in the cell. (F') Analysis by ITC of binding between 5924 ^NDD Q36A in the syringe and 5925 ^CDD in the cell. All panels show representative data of the triplicate experiments that were carried out, while the measured N and K_d values and their errors are based on the triplicate measurements.

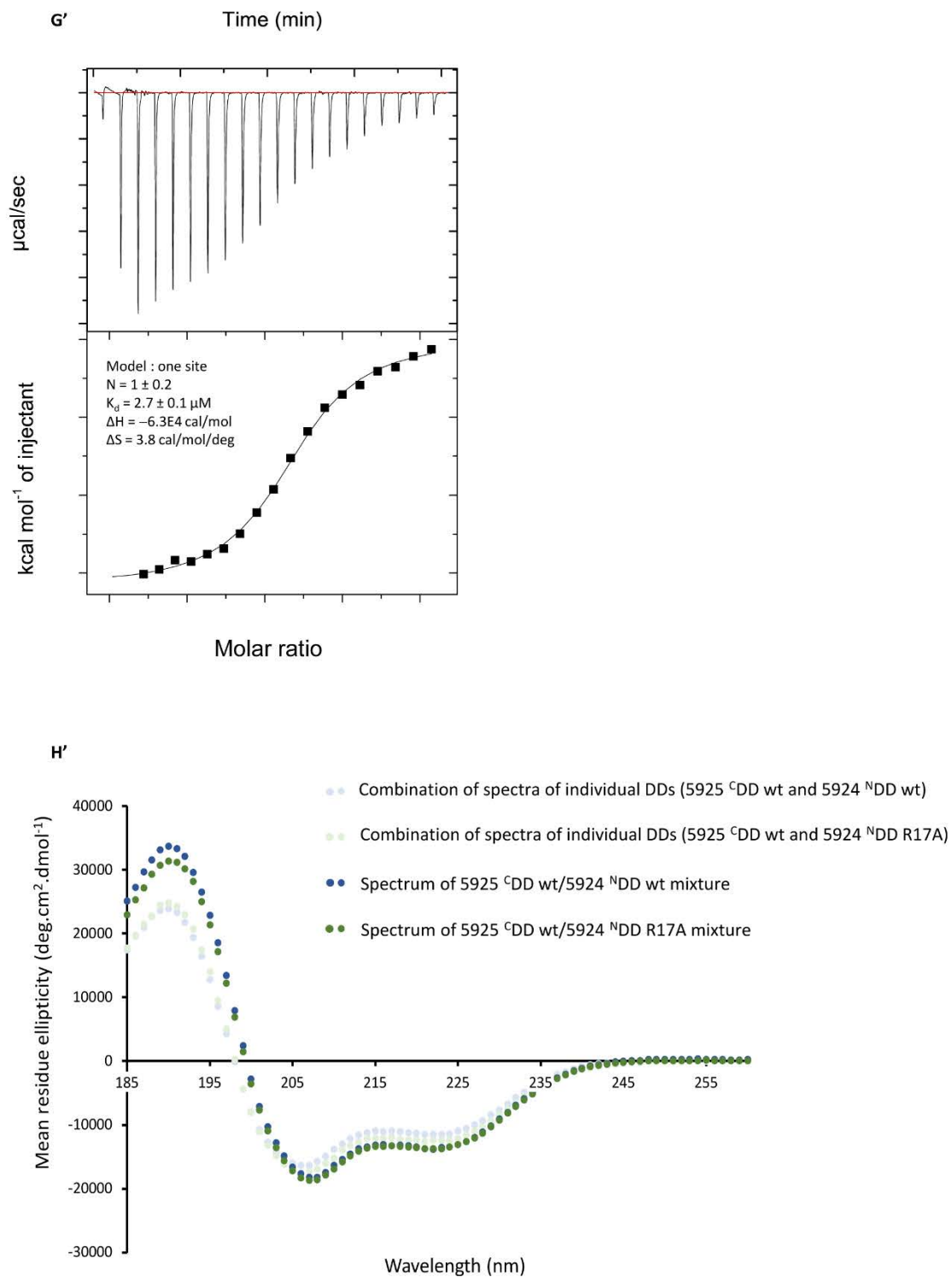


Figure S4, cont. (G') Analysis by ITC of binding between 5924 ^NDD Q36E in the syringe and 5925 ^CDD in the cell. Shown are representative data of the triplicate experiments that were carried out, while the measured N and K_d values and their errors are based on the triplicate measurements. (H') CD spectra showing structuration upon combining 5925 ^CDD and 5924 ^NDD R17A. A combined spectrum was obtained by simultaneously analyzing the two isolated DDs (both at 100 μM). The spectra are compared to those obtained with the wt 5924 ^NDD.

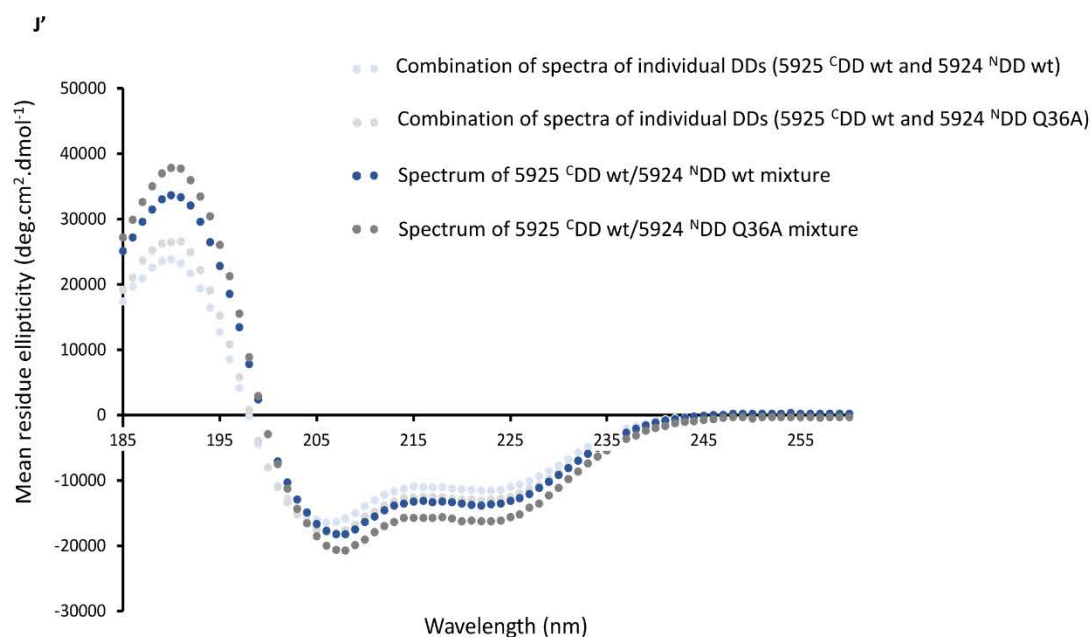
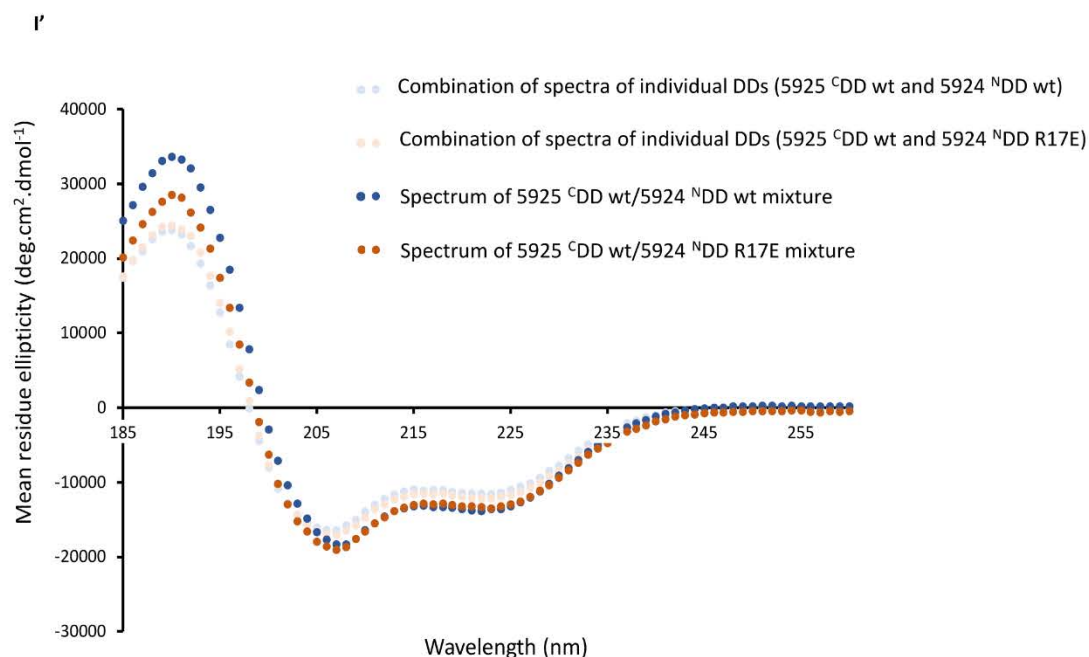


Figure S4, cont. (I') CD spectra showing structuration upon combining 5925 ^CDD and 5924 ^NDD R17E. A combined spectrum was obtained by simultaneously analyzing the two isolated DDs (both at 100 μM). The spectra are compared to those obtained with the wt 5924 ^NDD. (J') CD spectra showing structuration upon combining 5925 ^CDD and 5924 ^NDD Q36A. A combined spectrum was obtained by simultaneously analyzing the two isolated DDs (both at 100 μM). The spectra are compared those obtained with the wt 5924 ^NDD.

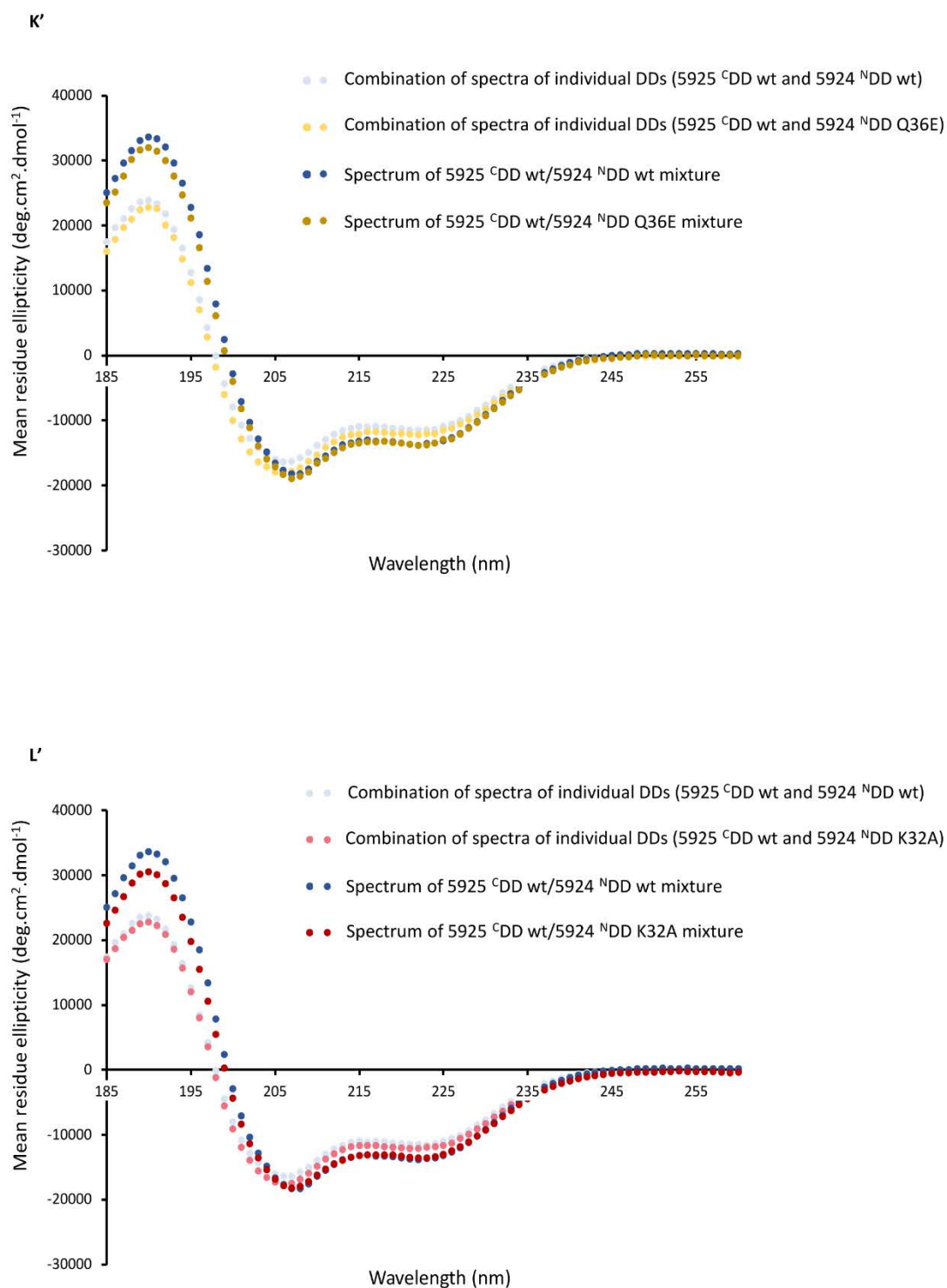


Figure S4, cont. (K') CD spectra showing structuration upon combining 5925 ^CDD and 5924 ^NDD Q36E. A combined spectrum was obtained by simultaneously analyzing the two isolated DDs (both at 100 μ M). The spectra are compared to those obtained with the wt 5924 ^NDD. (L') CD spectra showing structuration upon combining 5925 ^CDD and 5924 ^NDD K32A. A combined spectrum was obtained by simultaneously analyzing the two isolated DDs (both at 100 μ M). The spectra are compared to those obtained with the wt 5924 ^NDD.

Table S3. Summary of the SAXS data obtained on the DD constructs.

Construct	R_g^a Guinier (Å)	R_g (GNOM) (Å)	D_{max}^b (GNOM) (Å)	Calc'd MW (kDa)	Exper. MW (kDa) (SAXS MoW)	χ^2 ^c (DAMMIN)	χ^2 (CRY SOL)	χ^2 (SASREF)	nsd
5925 ^c DD short version	15.63 ± 0.03	16.33 ± 0.06	62	3.66	3.1	1.135	n.d.	n.d.	n.d.
5924 ^N DD	18.68 ± 0.01	19.33 ± 0.02	75	10.04 (dimer)	10.17 (dimer)	1.085	7.2	1.1	1.88
5924 ^N DD K32A	18.04 ± 0.03	18.77 ± 0.09	80	9.92 (dimer)	9.74 (dimer)	n.d.	n.d.	n.d.	n.d.
5924 ^N DD Q36E	17.71 ± 0.02	18.26 ± 0.07	77	10.04 (dimer)	9.24 (dimer)	n.d.	n.d.	n.d.	n.d.
5924 ^N DD R17A	17.55 ± 0.02	18.38 ± 0.07	82	9.87 (dimer)	8.40 (dimer)	n.d.	n.d.	n.d.	n.d.
5925 ^c DD–5924 ^N DD (G ₃ S) ₃ covalent fusion	20.87 ± 0.03	21.89 ± 0.08	96	18.28 (dimer)	23.08 (dimer)	1.092	9.4 (inter) ^f 14.4 (intra) ^f	1.2 (inter) 1.2 (intra to inter) ^g	1.93 (inter) 1.88 (intra)
5920 ^c DD–5919 ^N DD (G ₃ S) ₂ covalent fusion	17.74 ± 0.02	17.96 ± 0.02	60	13.45 (dimer)	12.61 (dimer)	1.171	2.9	n.d.	1.64
5920 ^c DD short version	12.43 ± 0.03	13.14 ± 0.05	49	2.17	21 ^d ; 3.31 ^e	1.318	n.d.	n.d.	n.d.

^a R_g is the radius of gyration derived from the atomic models using the program PRIMUS.¹⁴ For experimental data, R_g is given by the Guinier approximation.

^b D_{max} is the maximal particle diameter derived from the distance distribution function (P(r)) using the program GNOM.¹²

^c χ^2 is the discrepancy between the experimental SAXS profile and the theoretical calculated curve, obtained using the programs DAMMIN,¹³ SASREF¹⁷ or CRY SOL.¹⁸

^dThe SAXS MoW¹⁹ method by which the molecular weight is calculated is not adapted to this particular disordered region.

^eCalculated using the volume correlation (Vc) method.²⁰

^fThe designations 'inter' and 'intra' refer to whether the interaction between the docking domains within the artificially-fused complex occurs in an inter- or intramolecular fashion.

^g'Intra to inter' refers to a SASREF calculation carried out on a pdb file of the complex, in which the normally intramolecular interaction between 5925 ^cDD/5924^NDD was forced to be intermolecular via manipulation of the intervening linkers (**Figure S4 (S)**).

n.d.: not determined

Table S4. NMR and refinement statistics.

	5924 ¹ NDD	5925 ¹³ CD–5924 ¹ NDD	5920 ¹³ CD–5919 ¹ NDD
NMR distance and dihedral constraints			
Distance constraints			
Total NOE	1824	2875	2504
Intra-residue ($ i-j = 0$)	520	755	660
Inter-residue ($ i-j \neq 0$)	1304	2120	1844
Sequential ($ i-j = 1$)	466	812	670
Medium-range ($ i-j < 4$)	544	710	692
Long-range ($ i-j > 5$)	294	598	482
Hydrogen bonds	0	0	0
Total dihedral angle restraints	116	136	148
ϕ	58	68	74
ψ	58	68	74
Constraint violations, mean $\pm$ S.D			
Distances violations < 0.1 Å	5.12 $\pm$ 1.43	9.44 $\pm$ 2.17	9.12 $\pm$ 2.28
Distances violations > 0.1 Å	0.12 $\pm$ 0.21	0.96 $\pm$ 0.61	0.64 $\pm$ 0.61
Average maximum distance violation (Å)	0.08 $\pm$ 0.01	0.12 $\pm$ 0.03	0.10 $\pm$ 0.01
Dihedral angle violations (°)	0	0	0
Average maximum dihedral angle violation (°)	0	0	0
Ramachandran statistics (%)^a			
Most favoured regions	95.3	92.6	94.6
Additional allowed regions	3.6	4.9	5.0
Generously allowed regions	0.5	0.5	0.3
Disallowed regions	0.6	2.1	0.2
Precision, root mean square deviation from the mean (Å)			
	5924 ¹ NDD W10–Q37	5925 ¹³ CD E2621–E2640, 5924 ¹ NDD W10–Q37	5920 ¹³ CD P2960–G2973, 5919 ¹ NDD, M1–E26
All heavy atoms	0.99 $\pm$ 0.09	1.23 $\pm$ 0.12	1.17 $\pm$ 0.11
Backbone	0.40 $\pm$ 0.07	0.63 $\pm$ 0.11	0.66 $\pm$ 0.13
Molprobit^b			
Molprobit score	0.5 (100 th percentile)	1.33 (98 th percentile)	1.16 (99 th percentile)
Clashscore, all atom	0 (100 th percentile)	0 (100 th percentile)	0 (100 th percentile)

^aDetermined by PROCHECK-NMR²¹^bDetermined by MolProbit²² server on the first model (best representative) of each NMR ensembles.

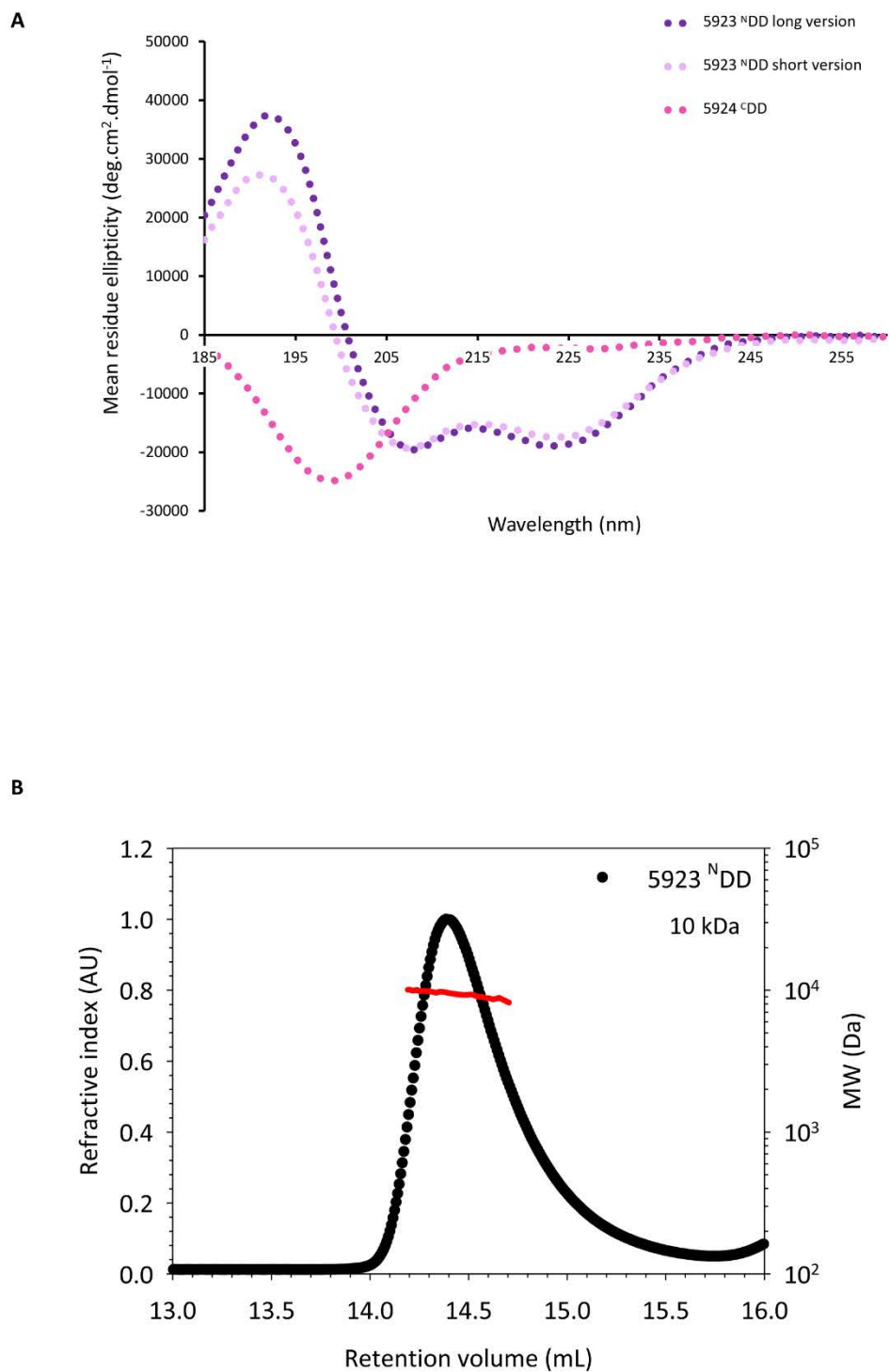


Figure S5. Characterization of the 5924/5923 interface. (A) Analysis of 5923 ^NDD long and short versions and 5924 ^CDD at 100 μ M. (B) Oligomeric state analysis of 5923 ^NDD long version by SLS. (Calc'd MW (monomer): 4.6 kDa)

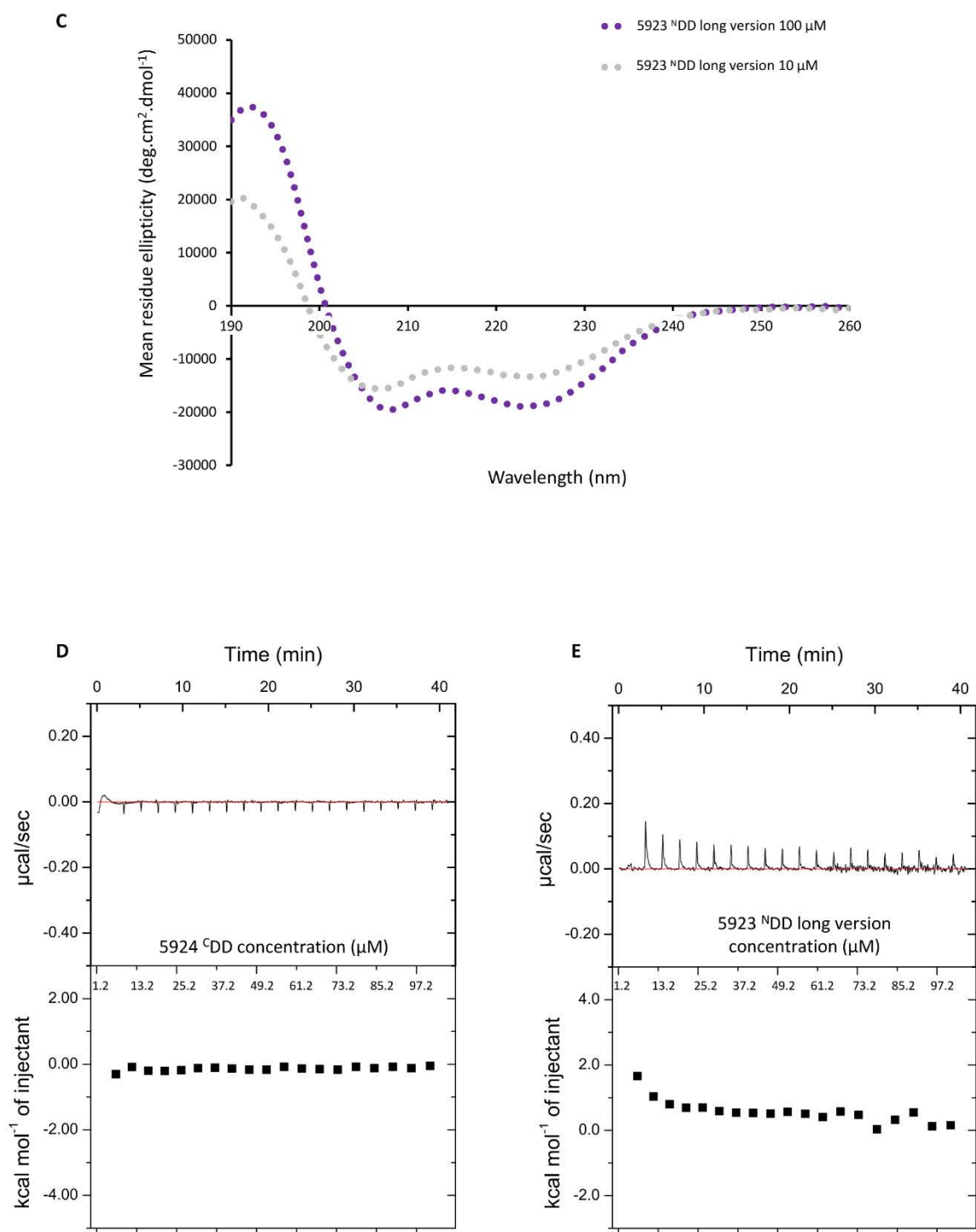
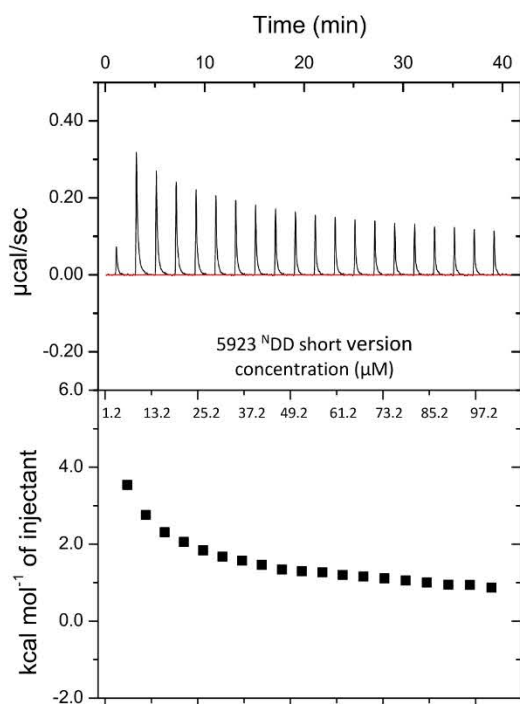


Figure S5, cont. (C) Analysis of 5923 ^NDD long version at concentrations above (100 μM) and below (10 μM) the point at which it is suspected to homodimerize. (D) Dilution analysis of 5924 ^CDD, which gave no change in the observed heat signal. (E) Dilution analysis of 5923 ^NDD long version, which gave changes in the observed heat signal.

F



G

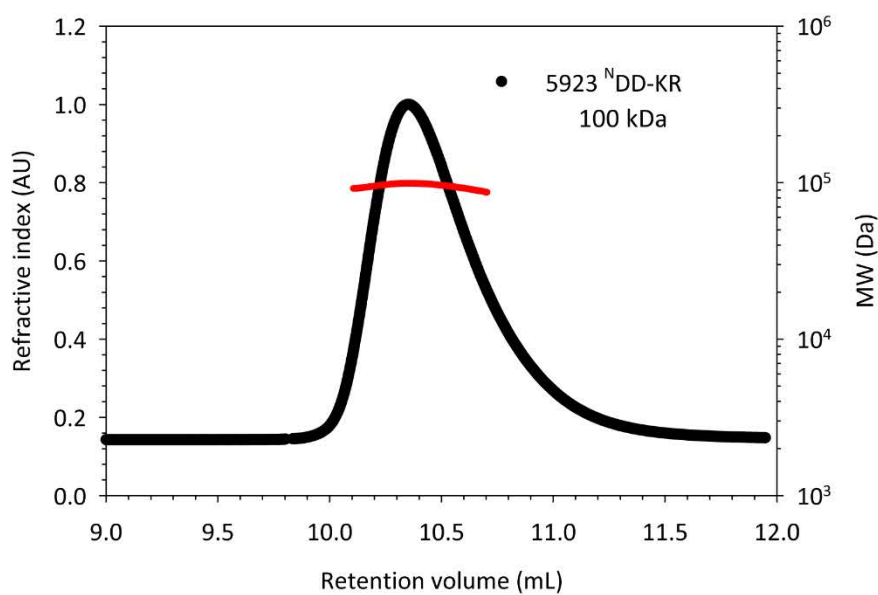


Figure S5, cont. (F) Dilution analysis of 5923 ^NDD short version, which gave changes in the observed heat signal. (G) Oligomeric state analysis of 5923 ^NDD-KR by SLS. (Calc'd MW (monomer): 57 kDa).

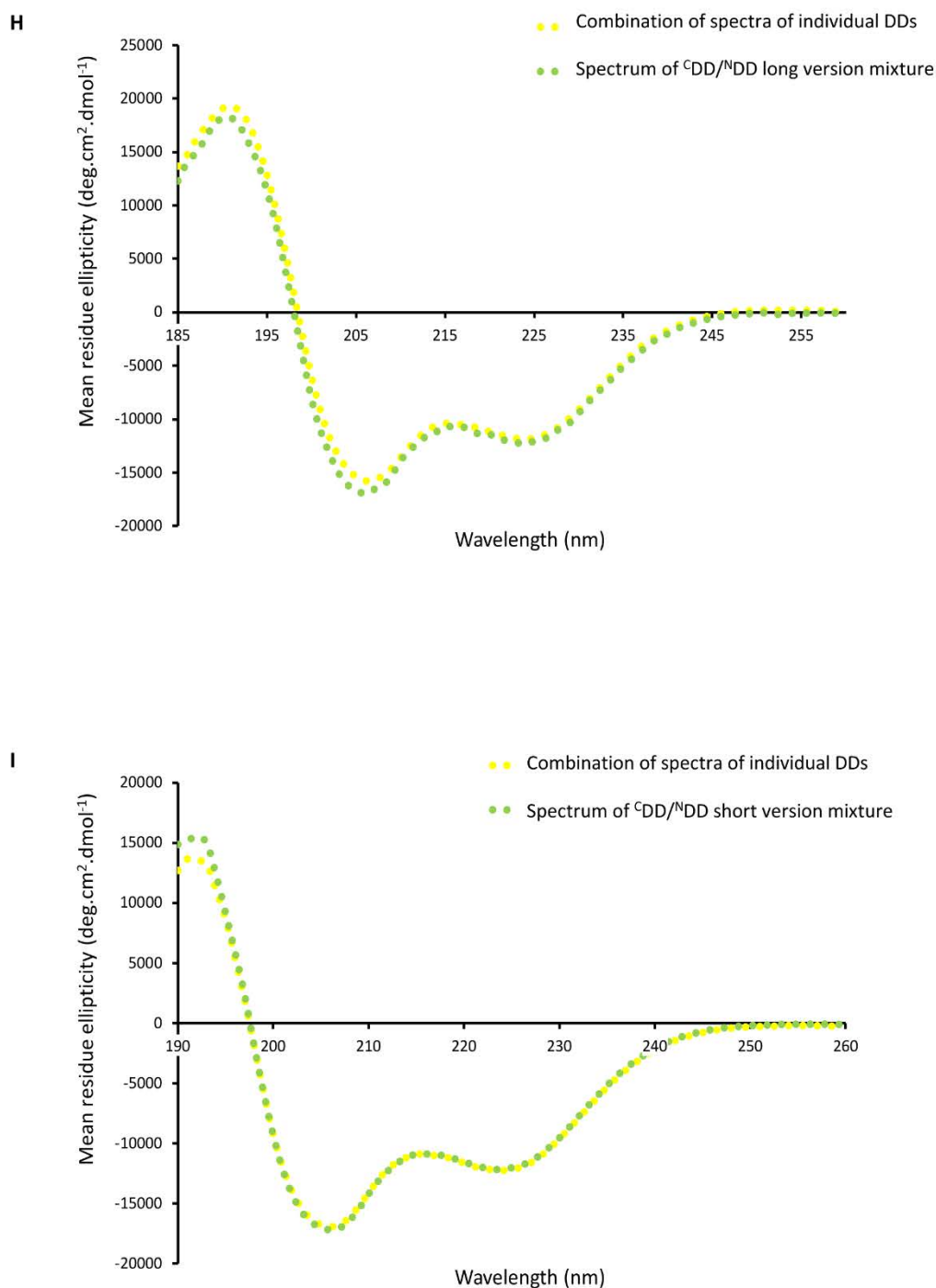


Figure S5, cont. (H) CD spectra showing no structuration upon combining 5924 ¹³CDD and 5923 ¹⁵NDD long. A combined spectrum was obtained by simultaneously analyzing the two isolated DDs (both at 100 μ M). Comparison with the spectrum of a mixture of the two docking domains, again at an overall concentration of 100 μ M, provides no evidence for increased structuration. (I) CD spectra showing no structuration upon combining 5924 ¹³CDD and 5923 ¹⁵NDD short. This experiment was performed as described in (H).

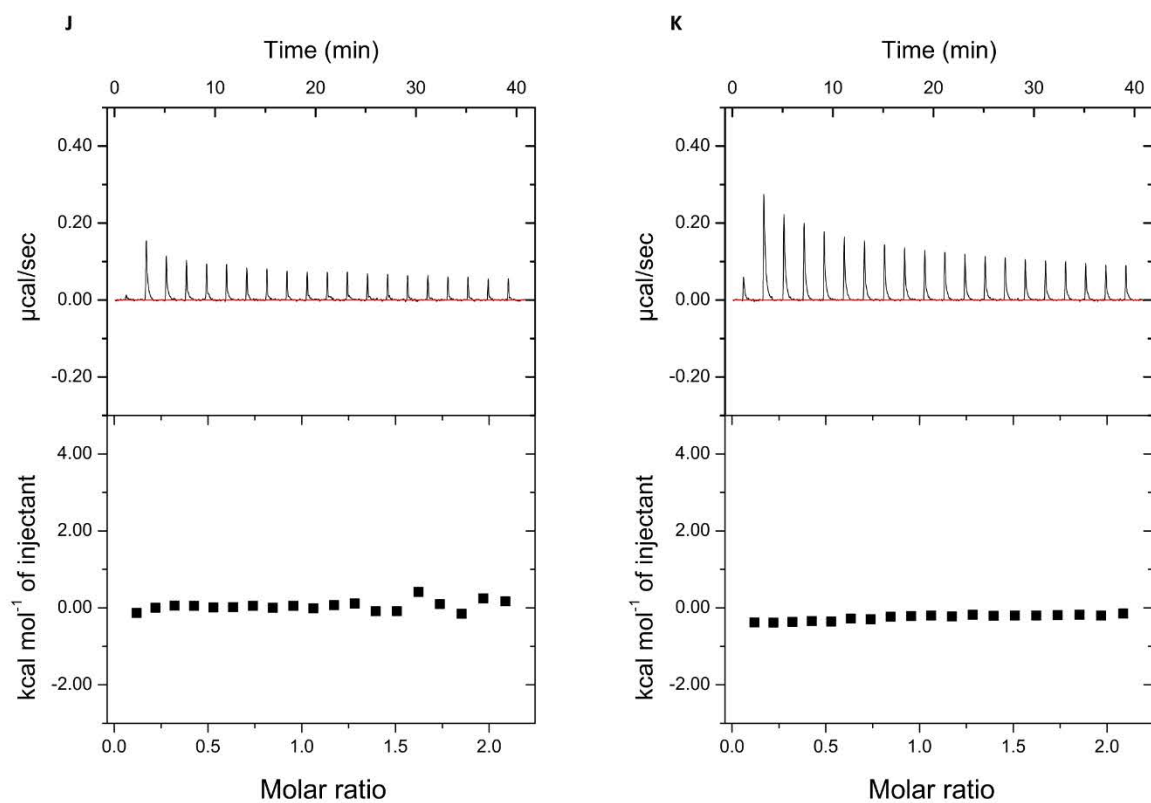


Figure S5, cont. (J) Analysis by ITC of binding between 5923 ^NDD long and 5924 ^CDD. (K) Analysis by ITC of binding between 5923 ^NDD short and 5924 ^CDD.

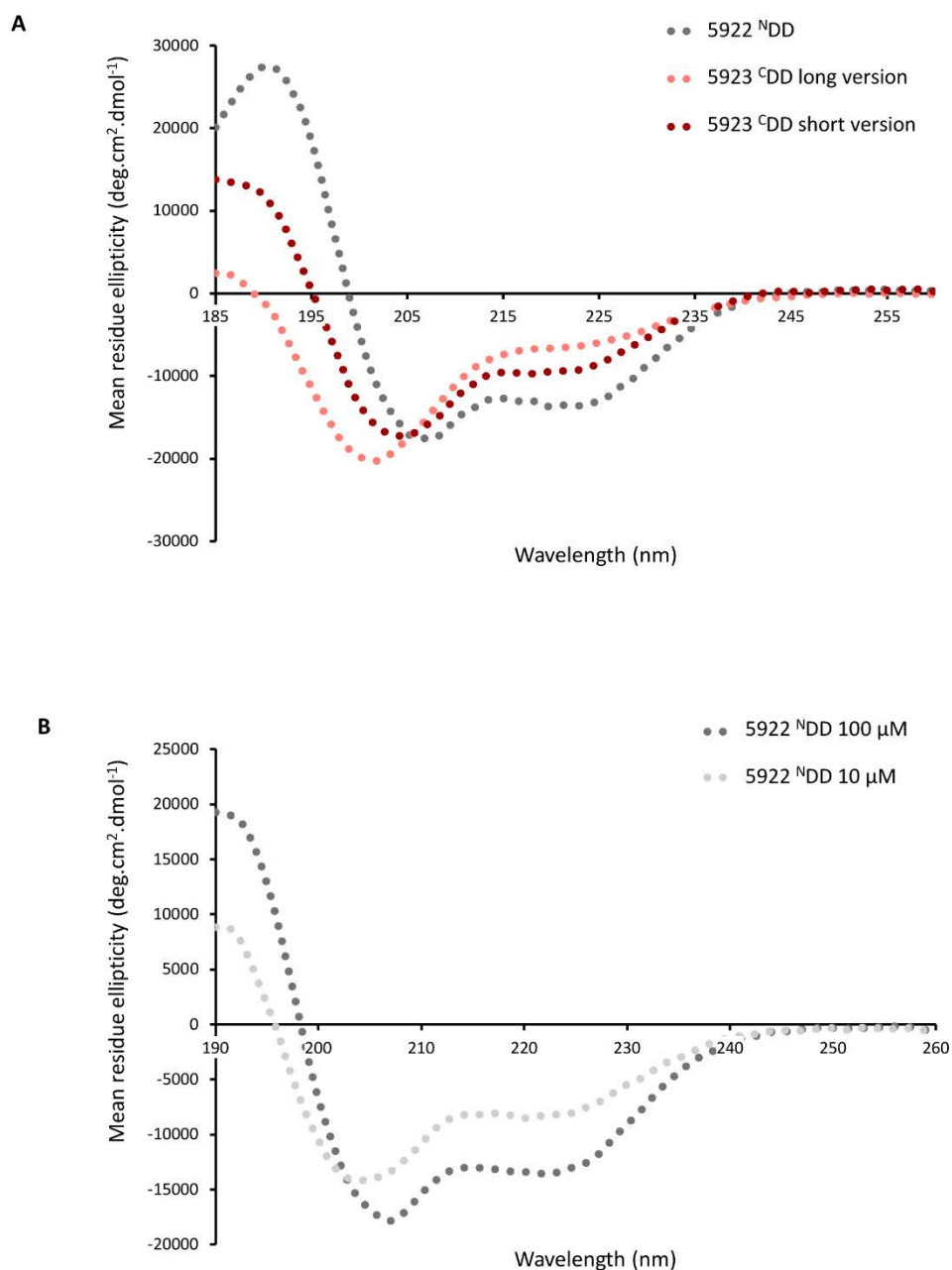


Figure S6. Characterization of the 5923/5922 interface. (A) CD analysis of 5923 ^CDD long and short versions and 5922 ^NDD at 100 μM. (B) CD analysis of 5922 ^NDD at concentrations above (100 μM) and below (10 μM) the point at which it is suspected to homodimerize.

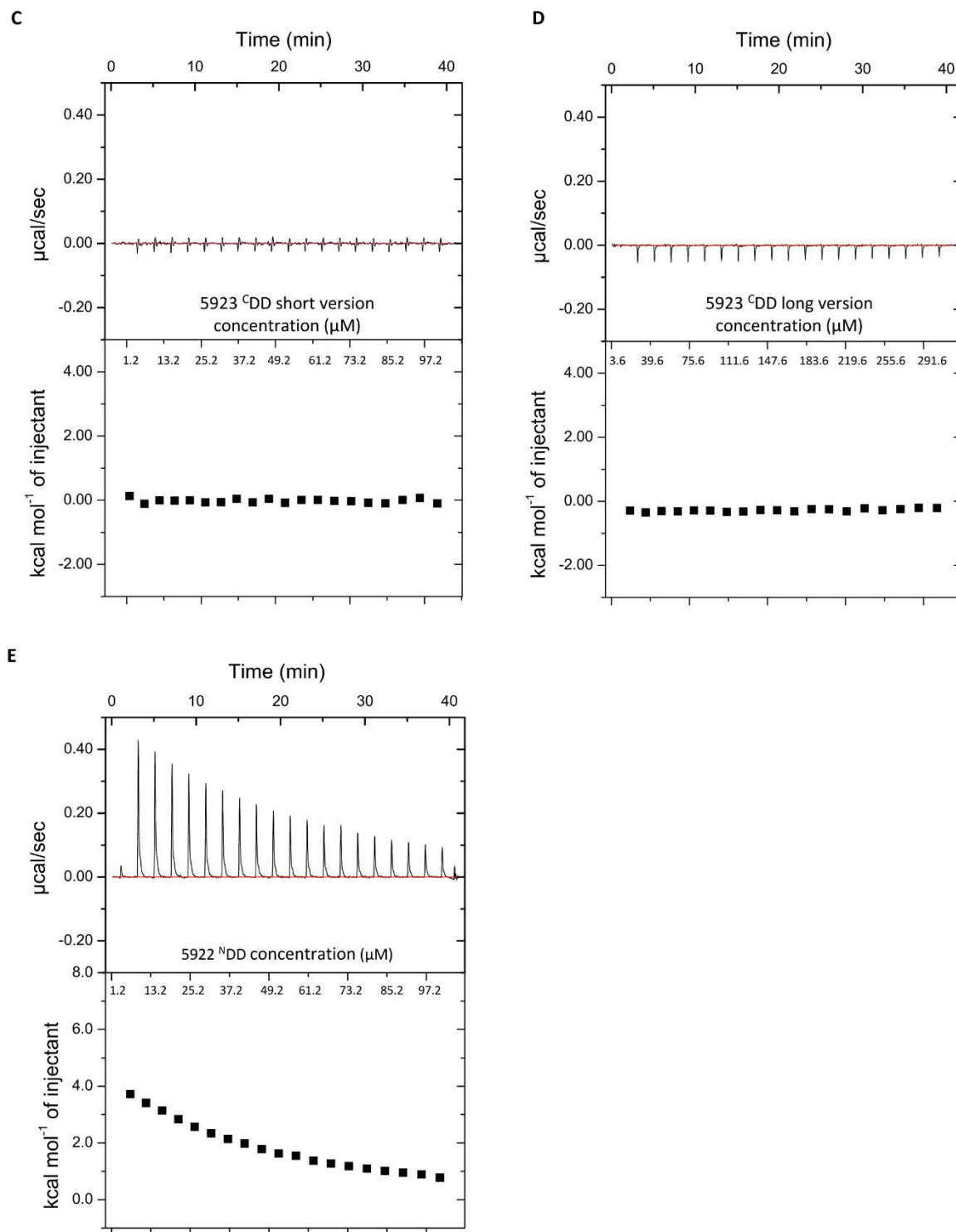


Figure S6, cont. (C) Dilution analysis by ITC of 5923 ^CDD short, which gave no change in the observed heat signal. (D) Dilution analysis by ITC of 5923 ^CDD long, which gave no change in the observed heat signal. (E) Dilution analysis by ITC of 5922 ^NDD long, which yielded changes in the heat signal.

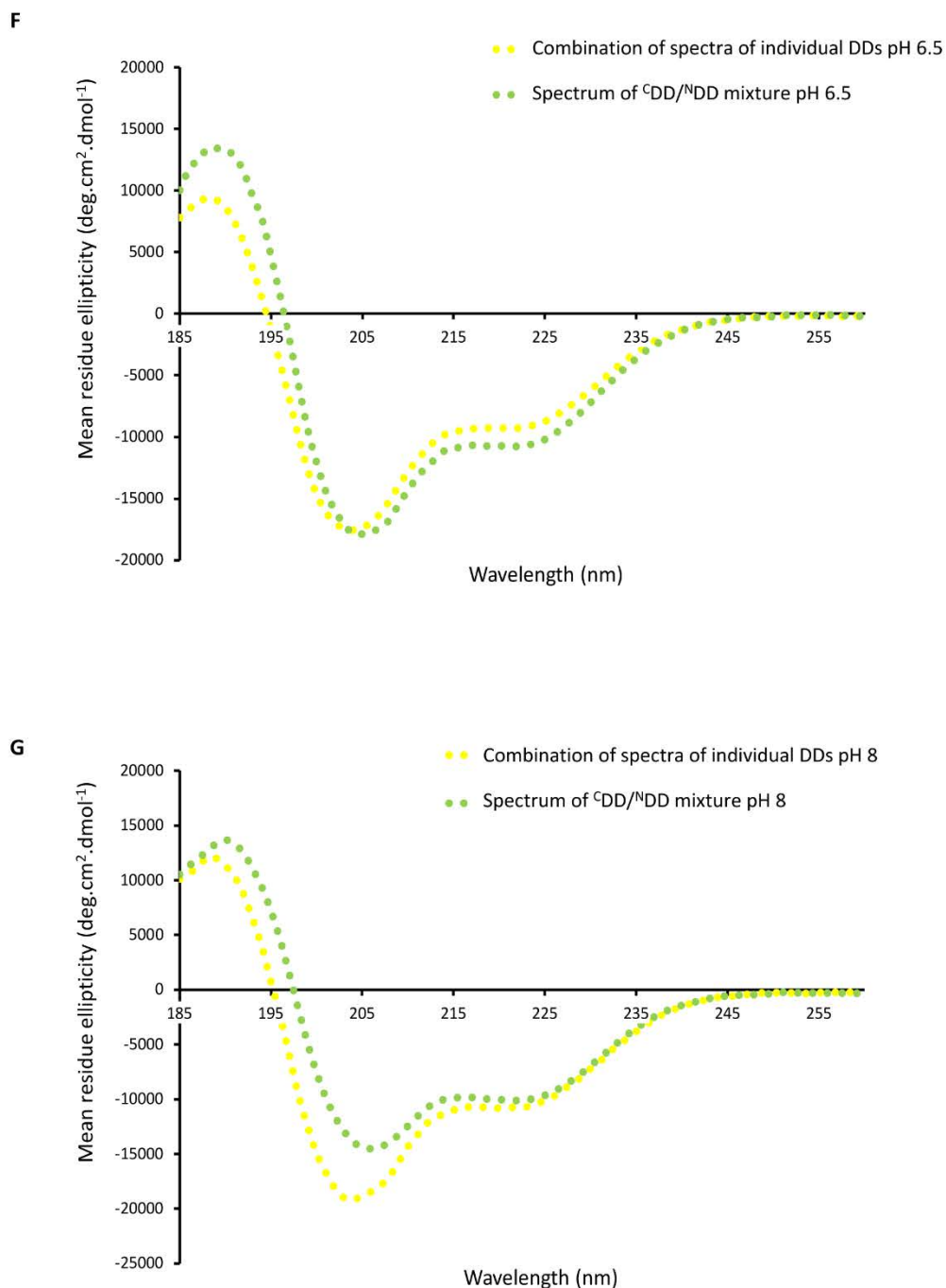


Figure S6, cont. (F) CD spectra showing net structuration upon combining 5923 ¹³DD long and 5922 ¹⁵DD (pH 6.5). A combined spectrum was obtained by simultaneously analyzing the two isolated DDs (both at 100 μ M). Comparison with the spectrum of a mixture of the two docking domains, again at an overall concentration of 100 μ M, provides evidence for increased structuration. (G) CD spectra showing a small extent of structuration upon combining 5923 ¹³DD long and 5922 ¹⁵DD (pH 8). The experiment was performed as in (F).

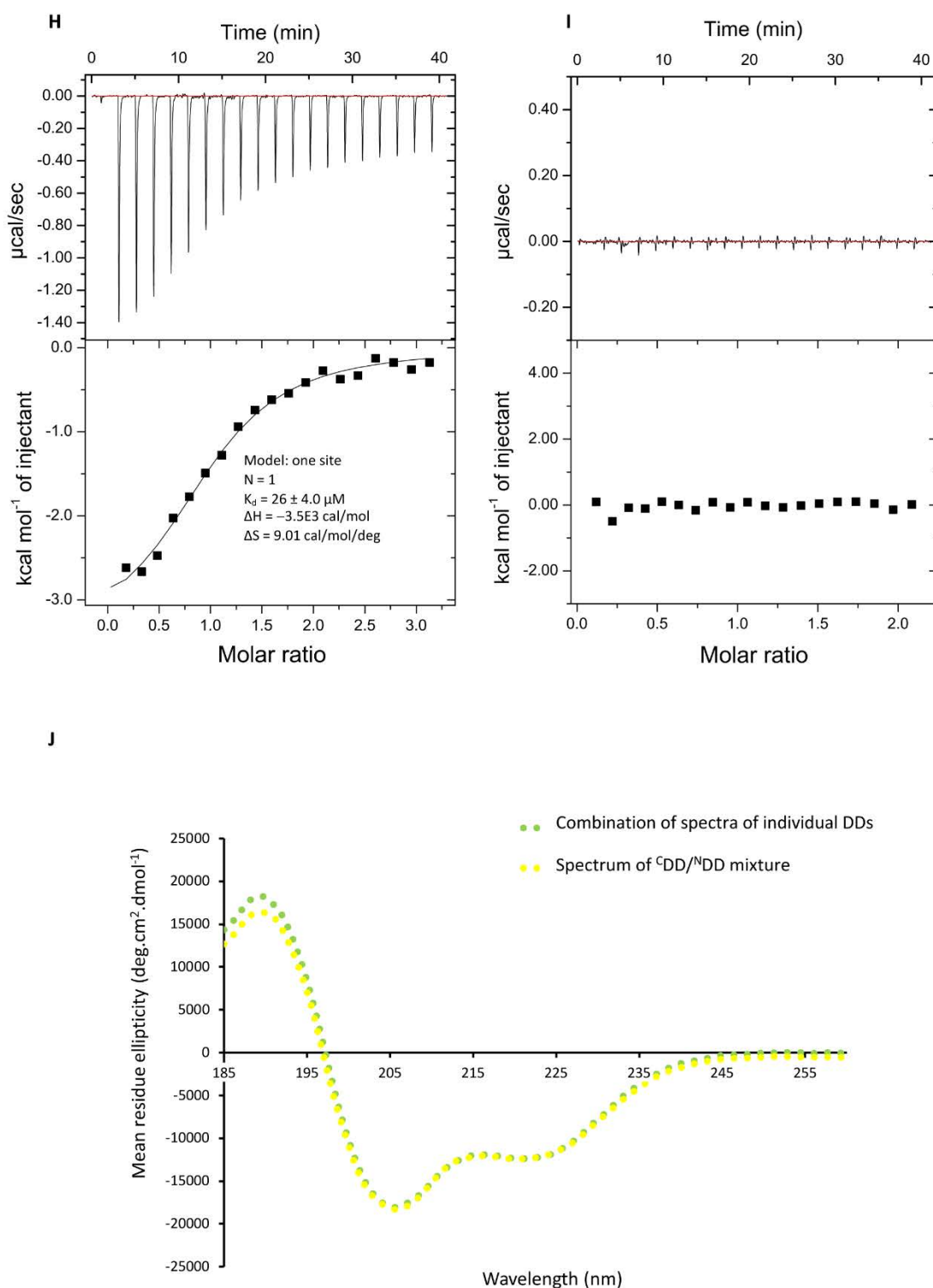


Figure S6, cont. (H) Analysis by ITC of binding between 5923 ¹³DD short and 5922 ¹⁵DD. The error on the K_d represents a single measurement. (I) Analysis by ITC of binding between 5923 ¹³DD long and 5922 ¹⁵DD. (J) Spectra showing the absence of induced structuration upon combining 5923 ¹³DD short and 5922 ¹⁵DD. A combined spectrum was obtained by simultaneously analyzing the isolated domains (at 100 μM). Comparison with the spectrum of a mixture of the two docking domains, again at an overall concentration of 100 μM , shows that no substantial change occurs.

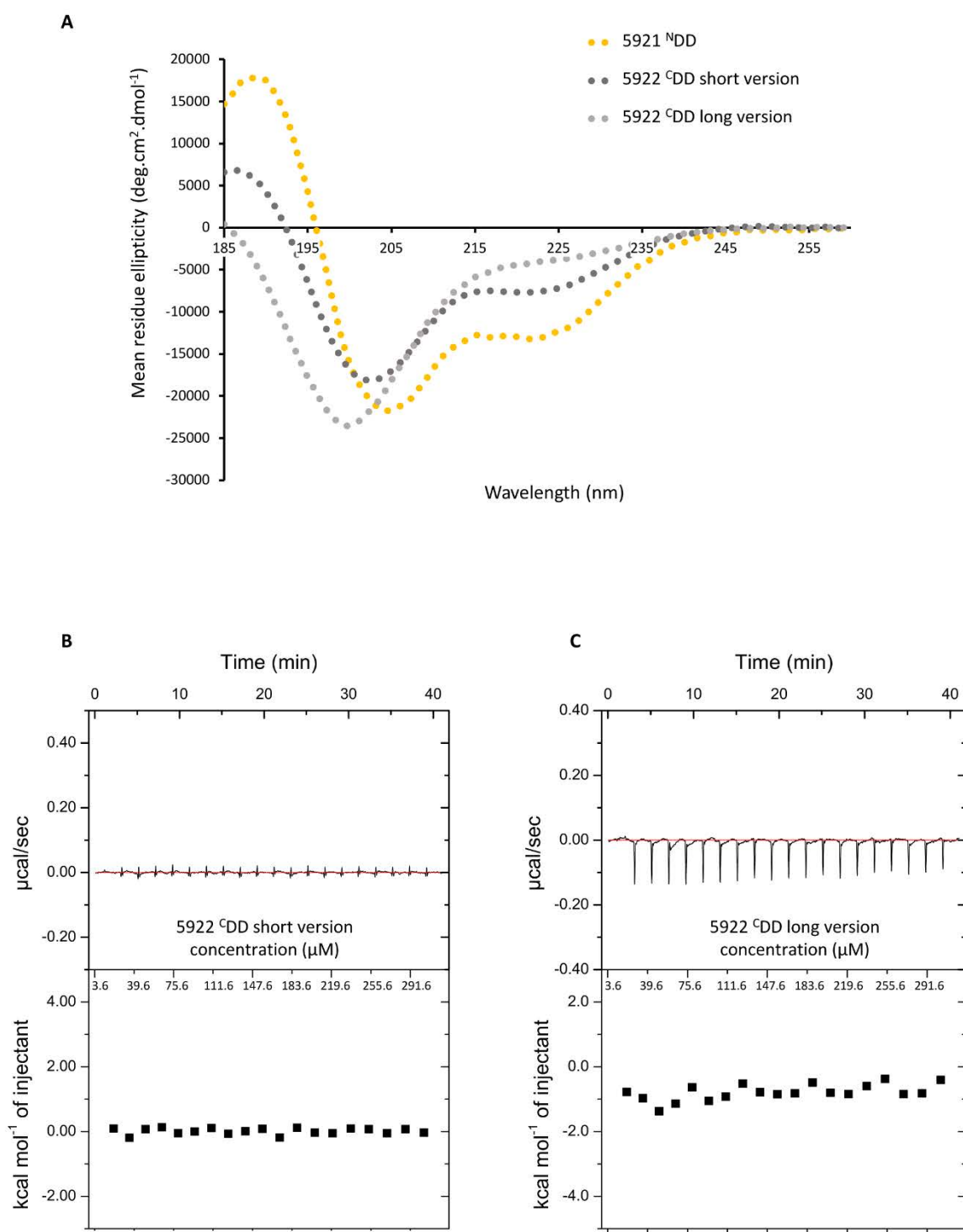


Figure S7. Characterization of the 5922/5921 interface. (A) CD analysis of 5921 ^NDD and 5922 ^CDD long and short versions at 100 μM. (B) Dilution analysis by ITC of 5922 ^CDD short, which gave no change in the observed heat signal. (C) Dilution analysis by ITC of 5923 ^CDD long, which gave no change in the observed heat signal.

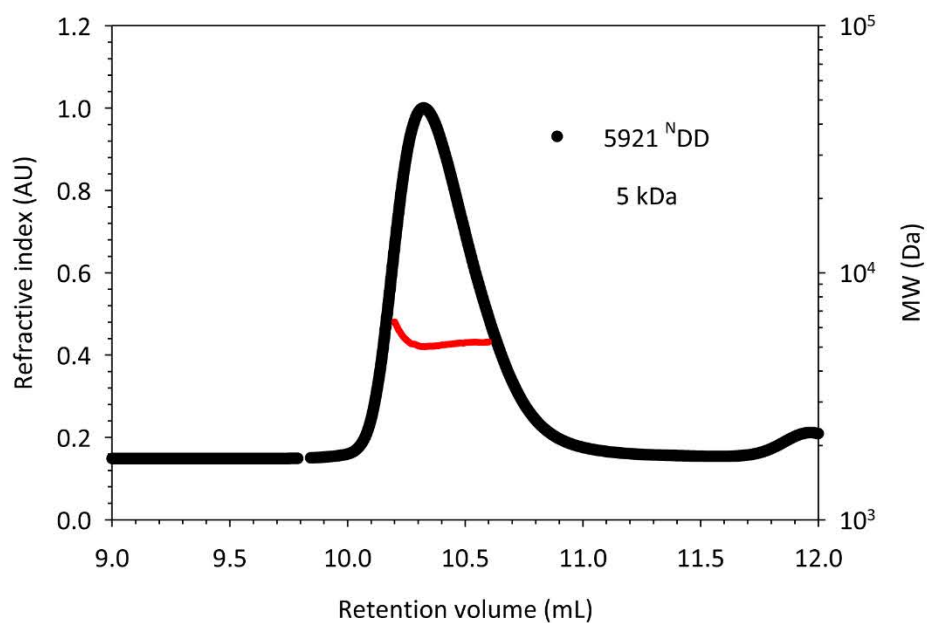
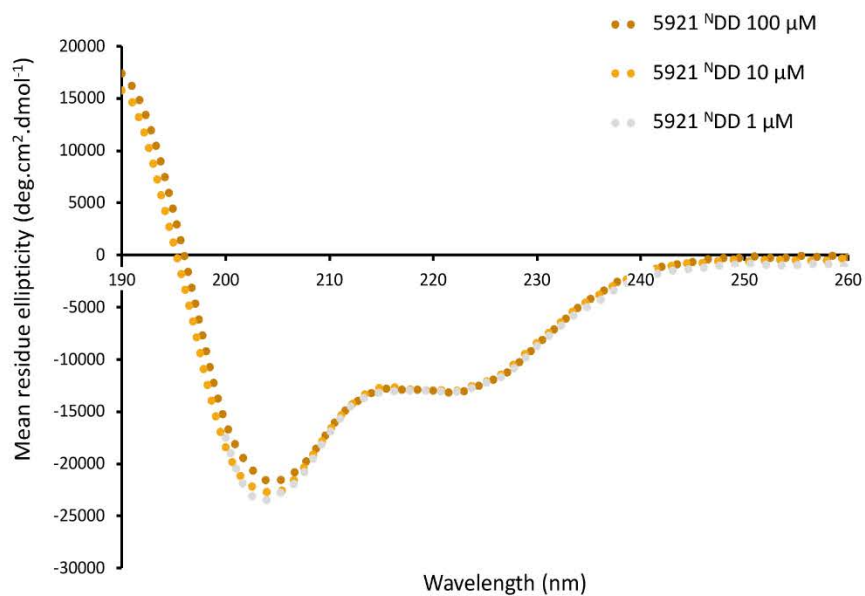
D**E**

Figure S7, cont. (D) Oligomeric state analysis of 5921^NDD by SLS, showing it to be a monomer (Calc'd MW (monomer): 4.2 kDa). (E) CD analysis of 5921^NDD at 100 μM, 10 μM and 1 μM, revealing no significant loss of signal corresponding to α-helix.

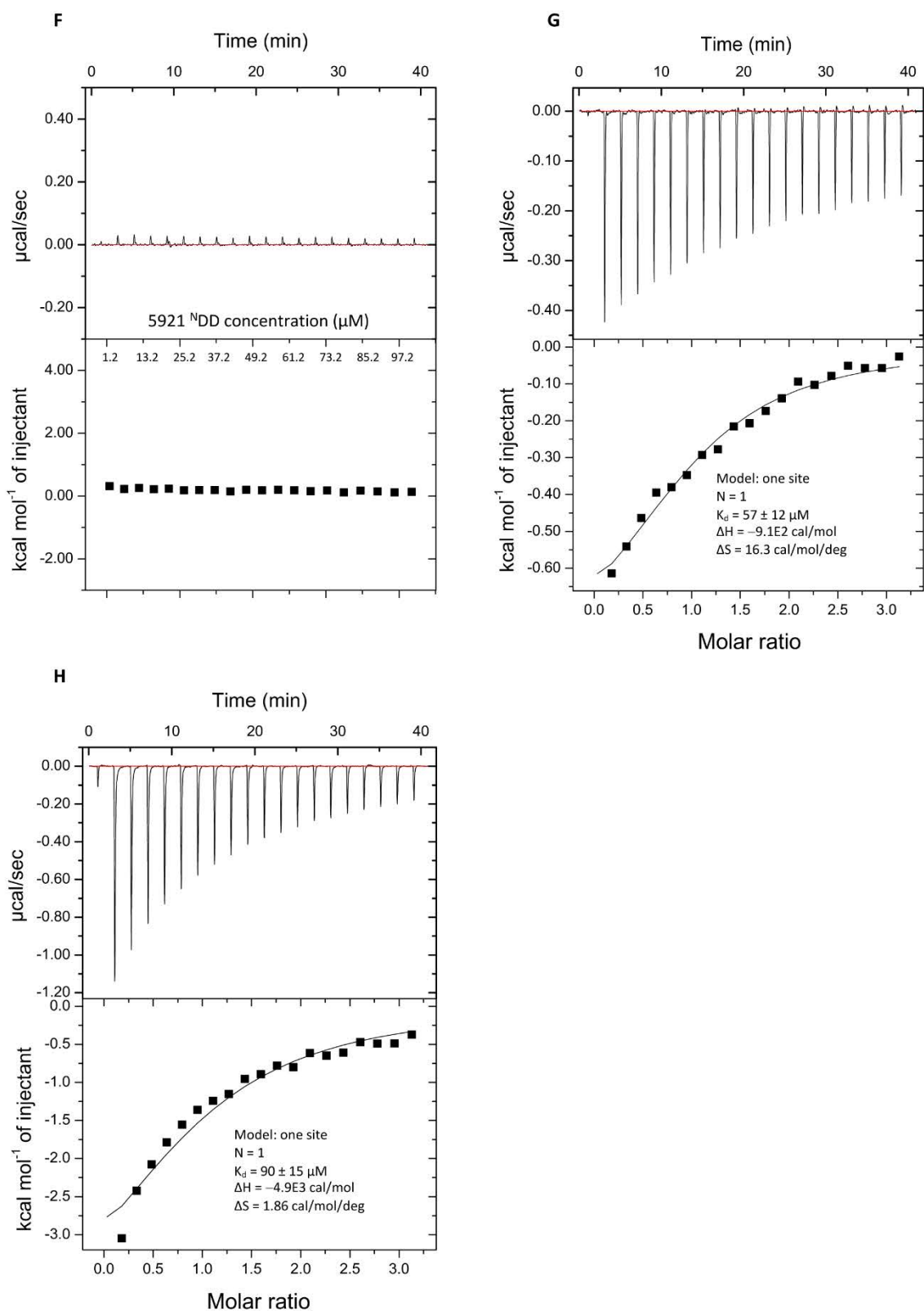


Figure S7, cont. (F) Dilution analysis by ITC of 5921 ^NDD, which gave no change in the observed heat signal. (G) Analysis by ITC of binding between 5922 ^CDD long and 5921 ^NDD. (H) Analysis by ITC of binding between 5922 ^CDD short and 5921 ^NDD. For (G) and (H), the error on the K_d is that of a single measurement.

I

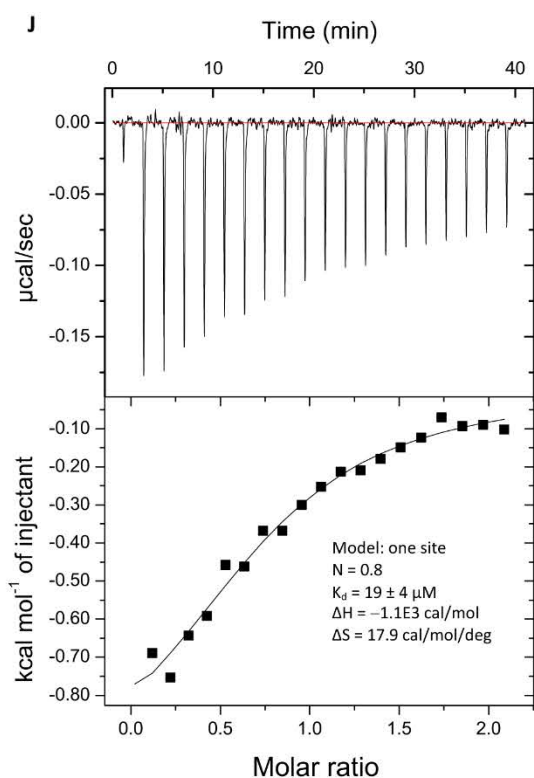
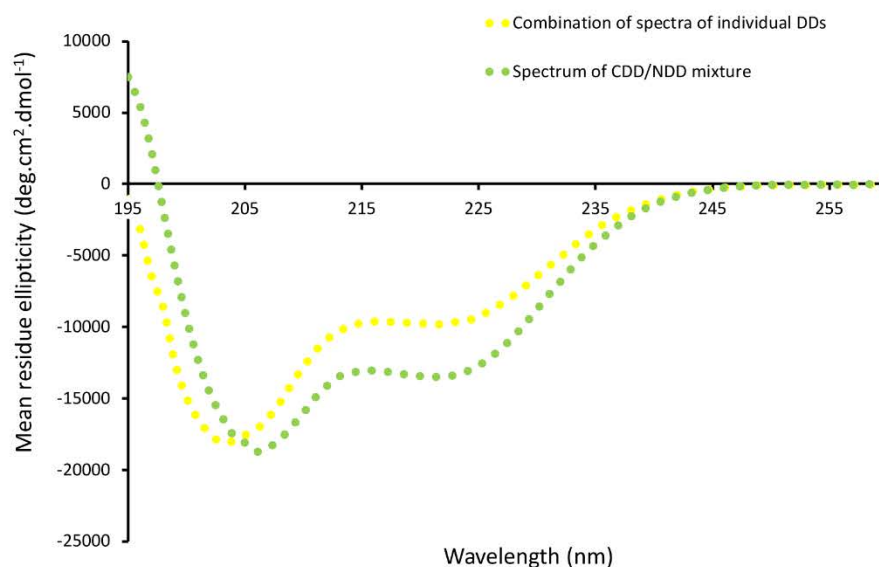


Figure S7, cont. (I) CD spectra showing structuration upon combining 5922 $^{\text{C}}$ DD short and 5921 $^{\text{N}}$ DD. A combined spectrum was obtained by simultaneously analyzing isolated domains (at $100 \mu\text{M}$). Comparison with the spectrum of a mixture of the two docking domains, again at an overall concentration of $100 \mu\text{M}$, shows structuration occurs. (J) Analysis by ITC of binding between 5922 ACP- $^{\text{C}}$ DD and 5921 $^{\text{N}}$ DD. The error on the K_d is that of a single measurement.

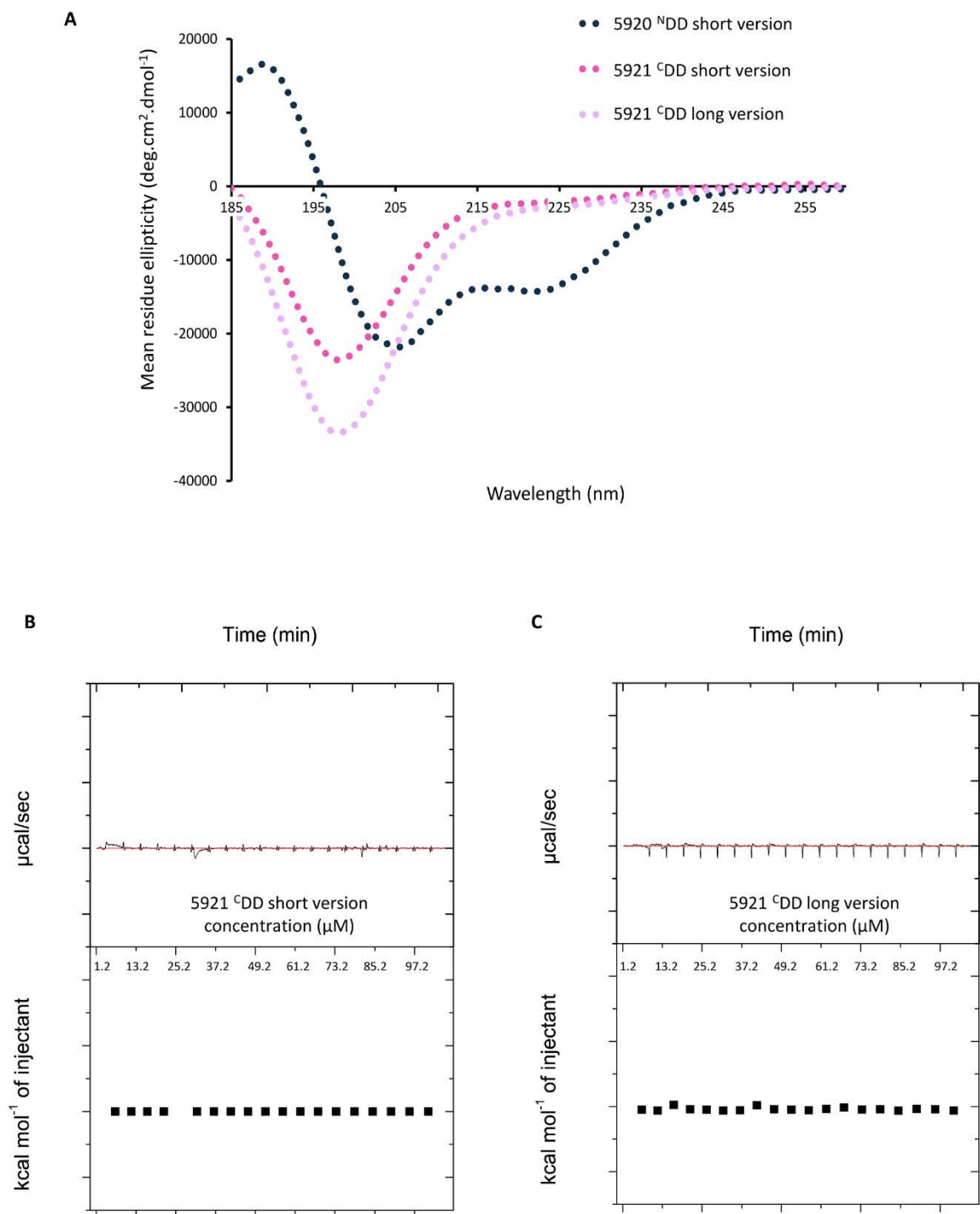
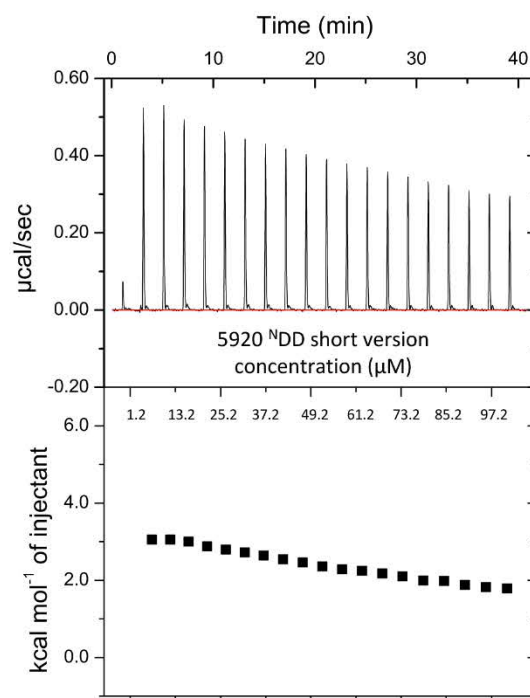


Figure S8. Characterization of 5921/5920 interface. (A) CD analysis of 5920 ^NDD and 5921 ^CDD long and short versions at 100 μM. (B) Dilution analysis by ITC of 5921 ^CDD short, which gave no change in the observed heat signal. (C) Dilution analysis by ITC of 5921 ^CDD long, which gave no change in the observed heat signal.

D



E

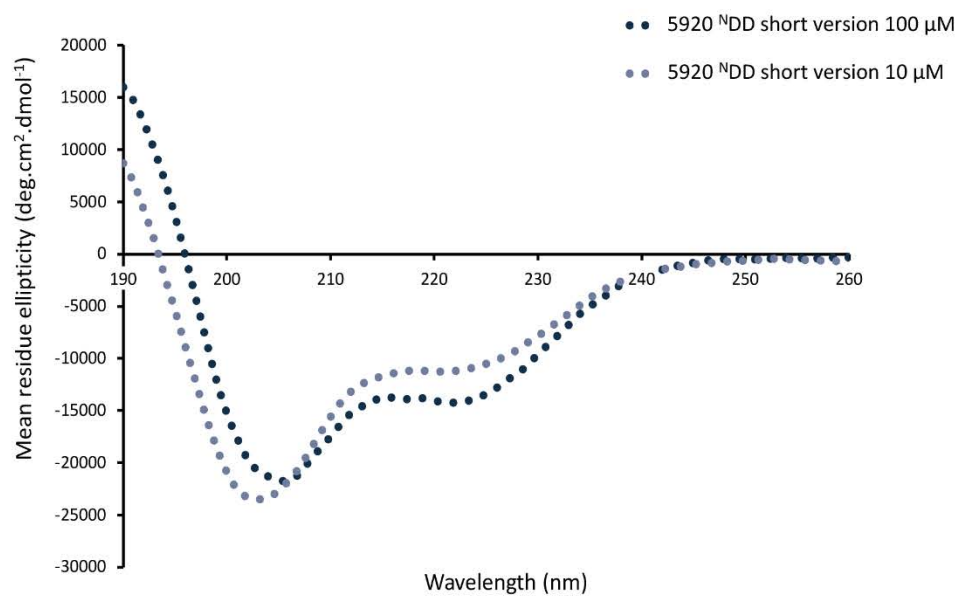


Figure S8, cont. (D) Dilution analysis by ITC of 5920^NDD short, which gave changes in the observed heat signal. (E) CD analysis of 5920^NDD short at concentrations above (100 μM) and below (10 μM) the point at which it is suspected to homodimerize.

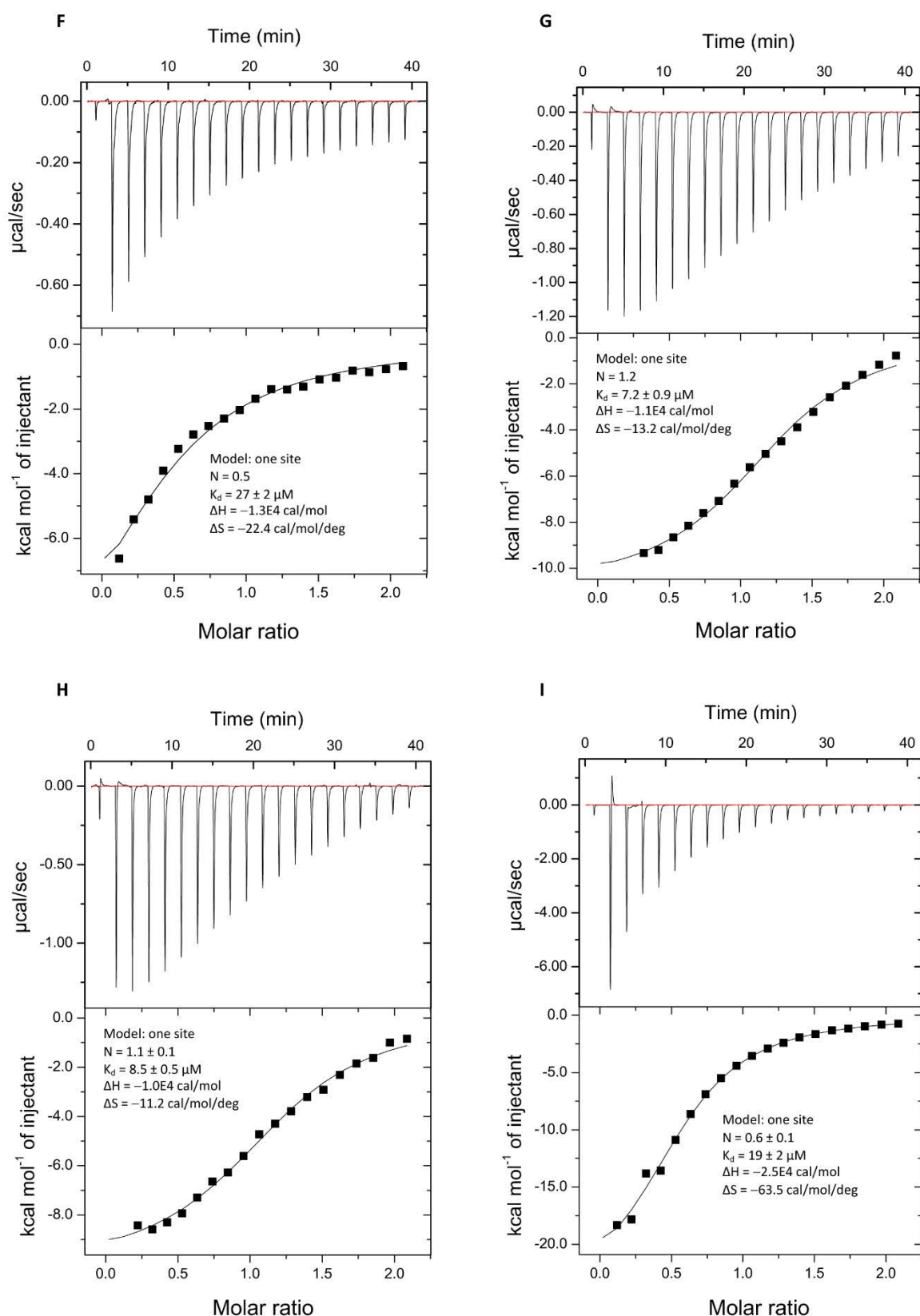


Figure S8, cont. (F) Analysis by ITC of binding between 5921 ^CDD long and 5920 ^NDD long. (G) Analysis by ITC of binding between 5920 ^NDD short and 5921 ^CDD long. (H) Analysis by ITC of binding between 5920 ^NDD short and 5921 ^CDD short. (I) Analysis by ITC of binding between 5921 ^CDD short and 5920 ^NDD short. For (F)–(I), the error on the K_d is that for a single measurement.

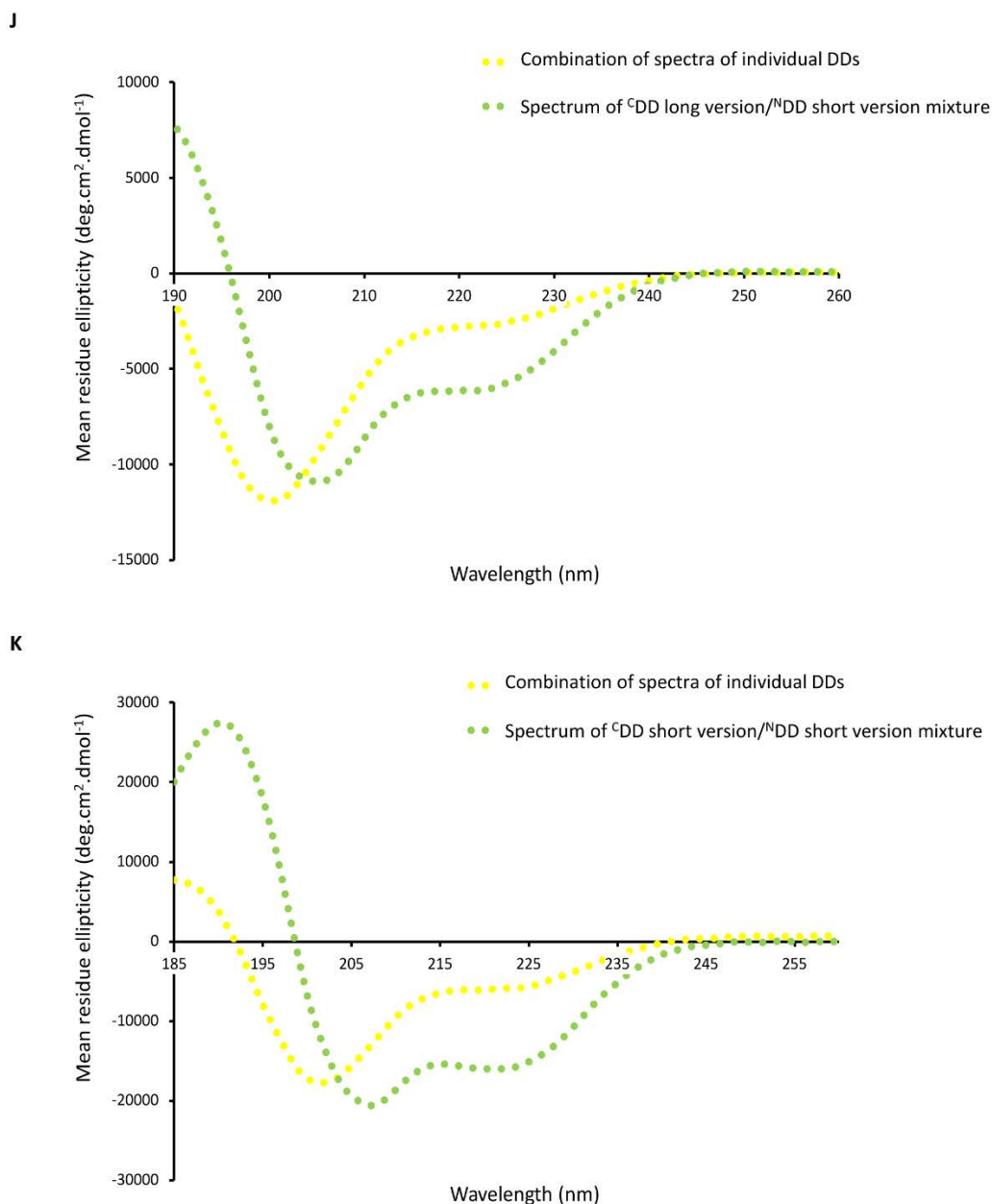


Figure S8, cont. (J) CD spectra showing net structuration upon combining 5921 ^CDD long and 5920 ^NDD short. A combined spectrum was obtained by simultaneously analyzing the two isolated DDs (both at 100 μ M). Comparison with the spectrum of a mixture of the two docking domains, again at an overall concentration of 100 μ M, provides evidence for increased structuration. (K) CD spectra showing structuration upon combining 5921 ^CDD short and 5920 ^NDD short. The experiment was performed as in (J).

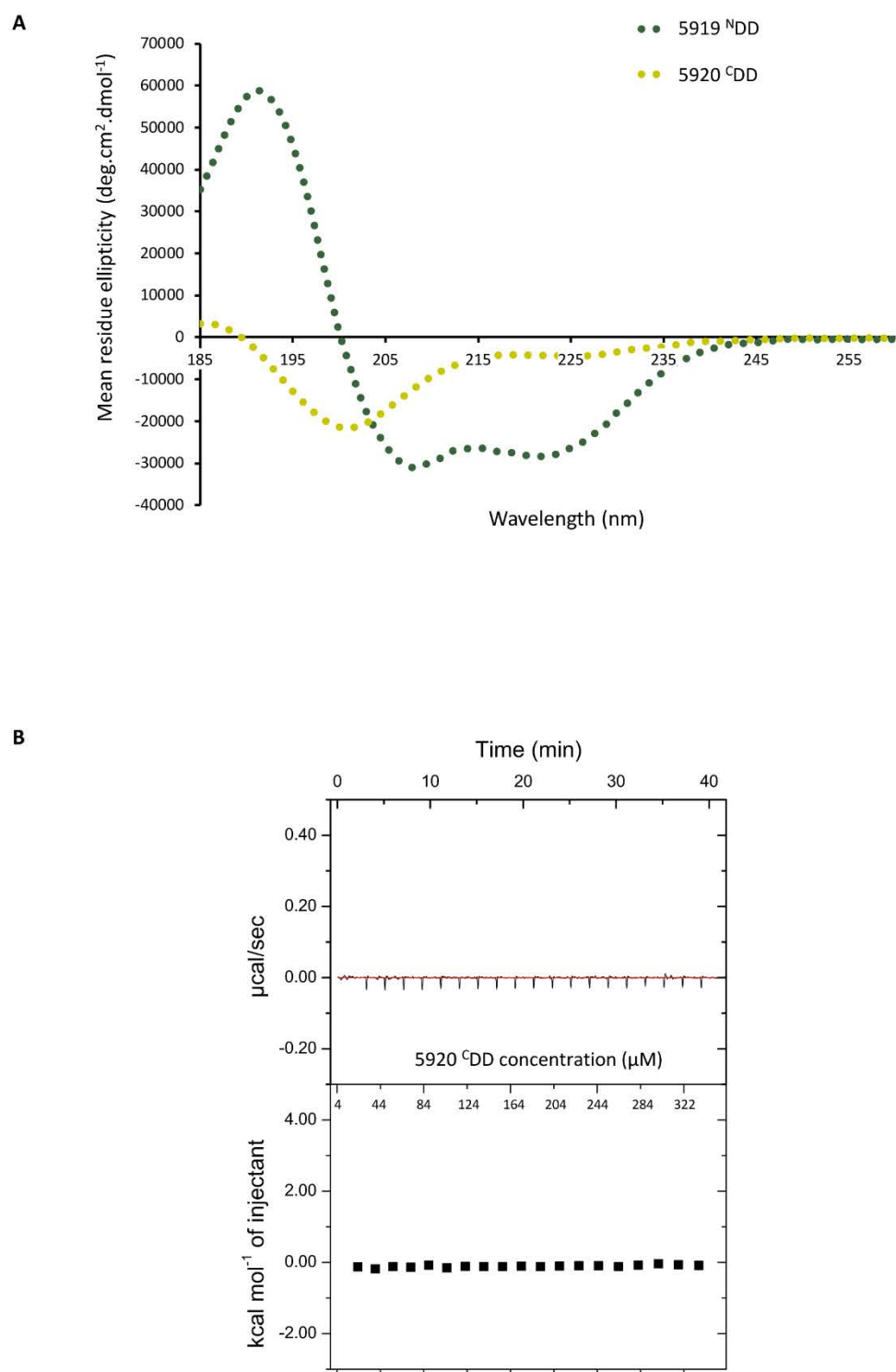
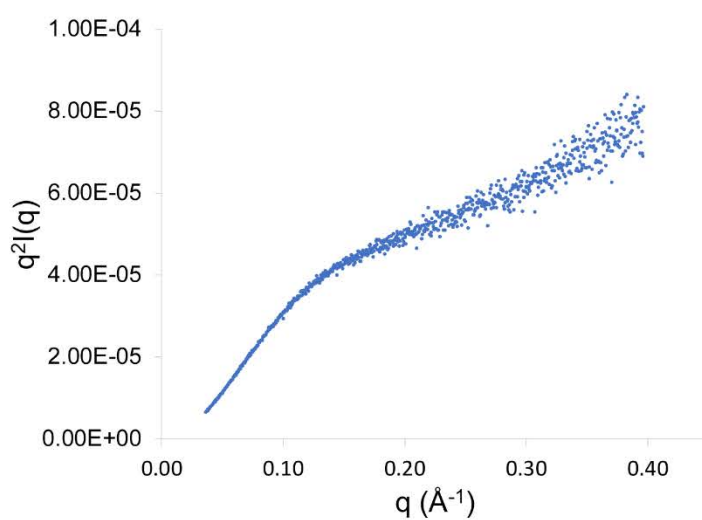


Figure S9. Characterization of the 5920/5919 interface. (A) CD analysis of 5919 ^NDD and 5920 ^CDD at 100 µM. (B) Dilution analysis by ITC of 5920 ^CDD, which gave no change in the observed heat signal.

C



D

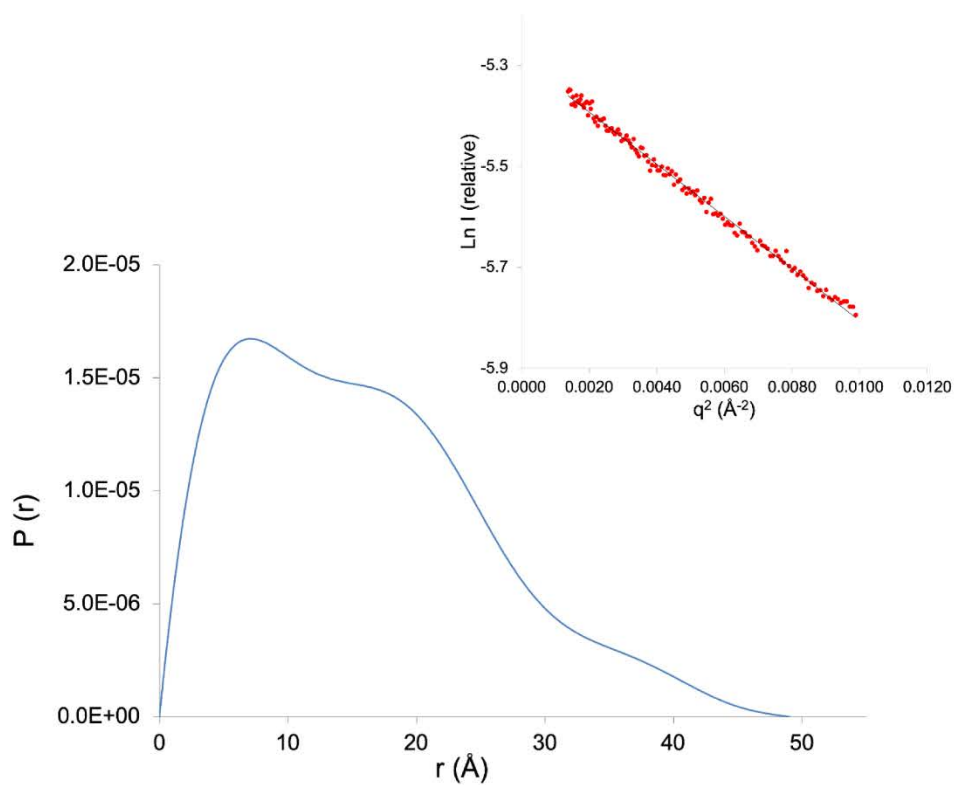


Figure S9, cont. (C) The Kratky plot (blue dots) for 5920 °C DD. (D) The distance distribution function for 5920 °C DD calculated with GNOM.¹² The Guinier plot is inset.

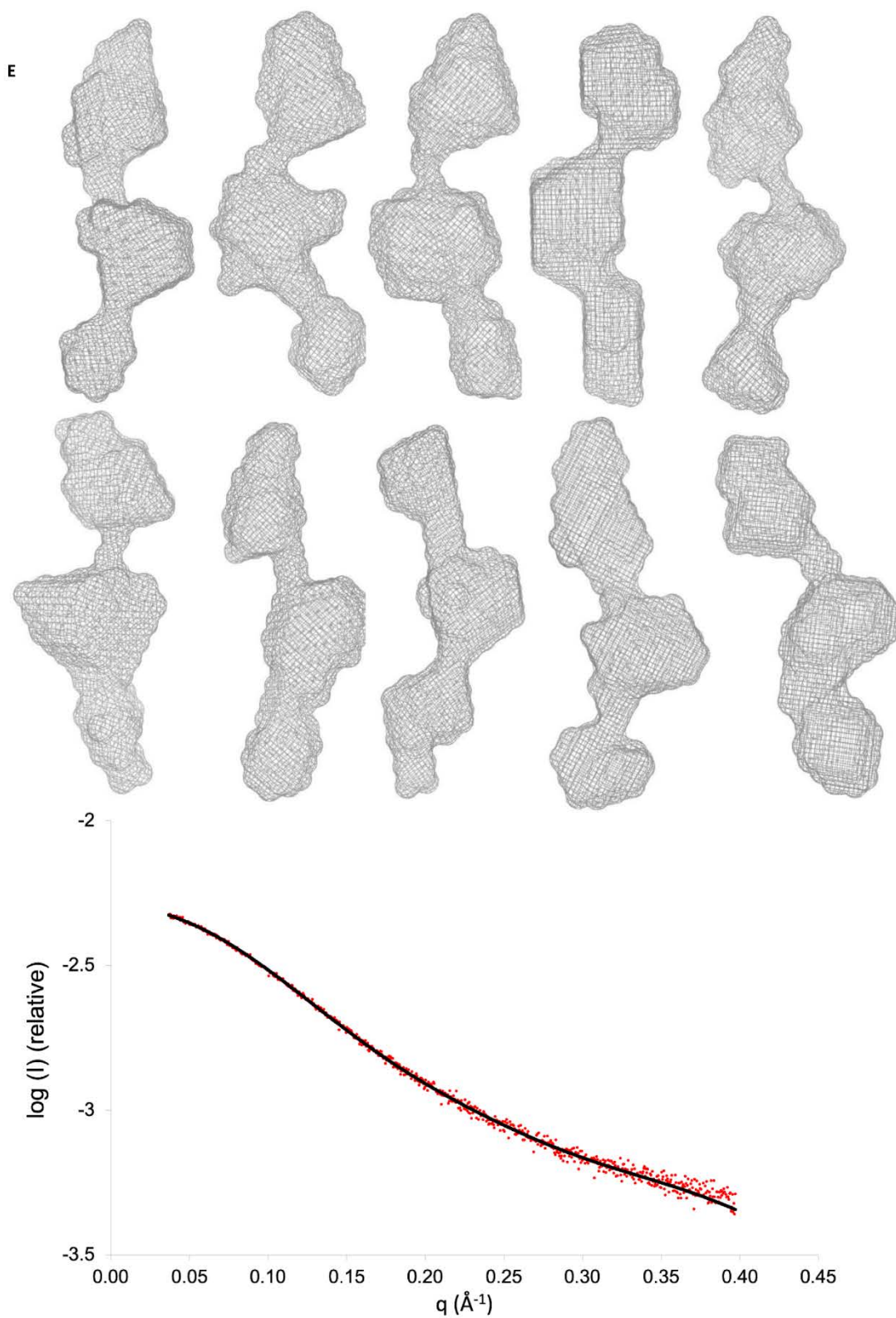
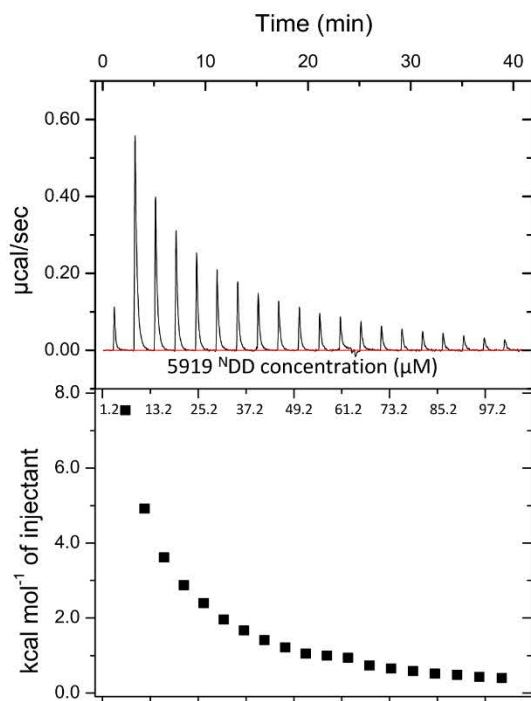


Figure S9, cont. (E) Panel of 10 representative *ab initio* form calculations for 5920 ^6DD calculated with DAMMIN.¹³ The fit between the SAXS data and the 10 form calculations is shown ($\chi^2 = 1.315\text{--}1.344$).

F



G

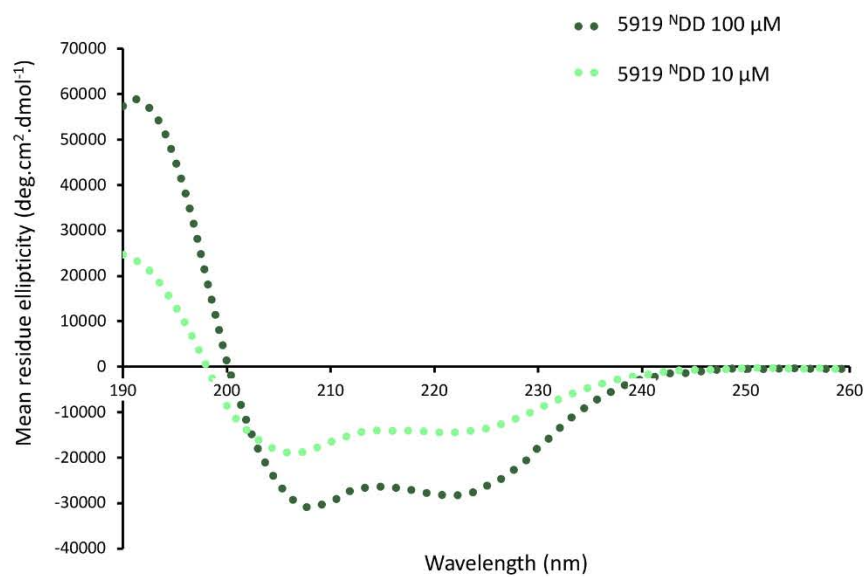
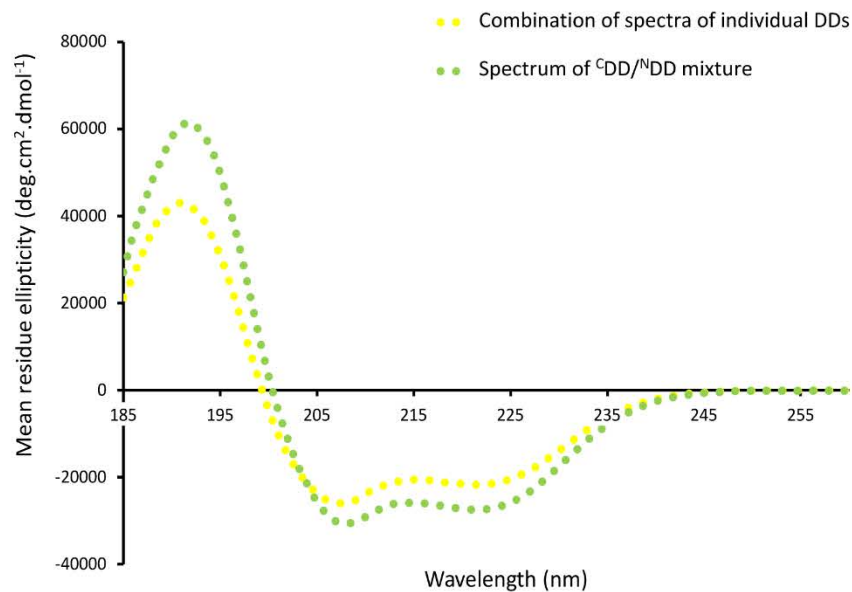


Figure S9, cont. (F) Dilution analysis by ITC of 5919^NDD, which gave changes in the observed heat signal. (G) CD analysis of 5919^NDD short at concentrations above (100 μM) and below (10 μM) the point at which it is suspected to homodimerize.

H



I

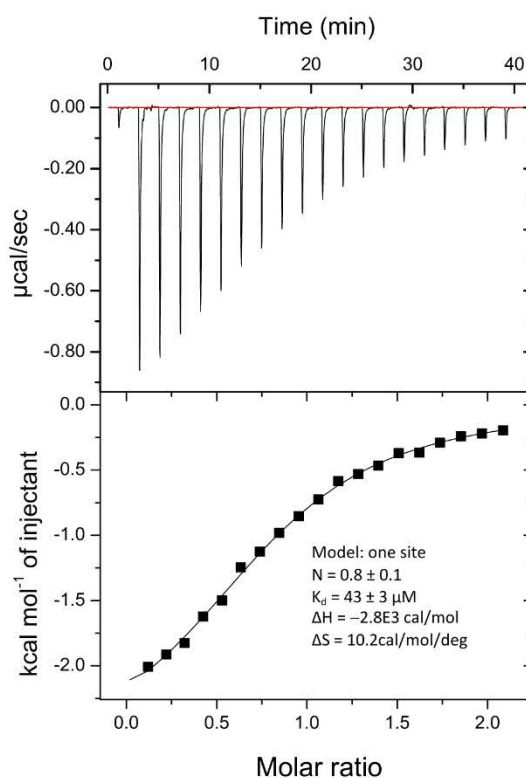
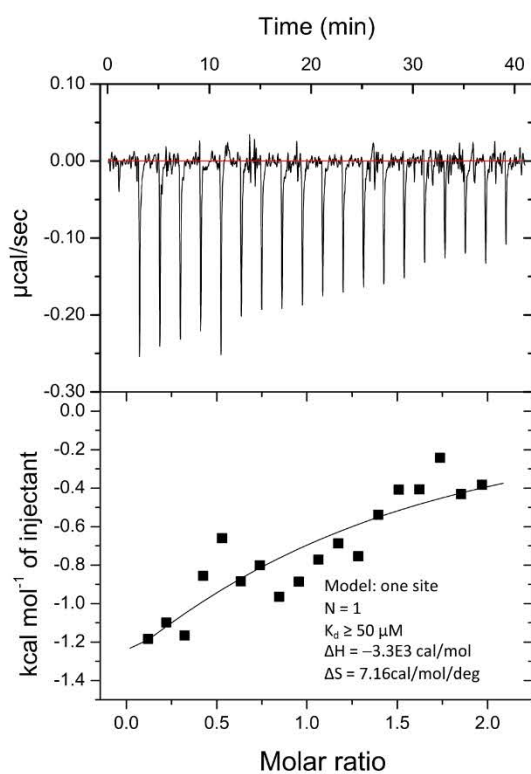


Figure S9, cont. (H) CD spectra showing net structuration upon combining 5920 $^{\text{C}}\text{DD}$ and 5919 $^{\text{N}}\text{DD}$. A combined spectrum was obtained by simultaneously analyzing the two isolated DDs (both at $100 \mu\text{M}$). Comparison with the spectrum of a mixture of the two docking domains, again at an overall concentration of $100 \mu\text{M}$, provides evidence for increased structuration. (I) Analysis by ITC of binding between 5920 $^{\text{C}}\text{DD}$ and 5919 $^{\text{N}}\text{DD}$. The error on the K_d is that for a single measurement.

J



K

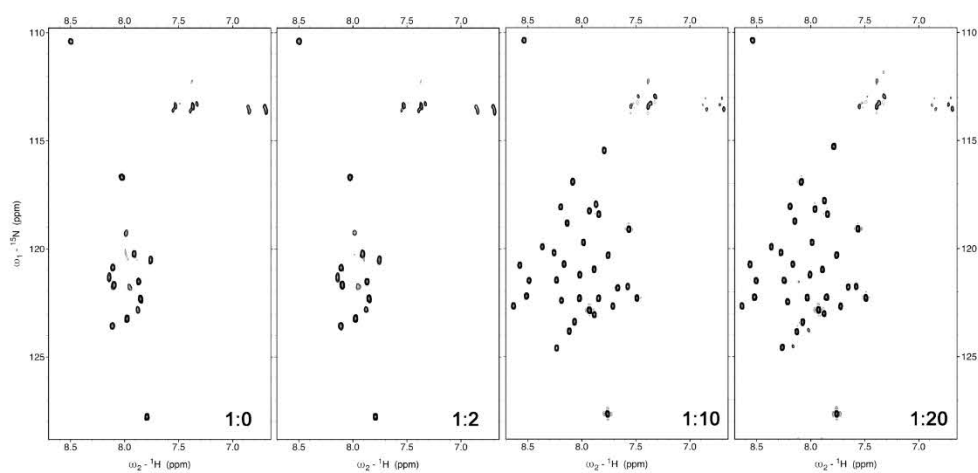


Figure S9, cont. (J) Analysis by ITC of binding between 5919 ¹⁵NDD and 5920 ACP-¹³CDD. The error on the K_d is that for a single measurement. (K) Saturation experiment with ¹⁵N-labelled 5919 ¹⁵NDD in the presence of unlabeled 5920 ¹³CDD.

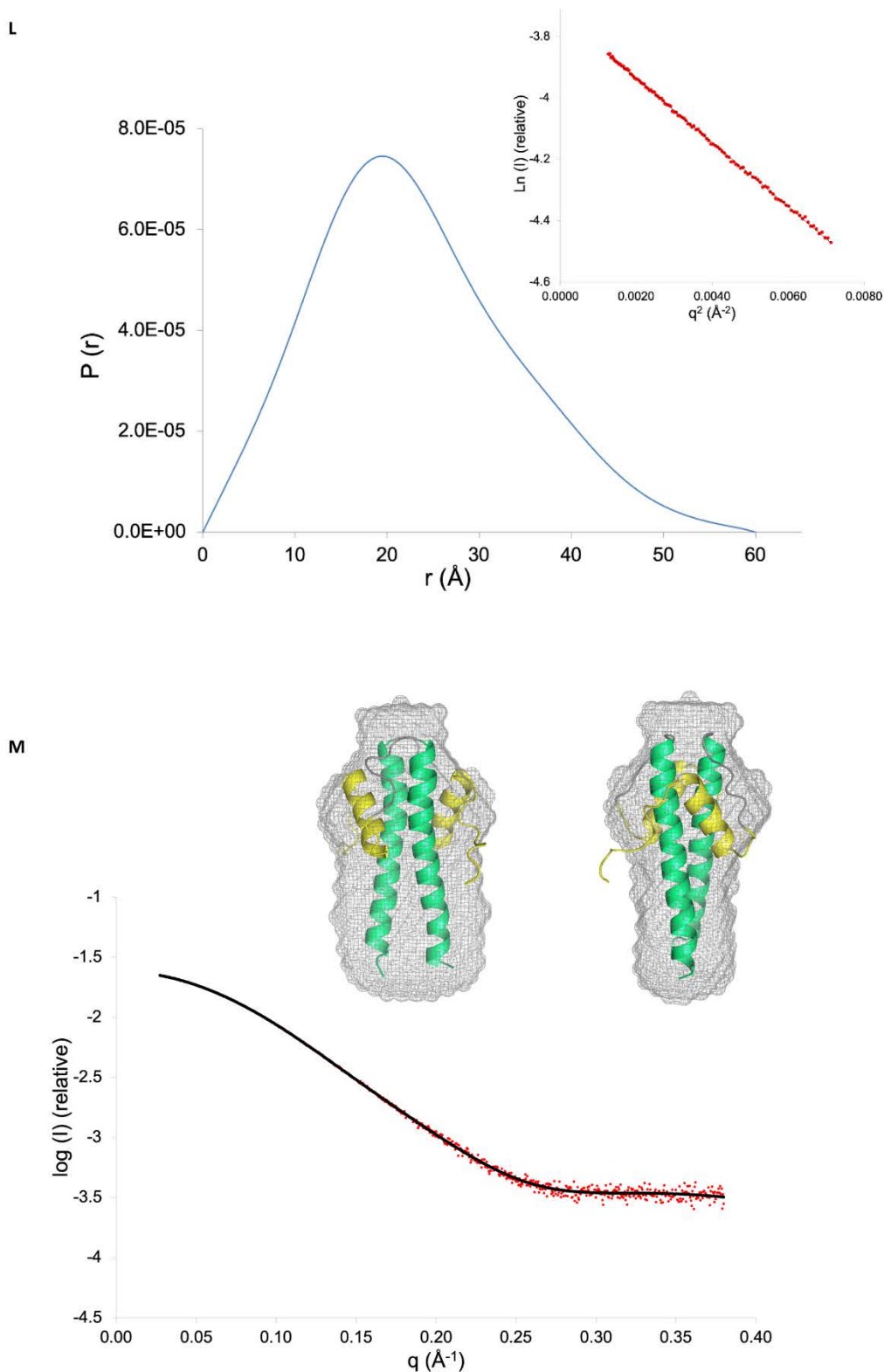
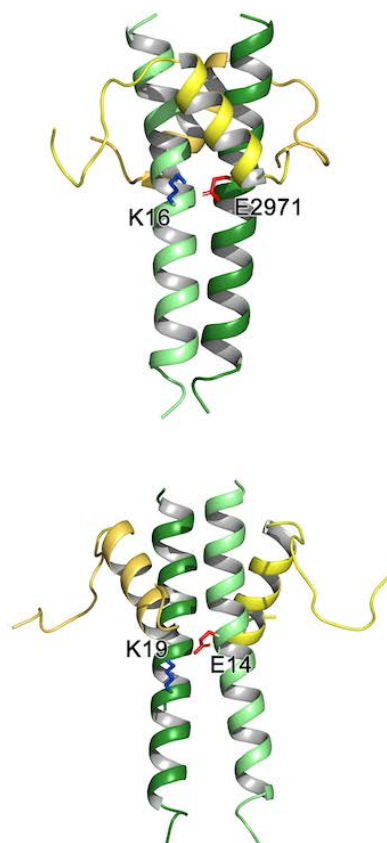


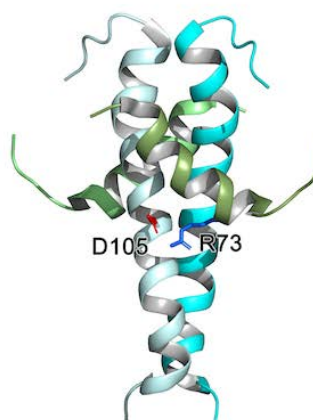
Figure S9, cont. (L) The distance distribution function for the 5920 ^CDD–5919 ^NDD covalent fusion calculated with GNOM. The Guinier plot is inset. (M) Analysis by SAXS of a covalent complex between 5920 ^CDD and 5919 ^NDD. Fit between the *ab initio* model computed with DAMMIN (solid black line) and the experimental SAXS data (red dots) ($\chi^2 = 1.172$). The NMR structure of the fusion complex (5920 ^CDD α -helices in yellow and 5919 ^NDD α -helices in green) was placed into the average molecular envelope (in grey) using SUPCOMB.¹⁶

N

5925 ^CDD–5924 ^NDD



DEBS 2 ^CDD–DEBS 3 ^NDD
Type 1a



PikAIII ^CDD–PikAIV ^NDD
Type 1b

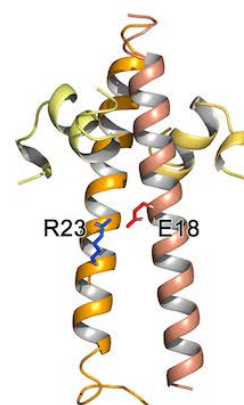
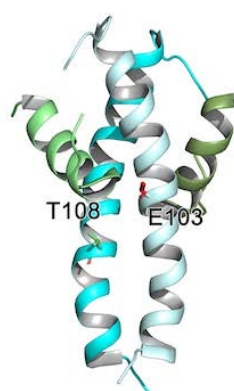
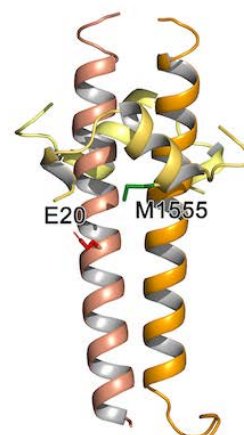
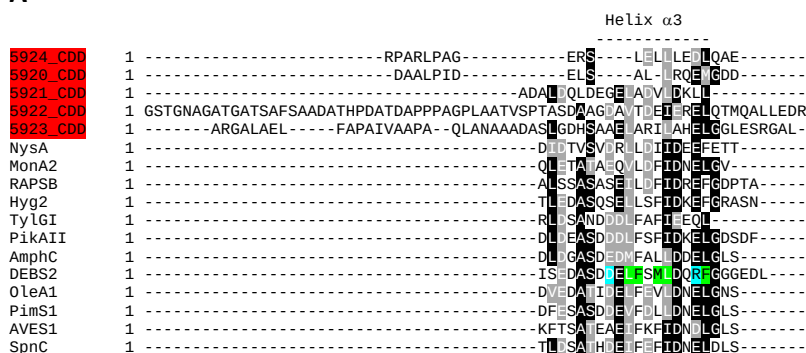
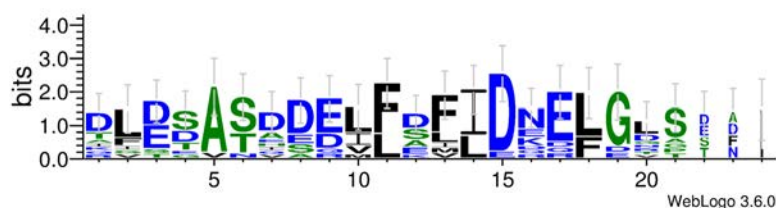


Figure S9, cont. (N) Key charge:charge interactions at the docking interface in the 5920 ^CDD–5919 ^NDD complex, relative to those in the prototypical type 1a (DEBS 2 ^CDD/DEBS 3 ^NDD)²³ and type 1b (PikAIII ^CDD/PikAIV ^NDD)²⁴ interfaces. This comparison shows the 5920 ^CDD/5919 ^NDD interface to exhibit both type 1a and 1b characteristics.

A



Sequence logo for the helix $\alpha 3$ region of type 1a^CDDs:



Alignment of full-length 5922^CDD against full-length type 1a^CDDs:

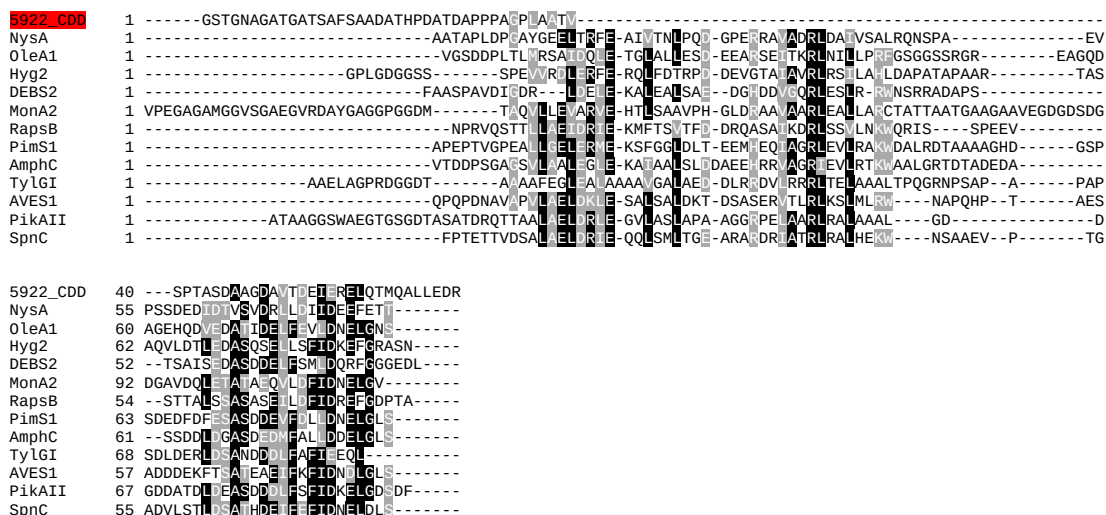


Figure S10. Comparative sequence analysis of the enacyloxin C-termini shown to participate in docking, with ^CDD classes 1a/1b, 2 and 4 α -helix bundle. In each case, we could exclude certain termini from the alignments based on their direct structural characterization or their lengths relative to previously-characterized ^CDDs, as specified below. (A) Alignment of the ^CDDs of subunits 5924–5920 with the interaction region (helix $\alpha 3$) of type 1a^CDDs²³ (based on the length of the enacyloxin ^CDDs, we could already exclude the presence of a four-helix bundle dimerization motif (helices $\alpha 1$ and $\alpha 2$ of this type of DD)). Below is a sequence logo produced from the type 1a^CDDs covering the helix $\alpha 3$ region. The unusually long 5922^CDD has also been aligned with the full-length type 1a^CDDs. Key: **green** = hydrophobic interface residues; **blue** = putative specificity-determinant interface residues. Also indicated are the α -helical regions identified in the solved structures of the fused DD complexes.^{23–26}

B

		Ha3'	Helix $\alpha 3''$	
5924_CDD	1	-----RPARLPAGERSLELL	EDLQAE	-----
5920_CDD	1	-----DAALPDELISAL	LR	-----QEIGDD
PikAIII	1	-----SIDDLDL	EL	-----RMAIGPRNT
NysB	1	-----SIDDMDAEAL		-----RLATENSAN
AmphB	1	-----DIDMDAEAL		-----RLAAENSAN
5922_CDD	1	-----	GSTGNAGATGA	-----TSAFSAADATHPDATDAPPPAGPLAATVSPTASDAAGDAVTDE
RapsA	1	-----FIDEMDADAL	KHVLEER	-----
Aves2	1	-----SIDEMDIDL	DLAHGHTAPAREPADAEDSSSSRNRTTH	-----THEGETA
Hyg1	1	-----LIDAMDTEAL	RHV	-----DGTGA
TyIGIV	1	-----LDDLIDGDAL	RLAIGEPGE	-----
NysI	1	-----SIDSMIVADL	RAALNGQSDL	-----
RifA	1	-----LIDALDISGL	QRAIGQTS	-----
PimS2	1	-----SIDAMDIAL	QAAFDGNPQER	-----
SpnA	1	-----SIDAMDVAGL	EAAIGERES	-----
ConA	1	-----LIDAMDVAEL	RMAREGIES	-----
5923_CDD	1	ARGALAEFAPAIVAAPQLANAAADAS	GDHSAELARILAHELGGLESRGAL	-----
5921_CDD	1	-----ADAQIDEGELADVLDKL		-----

5924_CDD	-----
5920_CDD	-----
PikAIII	-----
NysB	-----
AmphB	-----
5922_CDD	55 IERELQTMQALLED
RapsA	-----
Aves2	-----
Hyg1	-----
TyIGIV	-----
NysI	-----
RifA	-----
PimS2	-----
SpnA	-----
ConA	-----
5923_CDD	-----
5921_CDD	-----

		Ha3'	Helix $\alpha 3''$	
5923_CDD	1	-----ARGALAEFAPAIVAAPQLANAAADAS	GDHSAELARILAHELGGLESRGAL	-----
5922_CDD	1	-----	PAVTDELIERELQTMQALLED	-----
5924_CDD	1	-----RP	ARLPAGER	-----SL
5921_CDD	1	-----AGPAQAARAPASAPT	PRAAPADA	-----LDQIDEG
5920_CDD	1	-----LERLAAQARQA	PPGPAGQAAPADA	-----ALPIDE
PikAIII	1	-----	SIDDLDL	EL
NysB	1	-----	SIDDMDAEAL	RLATENSAN
AmphB	1	-----	DIDMDAEAL	RLAAENSAN
RapsA	1	-----	FIDEMDADAL	KHVLEER
Aves2	1	-----	SIDEMDIDL	DLAHGHTAPAREPADAEDSSSSRNRTTH
Hyg1	1	-----	LIDAMDTEAL	RHV
TyIGIV	1	-----	LDDLIDGDAL	RLAIGEPGE
NysI	1	-----	SIDSMIVADL	RAALNGQSDL
RifA	1	-----	LIDALDISGL	QRAIGQTS
PimS2	1	-----	SIDAMDIAL	QAAFDGNPQER
SpnA	1	-----	SIDAMDVAGL	EAAIGERES
ConA	1	-----	LIDAMDVAEL	RMAREGIES

(Here, 5922_CDD has been truncated to just the C-terminal predicted α -helix)

Sequence logo for the helix $\alpha 3$ region of type 1b c DDs:

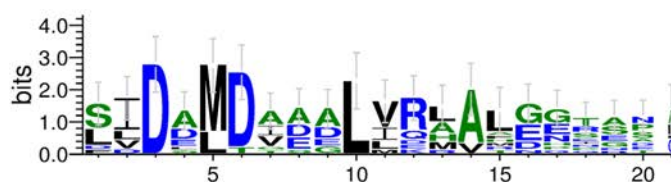


Figure S10, cont. (B) Alignment of the c DDs of subunits 5924–5920 with the interaction region (helix $\alpha 3$) of type 1b c DDs²⁴ (based on the length of the enacyloxin c DDs, we could already exclude the presence of a four-helix bundle dimerization motif (helices $\alpha 1$ and $\alpha 2$)). Below is a sequence logo produced from the type 1b c DDs covering the helix $\alpha 3$ region. 5922 c DD has also been aligned with the full-length type 1a c DDs (see next page).

[illegible]

		Ha3'	Helix	$\alpha 3''$
5922_CDP	49	DAVTDE	RE	LQTQALLDR
ConA	70	ELTDANDVAE	L	RLRARGIES
NysB	61	SESDDMDA	EA	LRLATNSAN
AmphB	53	ALDDMDA	EA	LRLAANSAN
RifA	62	AEITDA	DTSC	LQRALGQTS
Hyg1	59	GLIDAMDTE	L	RHVIGTGGA
AVES2	74	PVSIDEMDIDD	L	DLAUGHTAPAREPADAEDSSSRNRTHHGETA
RapsA	86	AFDDMDA	DA	LRLHVLFEER
TylGIV	56	ALDDLIDGA	EA	LRLALGEPGE
PikAIII	63	ASIDDD	AE	LRYALGP RNT
SpnA	57	ATSDAMDVAE	L	EALGERES
NysI	51	GESDSDVITAD	L	RLALNGQSDL
PjmS2	67	FESDAMDITAD	L	QAEEGNSPQFR

			Helix α_1	Helix α_2
StiA	1	-----	AAMLMKLEQFQLSNEILLGLSDDDKEDEVSS	-----
NidA4	1	-----	RPADAGSDALSPAEILIRLAKTGAGQ	-----
5923_CDB	1	ARGALAE	LFPATVAAPQAQLN	AADASGLHGAELARLAHLEGGLESRGAL
AjuB	1	-----	TALPAPSLENLSAEISDLLAAEL	SASAALMGPGMD
EpoC	1	-----	HTTDSAVETLEISQDDILITQLAAKFKAIT	-----
MelD	1	EAEPSLEN	ENQLPQELGALLDQKLADLEKLVGDA	-----
MxaD	1	-----	ALETEINNKQLSDAEALAEKLAALAE	-----
AufI	1	-----	EGQERLAEVEQSDAEQLLLASLESLTELK	-----
DknX	1	-----	AWDAGEDQLDALDQVLDLDDQQLARRRG	-----
5925_CDB	1	-----	APLDTLESTLE	EGQDDDLAALLGKEFTIRE
CyrD	1	-----	SSIIRIDNETDVLPSLEQSDDESLSTIQKLSTHLD	-----
CurK	1	-----	DGSSQSDITLSLELEASLQLEALEKLI	-----
CurG	1	-----	SLEKLIDITKQSNQLEESSDITLESIN	-----
JamJ	1	-----	QKOEKLATKLSQLEETNOELNLIN	-----

			Helix α1		Helix α2
			-----		-----
DkxN	1		-AWDAGEDQELDAL-	AVDLDLDDQQQAIRRRG-	
5925_Cdo	1		-APLDTET-	SEI-EGQDDDAALLGKEFIRE-	
CyrD	1		-SSTTRIT-	DNETDYLDSEQSDEDESLSLIQKLSTHLD-	
CurK	1		-DGSSOL-	SDTILSEIELEASVLQEFALEKLT-	
CurG	1		-SLEAKLL-	DIKOSNQNELESSIDQLLESIN-	
JamJ	1		-QKQEKLL-	ATKLSLEEQUEELNQEQLNLIN-	
5922_Cdo	1	GSTGNAGATGGSAFSAADATHPDATDAPPAGPLAATVSPTASDAQAVTDIERERISQTMAALLIEDR-			
StIA	1		-AAMLKLV-	EQFGLSNLGLIGLGDCKDEVSS-	
NidA4	1		-RPADA-	GEDALSPAEILRLAKTGAGQ-	
5923_Cdo	1	-ARGALAEI-	FAPAIVAAPAQLANLAAD-	ASGHSAALARLAHELGGLESRGAL-	
AjuB	1		-TALPAL-	PSLENISEALSIDLAAELASAAMLMGPMD-	
EpoC	1		-HTTDSA-	VTEEMSDYLUTLLAAKFKAIT-	
MelD	1		-EAEPSSL-	ENIQPOLUFGALDQKLADLEKLVGDA-	
MxaD	1		-ALEAEIV-	NNKQLSPDAEALAEKLAALAE-	
AufT	1		-EGQFRV-	AEVFQSDPAFOFLASIESISTEILK-	

60

D

		Helix α 1	Helix α 2
TstDEF	1	-----FDEDLYASVIDAL-----	-----DRTLDSTAWQTVRSGR-----
SorB	1	-----VNGGRLSAVLDAL-----	-----LADEIDHAAART-RALLAGGLS-----
5922_CDD	1	GSTGNAGATGATSAFSAADATHPDATDAPPPAGPLAATVSPTAS-----	-----DAAGDAVTDEIEREL--QT--QALLIEDR-----
VirA	1	-----PSQADIDALIS-----	-----AVRDNRIS--EQA-VTLTPRR-----
Diff	1	-----SPLETYERLLDELEKNEIS--EAAVKRT-IGFYRD-----	-----
PtzC	1	-----FSAMSMENLLKSLETGEIDKQAYQVD-INQDA-----	-----
ElaQ	1	-----RDDKEYYDLDQAVLSELSVQHATNEI-VKML-----	-----
RizD	1	-----DEDFETLALRGLSQCALSTEVDAR--TAGIKQ-----	-----
AtcE	1	-----ERATDVEDLRRLLAAGELTVAAYTN--GGPA-----	-----
ChiB	1	-----EGDEPLWRLLRGVESCAIDNACRL--ELP-----	-----
MlnD	1	-----TADGEIMS-----	-----LLEKAQTGEINTKETSEL--EELLFHE-----
5923_CDD	1	-----ARGALAE LFAPAIVAAPAQLAN-AAADASLGHSAA--LARI--AHELGGLESRGAL-----	-----
OzmH	1	-----APDELIAKLEARFAAGELSAAEVLDL--DAELATRERR-----	-----
RhiC	1	-----PPHPLDHQLEKQFLSGELSDSLNLV--SIGTVER-----	-----

Sequence logo for the helix α 1 and α 2 regions of 4 α -helix bundle type $^{\text{CDD}}$ s:

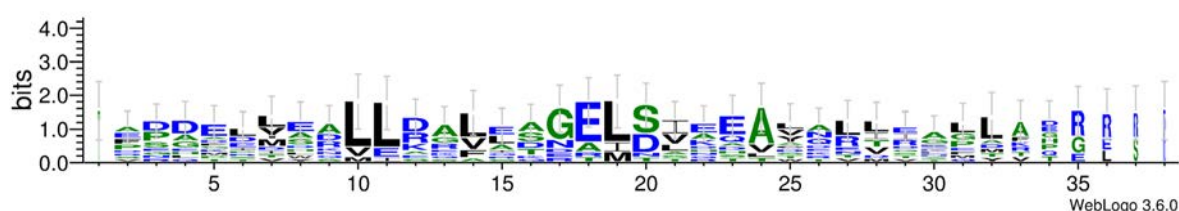


Figure S10, cont. (D) Alignment of the $^{\text{CDD}}$ of subunits 5923 and 5922 with the interaction regions (helices α 1 and α 2) of 4 α -helix bundle type $^{\text{CDD}}$ s (the $^{\text{CDD}}$ of subunit 5921 is too short to be this type).^{26,27} Below is a sequence logo produced from the 4 α -helix bundle $^{\text{CDD}}$ s covering the helices α 1 and α 2 regions.

Coiled-coil

```

SpnD      1  ----TVTTSYELVALLASLKNERLRGRDRFSAEKD-----
AmphD     1  ----DNEQKIVLYKLALV----LRARRRRGVEESAQ-----
OleA2     1  ----TNDKIVLYKLRAIV----LRKARRHWELED-----
NysB      1  QEPQQGQPDQQLKIVLYLVRVSG----LRARRRRGVEESKDN-----
RAPSC     1  ----PEQIKIVLYLEWVA----LITRALEEAANA-----
DEBS3     1  ----SGDNGMTEELKRRYLRKRLTE----LDSVTARLREHIRAGE-----
TyLGI1    1  ---------A----LVATRRKLGALEERAR-----
Hvg3      1  ----ADSKTITVLYKMWVA----LYRTREARLSEASA-----
PiKAI1    1  ----ANNEKIVLYLKRVA----LQONTRRRLREGRTH-----
Pim5      1  ----SNEKIVLYLKRATIA----LETROODLETAQKR-----
MonA3     1  ----VSEKIVLYLKRVA----LATRRRLREARERQ-----
AVES2     1  ----QLANEKIVLYLKRVA----LDRTRRLRYEREQ-----
5923_NDD  1  ----GRTVAHFLDRLEQLSKPQLQ----L-----ALARAMRDEILQLR-----
5919_NDD  1  ---------QDQLLAKSLTL-----LRLAANQAQEARRE-----
5921_NDD  1  ----DSKDRELLAQRAIGIQIK-----LQALSARLDQ-----
5922_NDD  1  ----TMDKLLQSQIT-----LELTRLQAQAQGH-----
5920_NDD  1  ----NDYKALKSSIRK-----LQEDDRRLRESGVDEP-----

```

bits

4.0
3.0
2.0
1.0
0.0

1 5 10 15 20 25 30

WebLogo 3.6.0

			Coiled-coil		

5921_NDD	1	----	DSKDRELLARQR	A	RGGQKIQALSRLDG----
5922_NDD	1	----	TMDKLLELLOSQTIREL	TRLQAQEGH----	
5920_NDD	1	----	NDYKALLSSKKIQEOD	RIRESLESVDEP----	
5923_NDD	1	GR--	TVA--HFLDRIEQLS	KPOQALARMRDLQLR---	
5919_NDD	1	----	QDQQLAKSLTIEIKRL	AAANQALEQRRE----	
RAPSB	1	----	RETQTLARKSKENARLR	ANTSLRAMD----	
Aves3	1	----	DTSSKIVVALRASLKANTLR	ARNEQLAAMEASS----	
Spn8	1	----	TVTTIYELVVALRASLKENVLR	GRDRFSAEKD----	
NysC	1	----	STNPKYKVIALRSSLKEIERLR	ONEOLIVAFAV----	
AmphC	1	----	STTKNYVIALRSKKEIERLR	QNEOLIVASAA----	
Ty1GV	1	----	AMAAERITIALRTLSKEAERLR	QNERELRAFDAAR----	
NysJ3	1	----	NAPENPETENNVALRAA	KETDRLRQNRMLVAIAK----	
P1mS3	1	----	TTSETTVMAALRASLKEAENLR	QNRKLVAAT----	
P1kAIIV	1	----	TSNGLVIALRSLEENEELR	ESRRRADRRQ----	
ConB	1	TRTPSSSSPSSSASSQS	PAASPSSSDTTVMALDALKETARLR	ENROLQAVG----	
RifB	1	----	SAPNQVIVALRASLKENVRL	QENSALAAFAA----	
Hvq2	1	----	INQVVALRVSAKENARLR	ENASLOEFAA----	

bits

4.0
3.0
2.0
1.0
0.0

5 10 15 20 25 30

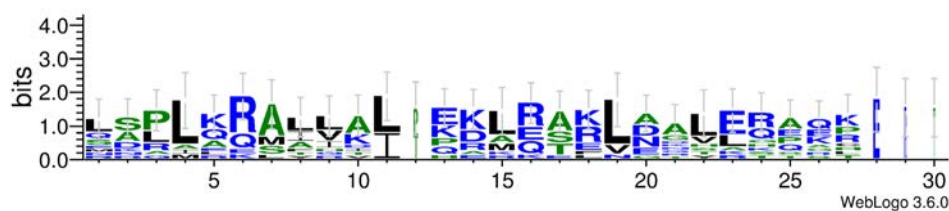
WebLogo 3.6.0

62

C

	Helix α A	Helix α B (coiled-coil)
NidA5	1 -----TRNPKMSPSIDEVLDAITTSKETERRRNQEVAAAR---	
5924_NDD	1 -NK--PTSSDGWKDDYLSRLSRISNQIMALA KLQQQTEQG---	
JamK	1 -----KDSSHNLQELSPDQSFVIEELKSDTLEKSQR---	
AufD	1 -----TASTQDGSQQQRALIEATITIKRAENARSAQK---	
EpoD	1 -----TTRGPTAQNPILQAAIIQILEELAGAQAELERT---	
CyrE	1 -----DEKLRTYERLLQSYHKLEAEAEINRKQTQCE---	
5922_NDD	1 -----TDMKDLLQSIQTIRELKTQLAQAEQGH---	
5921_NDD	1 -----DSKDRELLARQDAIRQGIQKQALSA RLQ---	
StiB	1 -----SDMMKQLARMIRDLPDRFAFIAD-----LRPEP---	
DxH	1 -----STQGSDTGYSDLKKHALKKQDAQS LDAHREERH---	
MxC	1 -----ANRNDESNELESPALMLVAEKIQT LNAEGAAK---	
AjuC	1 -----SNSTETPDQPARTLEAILAH LRT LDAQERQKT---	
MeIE	1 -----STEQNEHSARIALAVALAEKQALAEASREKR---	
CurH	1 LNKFTKKEQILSEKQIKQSPHQAAAKKLETLNNLTH---	
CurL	1 -----NLKQEKEQESLSALQALIAKDARSLEKYETQSK---	

Sequence logo for the helix α A and α B regions of type 2^NDDs:



D

	Helix α A	Helix α B
MInE	1 -----NEKTMQIEQITKKIKESLTPEEGLEILSLGK-----	
RizE	1 -----SGKRTQILQAQSCAMTPPEEGMROLALAV-----	
AtcF	1 -----NSAVRLELETRQISPEERRRRLTALGT-----	
DifG	1 -----KEFLRFSDIKKKMTKQALGILDYDM-----	
5922_NDD	1 -----TDMKDLLQSIQTIRELTRLAQAEQGH---	
5921_NDD	1 -----DSKDRELLARQDAIRQGIQKQALSA RLQ---	
PtZD	1 TQRTIKAGADAQKRVALADEMLSANALIREKQE-----	
TstGH	1 -----ANYHLDITFGQERNGLPRDAADLLRHYYR-----	
RhiD	1 -----SSQEFQSTTAGQNRQITVAEAKQRLQRLKA---	
OzmJ	1 -----STAELDILDRYARCA GAQEQARDLLAARRP-----	
VirFG	1 -----DKELITRFDGDRAAQALAGRTPE-----	
SorC	1 -----NNLVGYLKEIKDRLSKAVLELAEQGT-----	
Elar	1 -----KELVTYINELKADRTSRSEARLLQAGT-----	
ChiC	1 -----KRDPVITFEQIKKQILSRAEAHRLARSDE-----	

Sequence logo for the helix α A and α B regions of 4 α -helix bundle type^NDDs:

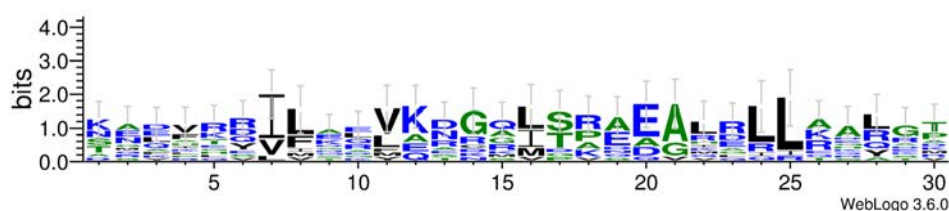


Figure S11, cont. (C) Alignment of the ^NDDs of subunits 5924 (control), 5922 and 5921 with the interaction regions (helices α A and α B) of type 2^NDDs.²⁵ Below is a sequence logo produced from the type 2^CDDs covering the helices α A and α B regions. (D) Alignment of the ^NDD of subunits 5922 and 5921 with the interaction regions (helices α A and α B) of 4 α -helix bundle type^NDDs.^{26,27} Below is a sequence logo produced from the 4 α -helix bundle^NDDs covering the helices α A and α B regions.

Based on the data shown in Figures S11 and S12, we classified the DDs as follows (*italics* = putative assignment):

Type 1-like

5924/5923, 5922/5921, 5021/5920, 5920/5919 (structure solved)

Type 2-like

5925/5924 (structure solved), 5923/5922

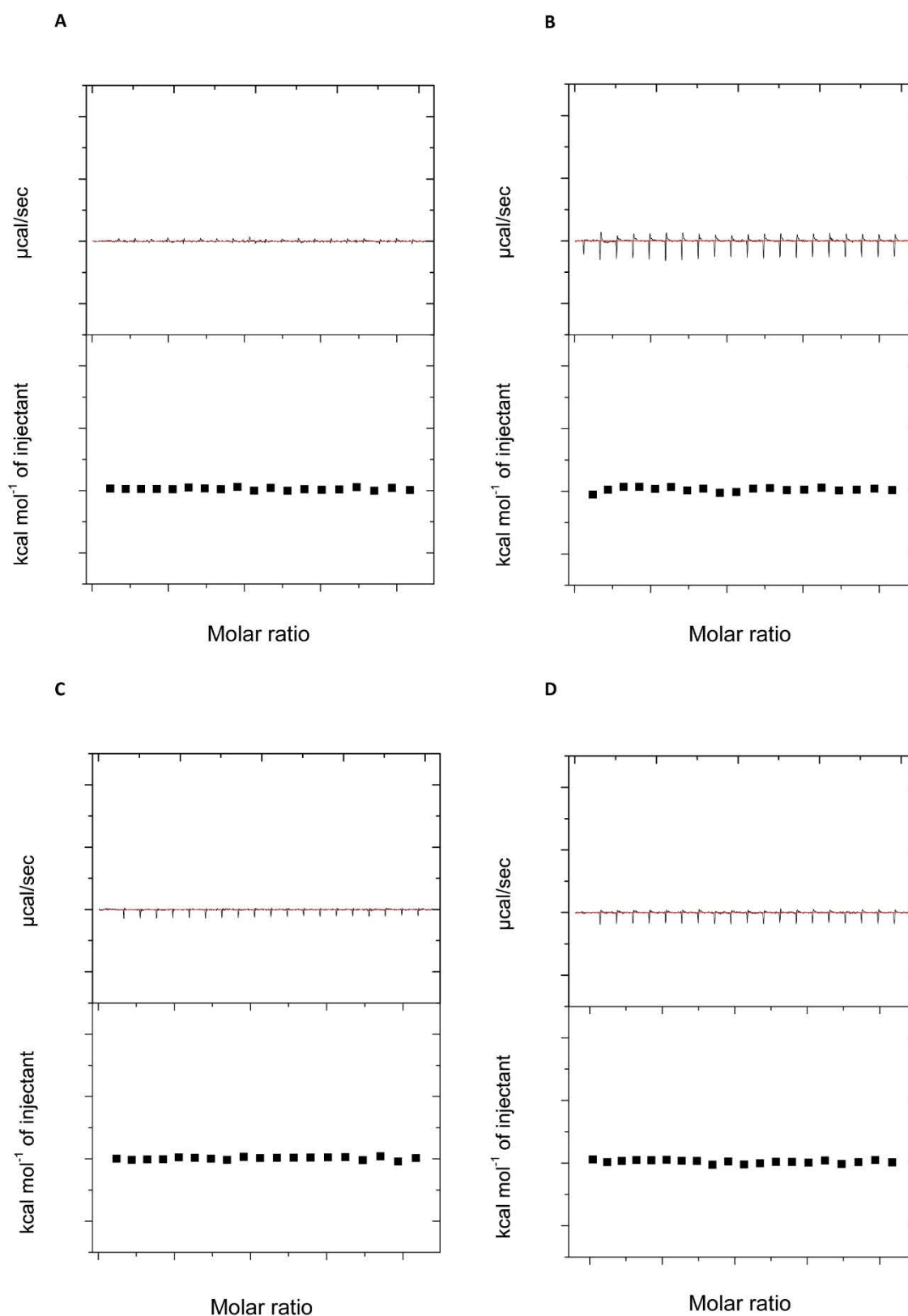


Figure S12. ITC experiments to detect cross-talk between mis-matched docking domain pairs. (A) Analysis by ITC of binding between 5925 $^{\text{C}}$ DD and 5923 $^{\text{N}}$ DD long version which gave no change in the observed heat signal. (B) Analysis by ITC of binding between 5925 $^{\text{C}}$ DD and 5922 $^{\text{N}}$ DD which gave no change in the observed heat signal. (C) Analysis by ITC of binding between 5925 $^{\text{C}}$ DD and 5921 $^{\text{N}}$ DD which gave no change in the observed heat signal. (D) Analysis by ITC of binding between 5925 $^{\text{C}}$ DD and 5920 $^{\text{N}}$ DD short which gave no change in the observed heat signal.

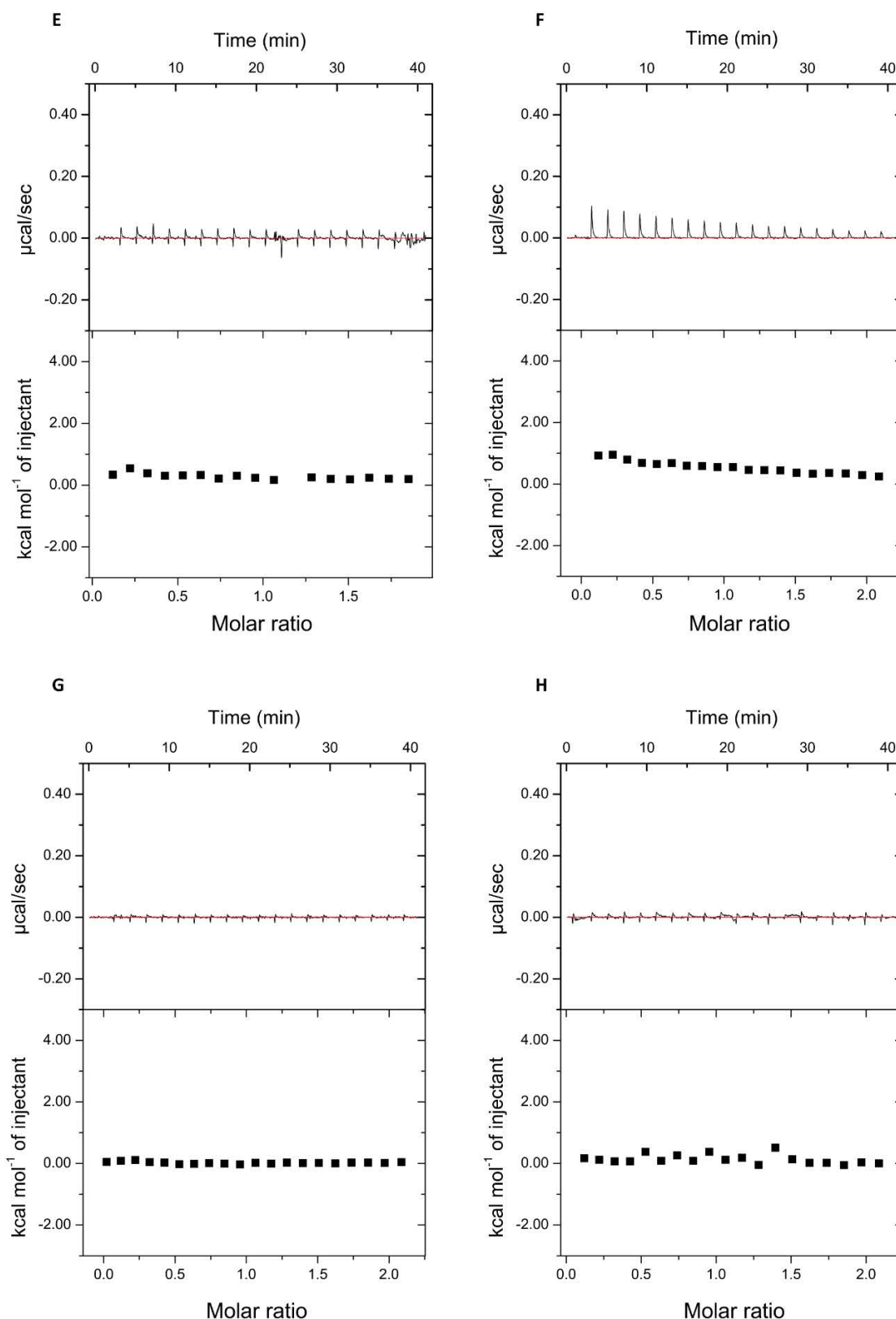


Figure S12, cont. (E) Analysis by ITC of binding between 5925 $^{\text{C}}\text{DD}$ and 5919 $^{\text{N}}\text{DD}$ which gave no change in the observed heat signal. (F) Analysis by ITC of binding between 5921 $^{\text{C}}\text{DD}$ short and 5924 $^{\text{N}}\text{DD}$ which gave a small change in the observed heat signal. (G) Analysis by ITC of binding between 5921 $^{\text{C}}\text{DD}$ short and 5923 $^{\text{N}}\text{DD}$ long which gave no change in the observed heat signal. (H) Analysis by ITC of binding between 5921 $^{\text{C}}\text{DD}$ short and 5922 $^{\text{N}}\text{DD}$ which gave no change in the observed heat signal.

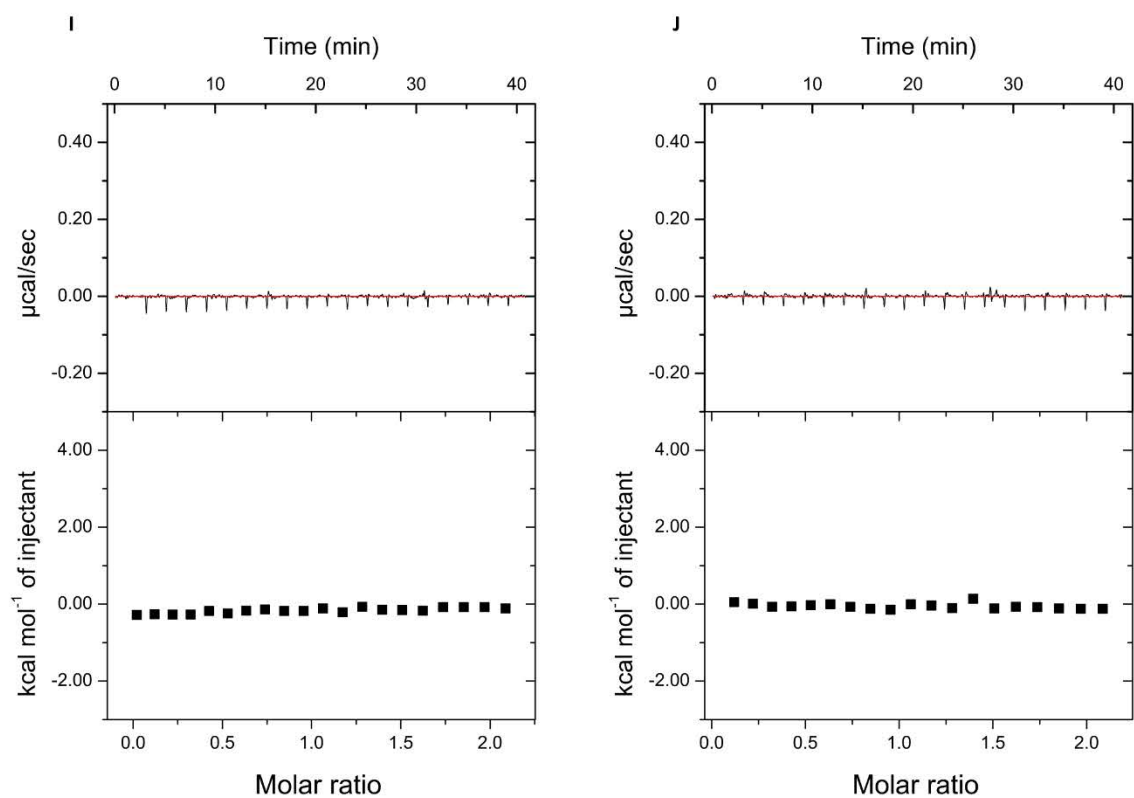


Figure S12, cont. (I) Analysis by ITC of binding between 5921 ^CDD short and 5921 ^NDD which gave no change in the observed heat signal. (J) Analysis by ITC of binding between 5921 ^CDD short and 5919 ^NDD which gave no change in the observed heat signal.

A

```

                                Helix α1                Helix α2
NysA      1  -----AATAPLDPGAYGEELTRFE-ATITNLPQD-GPERRAADRLDAVSALRQNSPA-----EV
KirAV_end 1  -----FDGGTGGADD-----LPAIFADFDRLA-AGLTEAATD-EVFRDRVERVRGLTALS PAT-----P
OleA1     1  -----VGSDDPPLTLRSAIDQLE-TGLALLESDEEARSEITKRLNILLPRFGSGGSSRGR-----EAGQD
Hyg2      1  -----GPLGDGGS-----SPEVVRQLRFE-RQLFDTRPD-DEVGTAIVRLRSILALDAPATAPAA-----TAS
DEBS2     1  -----FAASPAVDIGDR-----LDELE-KALEALSAE--DGHDDVGRLESILR-RWNSRRADAPS-----
MonA2     1  VPEGAGAMGGVSGAEGVRDAYGAGGPGGDM-----TAQVLLERARIE-HTLSAAVPH-GLDRAAARLEALIACTATTAAATGAAGAAVEGDDSDG
RapsB     1  -----NPRVQSTTLLEIDRIE-KMFTSTTFD-PROASA-KDRLSSVLRNKKQRIS----SPEEV-----
PimS1     1  -----APEPTVGP EALBELRIE-KSFGGLDLT-EMHEQVAGRLLEVLRKKDALRDTAAAAGHD-----GSP
AmphC     1  -----VTDDPSGAGSLALGLE-KATIAALSDDAEHRRAAGREVLRTKWAALGRDTADEDA-----PAP
TylGI     1  -----AAELAGPRDGGDT-----AAAAFEGEALAAAAALAEED-LRRDVRRLRLTEAAALTPQGRNPSPAP--A-----AES
AVES1     1  -----QPQPDNAVAPVLAELDLE-SALSALDKT-DSASERTLRKSLTLRW-----NAPQHP--T-----AES
PikAIII   1  -----ATAAGGSWAEGTSGGDTASATDRQTAAALDLRLE-GVLASLAPA-AGGRPELAARLRALAAAL----GD-----D
SpnC      1  -----FPTETTVDSALAEIDRIE-QQLSMLTGE-ARARDRLATRLRALHEKW-----NSAAEV--P-----TG
consensus 1  .....

```

```

                                Helix α3
NysA      55 PSSDIDIVSVDRLDIIDEFEFETI-----
KirAV_end 55 EAPPDLFAAATDDEIFALIDAEIGSPVRGQEDR
OleA1     60 AGEHODVDAIDIEFEVDNENLGS-----
Hyg2      62 AQVLTLLDASQSEILSFIDKEFGGRASN-----
DEBS2     52 --TSAISDASIDELISMIDRFGGEDL-----
MonA2     92 DGAVQLLEAIAQLDFIDNELGV-----
RapsB     54 --STTALSASASEILDFIDREFDPTA-----
PimS1     63 SDEDFDFESASIDFIDLDELGLS-----
AmphC     61 --SSDLDGASDEDNFALDDELGLS-----
TylGI     68 SDDLRLDANDDLFAFIEQL-----
AVES1     57 ADDDEKFTATEAEIKFIDNGLS-----
PikAIII   67 GDDATDLLEASDDLFSFIDKELGDSDF-----
SpnC      55 ADVLSTLDSATHDEIFIDNELDLS-----
consensus 101 .....

```

B

```

KirAV_end 1 FD-----GGTGGADLPALFADFDRAAAGLTEAATDEVFRDRVVERVRGLTALS-----PAT--PEAP-----PD
ConA      1 -----GADSLSA-----VLRNRLRATGLRLS-----ATVVFDPHT-----PTALAGRLRAELFPDRADDTVPDGPDH
NysB      1 FGTAPED-----ATGRPDPP-----ARTRELAIVPIGRLR-----KAGLDMLVLLADGDATDAPAP-----
AmphB     1 -----GRADPP-----ARTRELAIVPIGRLR-----KAGLDMLVLLAESDDTETSAT-----
RifA      1 -----LREDDGLGRE-----DILRRVLAAPVFA-RFR-----EAGVLDITLLGLADTGTEPGTDAETTE-----
Hyg1      1 -----SPAPAAQPDPPDE-----RILRRVLERVPIA-RFR-----EAGVLDITLLGLA-PDPART-----P-
AVES2     1 ---LPDQEATGEAQDQLSGGSE-----EDVRSLLTSPDGRLR-----KAGLDGLLLTADTGRGASGAAAGPEDAPPSGQ-----
RapsA     1 -----APEPEPRLHE-----QLRRRLAGSLSD-RFR-----EAGVLDITLLGLAMEGLAVP-KPDS-----
TylGI     1 -----T-PPGPAAGPRE-----BARALLARVPVLE-RLR-----EAGLDALLRLADESGATTPTRTSAASGAPRGREEP--DGRGEPDGGSH
PikAIII   1 -----L-APPEAPTWE-----GRVRRLAELPIDRLR-----EAGVLDITVLRITIEPEPGSGGSDGG-----
SpnA      1 -----LPE-SAGAGIPDEDE-----ARTRALTSPIFP-ALR-----EAGVSPLLALAHVPVDSG-----
NysI      1 ---FPDE-----PAGSDDE-----TIRALLASVPID-QLR-----ETGVLP LLQLARGGRAA-----
PimS2     1 ---VPDEDPG-EGTADDGPDSS-----SATRALLASVIA-QLR-----ETGVLP LLQLAEQTAASG-AEAA-----
consensus 1 .....

```

```

                                Ha3' Helix α3"
KirAV_end 60 --LFAAATDDE-----VFAIDAEIGSPVRGQEDR-----
ConA      60 AADAEELEAEELIDAMDIAELIRAREGIES-----
NysB      56 --EAD--APSESIDMDAELRLAENSAN-----
AmphB     46 --ATDDADALDDDAELRLAENSAN-----
RifA      55 --AAPADAEIDADISGLQRAIGQTS-----
Hyg1      52 --DAAEPDGLIDAMDTEALRHVLDGTGA-----
AVES2     70 -----DTPAPVSDENDDDLDLAHGHTAPAREPADAEDSSSSNRNTHHTEGETA
RapsA     52 -----ESDFAFIDNDADALRHVLEER-----
TylGI     76 RESPDAAGSALDDLDDGDLRLALGEPGE-----
PikAIII   55 --AADPGLEPASTDDID-AELIRVALGPRNT-----
SpnA      50 ---ISSDDAAATSIDAMDIAELAEALGERES-----
NysI      47 -----DIDGESIDSNITADLRALNGQSDL-----
PimS2     59 --AATAADAYESIDAMDIAELQAAPDGNPQER-----
consensus 101 .....

```

Figure S13. Comparative sequence analysis of the putative kirromycin DDs operating at the KirAV/KirAVI junction. (A) Alignment of the putative ^CDD of KirAV with full-length type 1a ^CDDs.²³ The first two α -helices form a four-helix bundle dimerization motif. (B) Alignment of the putative ^CDD of KirAV with full-length type 1b ^CDDs.²⁴

C

		Coiled-coil
SpnD	1	-----TVTTSYEEIVVIALRSLKENERLRRGRDRFSAEKD----
AmphD	1	-----DNECKLRVYLKLTAD-----LRARRRVGEESASQ----
OleA2	1	-----TNDEKLVVYLLKRAIVD-----LRKARRRMELED-----
NysB	1	QEPQQGQPDQQEKLVVYLLKRVTS-----LRARRRVGEESND-----
RAPSC	1	-----PEQCKLVVYLLWATA-----LITTRAKLEAATAANT-----
TyIGII	1	-----T-----LVATRRLGAEAR-----
Hyg3	1	-----ADSKLTIDYLLKMTAD-----LYRTRERLAEESASA-----
DEBS3	1	---SGDNGMTIEKLRVYLKRVIT-----LDSVTARLRREHAGE-----
PimS2	1	-----SNEKLRVYLKRAIAD-----LETROQLDETEAQR-----
AVES2	1	-----QLANEAKLLVYLKRVITAD-----LDRTRRLRYEVEREQ-----
MonA3	1	-----VSEKLVVYLLKRVITAD-----LATRQRLREAEERQ-----
KirAVI	1	-----SSEKLRVYLKRVITAD-----LQVTRRLREYESSG-----
PikAIII	1	-----ANNEKLRVYLKRVITAD-----LQNTRERLREIETHT-----
consensus	1	*.....

D

		Coiled-coil
KirAVI	1	-----SSEKLRVYLKRVITADLQVTRRLREYESSG-----
RAPSB	1	-----RECKLVVYLLKRAIVD-----LRARRRVGEESASQ-----
Aves3	1	-----DTSSEKLVVYLLKRAIVD-----LRARRRVGEESASQ-----
SpnB	1	-----TVTTSYEEIVVIALRSLKENERLRRGRDRFSAEKD-----
NysC	1	-----STNPKYVIALRSLKENERLRRGRDRFSAEKD-----
AmphC	1	-----STNPKYVIALRSLKENERLRRGRDRFSAEKD-----
TyIGV	1	-----AMSAERITVIALRSLKENERLRRGRDRFSAEKD-----
NysJ	1	-----NAPENPETPENNIALRSLKENERLRRGRDRFSAEKD-----
PimS3	1	-----TTSSETVIALRSLKENERLRRGRDRFSAEKD-----
PikAIV	1	-----TSSNEQVIALRSLKENERLRRGRDRFSAEKD-----
ConB	1	TRTPSSSSPSSASSQSPAASPSSTTTVIALRSLKENERLRRGRDRFSAEKD-----
RifB	1	-----PAPNEQVIALRSLKENERLRRGRDRFSAEKD-----
Hyg2	1	-----TNEQVIALRSLKENERLRRGRDRFSAEKD-----
consensus	1	**.....

Figure S13, cont. (C) Alignment of the putative ^NDD of KirAIV with full-length type 1a ^NDDs.²³ (D) Alignment of the putative ^NDD of KirAVI with full-length type 1b ^NDDs.²⁴

Based on this analysis, the KirAV/KirAIV pair are putatively type 1a domains.

SUPPLEMENTARY REFERENCES

- (1) 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.* **2014**, *5* (8), 3081–3095. <https://doi.org/10.1039/C3SC53511H>.
- (2) 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.* **2006**, *103* (30), 11124–11129. <https://doi.org/10.1073/pnas.0601924103>.
- (3) 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.* **2011**, *18* (5), 665–677. <https://doi.org/10.1016/j.chembiol.2011.01.020>.
- (4) 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.* **2007**, *14* (8), 931–943. <https://doi.org/10.1016/j.chembiol.2007.07.012>.
- (5) Buchan, D. W. A.; Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz297>.
- (6) Piasecki, S. K.; Zheng, J.; Axelrod, A. J.; Detelich, M. E.; Keatinge-Clay, A. T. Structural and Functional Studies of a Trans-Acyltransferase Polyketide Assembly Line Enzyme That Catalyzes Stereoselective α - and β -Ketoreduction. *Proteins* **2014**, *82* (9), 2067–2077. <https://doi.org/10.1002/prot.24561>.
- (7) Keatinge-Clay, A. T. A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol.* **2007**, *14* (8), 898–908. <https://doi.org/10.1016/j.chembiol.2007.07.009>.
- (8) 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. *Struct. Lond. Engl. 1993* **2010**, *18* (8), 913–922. <https://doi.org/10.1016/j.str.2010.04.015>.
- (9) 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. *Struct. Lond. Engl. 1993* **2006**, *14* (4), 737–748. <https://doi.org/10.1016/j.str.2006.01.009>.
- (10) Musiol, E. M.; Härtner, T.; Kulik, A.; Moldenhauer, J.; Piel, J.; Wohlleben, W.; Weber, T. Supramolecular Templating in Kirromycin Biosynthesis: The Acyltransferase KirCII Loads Ethylmalonyl-CoA Extender onto a Specific ACP of the Trans-AT PKS. *Chem. Biol.* **2011**, *18* (4), 438–444. <https://doi.org/10.1016/j.chembiol.2011.02.007>.
- (11) Mészáros, B.; Erdos, G.; Dosztányi, Z. IUPred2A: Context-Dependent Prediction of Protein Disorder as a Function of Redox State and Protein Binding. *Nucleic Acids Res.* **2018**, *46* (W1), W329–W337. <https://doi.org/10.1093/nar/gky384>.
- (12) Svergun, D. I. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J. Appl. Crystallogr.* **1992**, *25* (4), 495–503. <https://doi.org/10.1107/S0021889892001663>.

- (13) Svergun, D. I. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophys. J.* **1999**, *76* (6), 2879–2886. [https://doi.org/10.1016/S0006-3495\(99\)77443-6](https://doi.org/10.1016/S0006-3495(99)77443-6).
- (14) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. PRIMUS: A Windows PC-Based System for Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2003**, *36* (5), 1277–1282. <https://doi.org/10.1107/S0021889803012779>.
- (15) Volkov, V. V.; Svergun, D. I.; IUCr. Uniqueness of ab initio shape determination in small-angle scattering <https://scripts.iucr.org/cgi-bin/paper?ks7125> (accessed May 9, 2019). <https://doi.org/10.1107/S0021889803000268>.
- (16) Kozin, M. B.; Svergun, D. I. Automated Matching of High- and Low-Resolution Structural Models. *J. Appl. Crystallogr.* **2001**, *34* (1), 33–41. <https://doi.org/10.1107/S0021889800014126>.
- (17) Petoukhov, M. V.; Svergun, D. I. Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data. *Biophys. J.* **2005**, *89* (2), 1237–1250. <https://doi.org/10.1529/biophysj.105.064154>.
- (18) Svergun, D.; Barberato, C.; Koch, M. H. J. CRY SOL – a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28* (6), 768–773. <https://doi.org/10.1107/S0021889895007047>.
- (19) Fischer, H.; de Oliveira Neto, M.; Napolitano, H. B.; Polikarpov, I.; Craievich, A. F. Determination of the Molecular Weight of Proteins in Solution from a Single Small-Angle X-Ray Scattering Measurement on a Relative Scale. *J Appl Cryst* **2010**, *43*, 101–109.
- (20) Rambo, R. P.; Tainer, J. A. Accurate Assessment of Mass, Models and Resolution by Small-Angle Scattering. *Nature* **2013**, *496* (7446), 477–481. <https://doi.org/10.1038/nature12070>.
- (21) Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR* **1996**, *8* (4), 477–486.
- (22) 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. D Biol. Crystallogr.* **2010**, *66* (Pt 1), 12–21. <https://doi.org/10.1107/S0907444909042073>.
- (23) 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.* **2003**, *10* (8), 723–731.
- (24) Buchholz, T. J.; Geders, T. W.; Bartley, 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.* **2009**, *4* (1), 41–52. <https://doi.org/10.1021/cb8002607>.
- (25) 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.* **2013**, *20* (11), 1340–1351. <https://doi.org/10.1016/j.chembiol.2013.09.015>.
- (26) Dorival, J.; Annaval, T.; Risser, F.; Collin, S.; Roblin, P.; Jacob, C.; Gruez, A.; Chagot, B.; Weissman, K. J. Characterization of Intersubunit Communication in the Virginiamycin Trans-Acyl Transferase Polyketide Synthase. *J. Am. Chem. Soc.* **2016**. <https://doi.org/10.1021/jacs.5b13372>.

- (27) 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.* **2016**, *11* (9), 2466–2474. <https://doi.org/10.1021/acscchembio.6b00345>.

Caractérisation d'une enzyme post-PKS, LkcE

Dorival, J., **Risser, F.**, Jacob, C., Collin, S., Dräger, G., Paris, C., Chagot, B., Kirschning, A., Gruez, A. & Weissman, K. J. Insights into a dual function amide oxidase/macrocyclase from lankacidin biosynthesis. *Nat. Commun.* **9** : 3998 (2018).

Bien que peu exploitées à ce jour, les enzymes post-PKS présentent un intérêt certain pour les expériences d'ingénierie. Notre équipe s'est intéressée à une de ces enzymes, LkcE, qui est capable de catalyser une réaction de macrocyclisation d'un intermédiaire linéaire de la lankacidine, LC-KA05. Cette réaction fait appel à un mécanisme enzymatique rare, puisqu'elle implique une oxydation du groupement amide du substrat. LkcE serait la deuxième enzyme caractérisée capable de catalyser ce type de réaction (Ames et al., 2011). Cette oxydation est suivie d'une réaction de Mannich intramoléculaire et aucune autre enzyme n'utiliserait ce type de chimie à l'exception de cas particuliers de promiscuité d'activité (Wu et al., 2016).

Un premier travail bio-informatique met en évidence que LkcE possède 63% d'identité de séquence avec une amine oxydase caractérisée chez la bactérie *Hahella chejuensis*. Les plus proches homologues identifiés dans la PDB sont les monoamines oxydases humaines (hMAO) A (PDB : 2Z5X) et B (PDB : 1GOS), avec 26% et 27% d'identité de séquence respectivement.

La protéine a été clonée et purifiée à homogénéité et son cofacteur redox, le FAD, a été identifié par spectrométrie de masse. Les structures de la protéine sous forme *holo* ainsi qu'en complexe avec deux analogues de substrat : l'éthyl 2-méthylacétoacétate (EMAA) et le *N,N'*-diallyl-L-tartardiamide (DATD) (chacun mimant une partie du substrat naturel) (Figure 51 page 96), ont pu être obtenues par cristallisation et diffraction des rayons X.

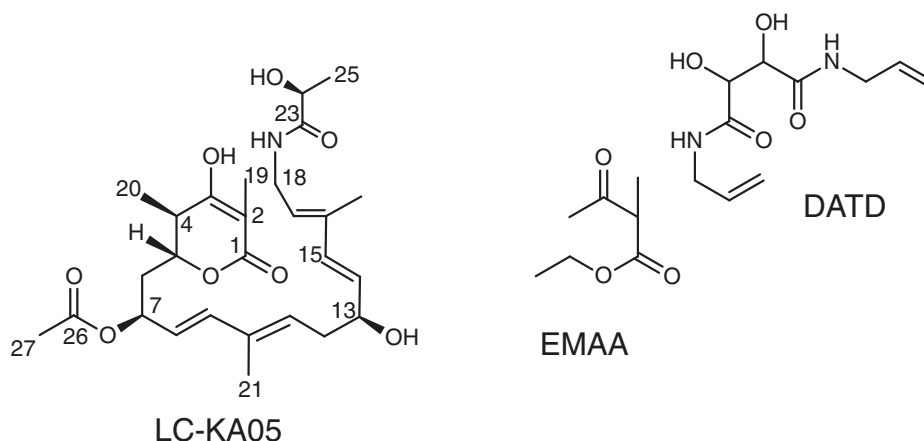


Figure 51 : Structures du substrat naturel de LkcE, LC-KA05, et des deux analogues : l'éthyl 2-méthylacétoacétate (EMAA) et le N,N'-diallyl-L-tartardiamide (DATD), utilisés pour mimer une partie de ce substrat naturel.

La première structure sous forme *holo* de la protéine sélénée a été résolue par diffusion anormale à une seule longueur d'onde (SAD) à 3,15 Å de résolution. La seconde, en présence des analogues de substrat, a été résolue par remplacement moléculaire à 2,8 Å de résolution. La comparaison des structures des cinq homologues structuraux les plus proches permettent de confirmer que LkcE appartient à la famille des MAO, le *rmsd* est de 4,72 Å avec la hMAO A (254 Cα) et de 5,00 Å avec la hMAO B (267 Cα). La protéine se présente sous la forme d'un homodimère, comme la plupart des enzymes de cette famille. Chaque monomère comporte deux domaines, le premier fixe le FAD, la séquence motif caractéristique xhxhGxGxxGxxxhxxh(x)₈hxhD(E) y est présente (x : acide aminé ; h : acide aminé hydrophobe) (Sillitoe et al., 2015). Ce domaine diffère d'un Rossmann *fold* classique de par la topologie adoptée par les éléments de structure secondaire. Le second domaine catalytique assure la fixation des analogues de substrat. Les superpositions des structures de LkcE avec celles des homologues les plus proches indiquent que bien que le domaine de fixation du FAD soit bien conservé, le domaine de fixation du substrat est significativement différent en termes de composition en hélices α et d'orientation de ces dernières. Ceci peut être expliqué par le fait que LkcE accommode au sein de son site actif un substrat bien plus gros comparé aux petites molécules typiquement prises en charge par les MAO.

Un autre aspect intéressant de la structuration adoptée par LkcE se situe dans son interface d'homodimérisation différente des MAO les plus proches, lui donnant une structure quaternaire différente. La hMAO B dimérise à la fois *via* son site de fixation du FAD et son site catalytique alors que LkcE dimérise uniquement *via* son site catalytique. Ce mode de

dimérisation alternatif ainsi que la forte présence de boucles à l'interface favoriseraient des mouvements de chaque monomère l'un par rapport à l'autre. La structure en présence des analogues de substrat révèle que ces derniers adoptent une position équivalente au sein de chaque site actif : le DATD se fixe au monomère A et l'EMAA au monomère B. Les deux molécules sont principalement stabilisées par des interactions hydrophobes.

L'analyse de ces structures, couplée aux alignements de séquences et à des considérations mécanistiques (oxydation de la fonction amide en ion iminium, suivie de l'attaque de l'énol porté par le C2) permettent de proposer des résidus potentiellement impliqués dans la catalyse : Glu64, Tyr182 et Arg326. Les mutants Glu64Ala, Glu64Gln, Tyr182Phe, Arg326Leu et Arg326Glu obtenus par mutagenèse dirigée ont donc été produits et purifiés à homogénéité. Une souche de *Streptomyces rochei*, dans laquelle le gène *lkce* a été supprimé, a permis l'accumulation, l'extraction et la purification de l'intermédiaire linéaire de la lankacidine. Ce substrat étant peu stable, deux produits de dégradation sont aussi obtenus. Le premier, correspondant à une élimination du groupement acétate en C7, est co-élué avec LC-KA05 dans un rapport 1:2 (ce composé est appelé « produit d'élimination »). Le deuxième produit de dégradation, correspondant à une hydrolyse de ce même groupement, a pu être séparé par HPLC des deux premières populations (Figure 52 page 97).

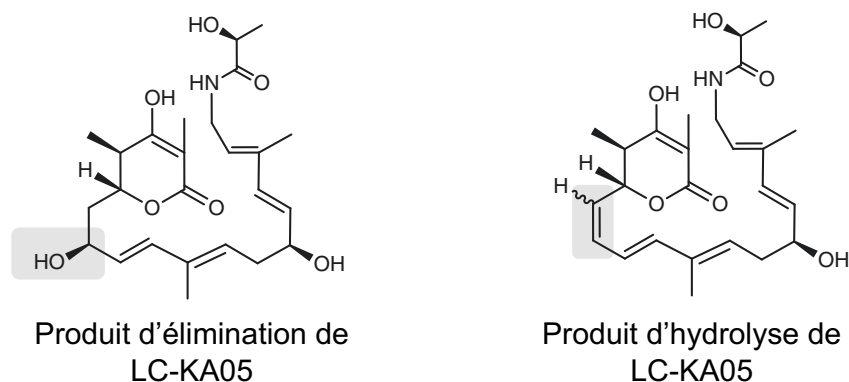


Figure 52 : Structures du produit d'élimination et d'hydrolyse du substrat naturel de LkcE.

Une analyse HPLC-MS atteste que LkcE est capable de cycliser les trois substrats : LC-KA05, le produit d'élimination et celui d'hydrolyse. Ce résultat nous démontre que LkcE est bien active et qu'elle a la capacité d'accommoder des substrats présentant des variations aux positions C6 et C7. De plus, cette enzyme est un excellent exemple de gain de fonction au cours de l'évolution : elle appartient à la famille des MAO mais est capable de réaliser une

réaction de cyclisation. Cette nouvelle fonction pourrait avoir émergé grâce au mode de dimérisation alternatif par rapport aux autres membres de la famille.

Un système enzymatique couplé a été mis au point afin de suivre de manière plus sensible la catalyse et de caractériser les mutants. Le système repose sur la production de H_2O_2 lors de la régénération du cofacteur FAD. Le peroxyde d'hydrogène est pris en charge par la NADH peroxydase de *Streptococcus faecalis* qui catalyse la réduction du H_2O_2 en parallèle à l'oxydation du NADH. De cette manière la réaction catalysée par LkcE est suivie par la perte d'absorbance à 340 nm, correspondant à l'oxydation du NADH. Une constante catalytique (k_{cat}) de $3,4 \pm 0,2 \text{ min}^{-1}$ et une constante de Michaelis (K_M) de $5 \pm 2 \mu\text{M}$ ont été calculées pour l'enzyme sauvage en présence de la mixture LC-KA05/produit d'élimination. Ce *turn-over* est comparable à celui rapporté par un système PKS *in vitro* et qui est de $1,1 \text{ min}^{-1}$ (Schlippe et al., 2005). L'analyse cinétique de la protéine sauvage en présence du produit d'hydrolyse du substrat permet l'obtention de paramètres cinétiques similaires : $k_{\text{cat}} = 2,4 \pm 0,1 \text{ min}^{-1}$ et $K_M = 5 \pm 1 \mu\text{M}$. Les mutants Glu64Ala, Glu64Gln et Arg326Leu sont inactifs alors que les mutants Tyr182Phe et Arg326Gln présentent une activité comparable à l'enzyme sauvage en présence du substrat déacétylé.

Des structures des mutants Glu64Gln et Arg326Gln, à 3,03 Å et 2,50 Å de résolution respectivement, ont été obtenues par remplacement moléculaire à partir de cristaux trempés dans la mixture LC-KA05/produit d'élimination. LC-KA05 est positionné à l'interface entre le domaine de fixation du substrat et le domaine de fixation du FAD et adopte une conformation étendue. Des résidus hydrophobes et des liaisons hydrogènes permettent l'accommodation de cette molécule au sein d'une cavité profonde de 20 Å de long sur 6 à 8 Å de large. Les structures adoptées par les deux mutants ne sont pas favorables à la cyclisation du substrat : le polycétide est présent dans cette cavité sous une forme étendue, 13,2 Å séparent les C2 et C18 impliqués dans la réaction de cyclisation (Figure 52 page 97). La conformation adoptée par le substrat serait une conformation de pré-cyclisation. Afin de créer l'espace nécessaire à une telle réaction l'enzyme doit subir un changement conformationnel important. En nous basant sur l'analyse structurale, nous avons émis l'hypothèse que ces changements reposent sur un mouvement charnière entre les domaines de fixation du cofacteur et de fixation du substrat qui serait rendu possible par le mode alternatif d'homodimérisation de LkcE.

L'ensemble de ces résultats permet de proposer le mécanisme catalytique suivant : LC-KA05 se fixe au site actif *via* formation d'une liaison hydrogène entre son groupement énol et le groupement amide de l'Arg326. Le transfert d'hydride a ensuite lieu avec le FAD

engendrant la formation de l'iminium. La Glu64 joue ensuite le rôle de catalyseur base général et active le groupement émol permettant la cyclisation de la molécule par la réaction de Mannich intramoléculaire (Figure 53 page 99).

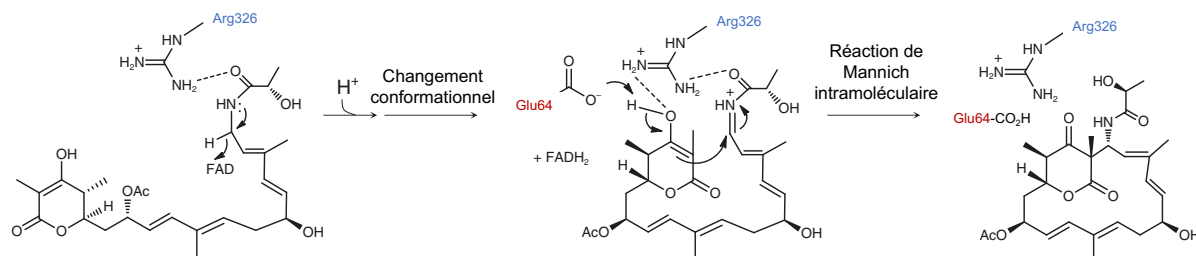


Figure 53 : Mécanisme réactionnel proposé de la catalyse réalisée par LkcE.

Dans ce travail collaboratif au sein de l'équipe, le docteur Jonathan DORIVAL a cloné, produit et purifié l'enzyme sauvage et certains mutants. Il a réalisé avec l'aide du docteur Arnaud GRUEZ et de Sabrina COLLIN l'ensemble des expériences de cristallographie et résolu les structures avec la participation du docteur Arnaud GRUEZ. J'ai participé au clonage, à la production et à la purification des mutants, et ai mis au point le système couplé afin de suivre l'activité de l'enzyme sauvage et de ses mutants. La souche mutante de *Streptomyces rochei* a été obtenue par le docteur Christophe JACOB, la production et l'extraction des substrats linéaires ont été réalisées par Jonathan DORIVAL, Christophe JACOB, Benjamin CHAGOT, Arnaud GRUEZ et Sabrina COLLIN. La purification et la caractérisation de ces substrats ont été réalisées par le docteur Gerald DRÄGER appartenant à l'équipe collaboratrice du professeur Andreas KIRSCHNING. Le docteur Cédric PARIS a réalisé les expériences de HPLC-MS.

ARTICLE

DOI: 10.1038/s41467-018-06323-w

OPEN

Insights into a dual function amide oxidase/macrocyclase from lankacidin biosynthesis

Jonathan Dorival^{1,4}, Fanny Risser¹, Christophe Jacob¹, Sabrina Collin¹, Gerald Dräger², Cédric Paris³, Benjamin Chagot¹, Andreas Kirschning², Arnaud Gruez¹ & Kira J. Weissman¹

Acquisition of new catalytic activity is a relatively rare evolutionary event. A striking example appears in the pathway to the antibiotic lankacidin, as a monoamine oxidase (MAO) family member, LkcE, catalyzes both an unusual amide oxidation, and a subsequent intramolecular Mannich reaction to form the polyketide macrocycle. We report evidence here for the molecular basis for this dual activity. The reaction sequence involves several essential active site residues and a conformational change likely comprising an interdomain hinge movement. These features, which have not previously been described in the MAO family, both depend on a unique dimerization mode relative to all structurally characterized members. Taken together, these data add weight to the idea that designing new multifunctional enzymes may require changes in both architecture and catalytic machinery. Encouragingly, however, our data also show LkcE to bind alternative substrates, supporting its potential utility as a general cyclization catalyst in synthetic biology.

¹UMR 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, BP 20199, 54505 Vandœuvre-lès-Nancy Cedex, France. ²Institut für Organische Chemie, Leibniz Universität Hannover, Schneiderberg 1B, Hannover 30167, Germany. ³Laboratoire d'Ingénierie des Biomolécules, Ecole Nationale Supérieure d'Agronomie et des Industries Alimentaires (ENSAIA), Université de Lorraine, 2 Avenue de la Forêt de Haye, BP 172, 54518 Vandœuvre-lès-Nancy Cedex, France. ⁴Present address: Sorbonne Universités, UPMC Univ. Paris 06, CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074 Roscoff, Bretagne, France. These authors contributed equally: Fanny Risser, Christophe Jacob. Correspondence and requests for materials should be addressed to A.G. (email: arnaud.gruez@univ-lorraine.fr) or to K.J.W. (email: kira.weissman@univ-lorraine.fr)

Enzyme activity within superfamilies frequently evolves via changes in substrate specificity, whereas reaction chemistry is usually retained^{1,2}. There are also well-documented examples of a conserved protein fold housing diverse catalytic activities: the $\alpha\beta$ hydrophobic barrel fold accommodates at least six different types of acid–base chemistry³, the fumarylacetoacetate hydrolase superfamily includes decarboxylases, isomerases, hydratases, and hydrolases⁴, whereas the enolase superfamily has > 30,000 members, which share an interdomain active site architecture and the mechanistic strategy of proton abstraction α to a carboxylate and stabilization of the enolate by a magnesium ion⁵. This evolution of novel activity can occur via several different mechanisms, including the insertion or deletion of residues, changes in oligomerization state, and the recruitment of metal ions and cofactors¹. The enzyme LkcE from the lankacidin polyketide biosynthetic pathway of the bacterium *Streptomyces rochei* represents an interesting test case for understanding such gain-of-function, as it exhibits a monoamine oxidase (MAO) protein fold yet also carries out a macrocyclization reaction.

The core of the lankacidin molecule is constructed from simple acyl-CoA and amino-acid-building blocks by a series of gigantic multimodular polypeptides, which operate in assembly-line fashion (polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), respectively)⁶. The full-length chain is released from the last multienzyme LkcG by a dedicated thioesterase domain⁷. Unusually for this type of pathway, the product is liberated from the multienzyme not as a macrolide, but as a β -oxo- δ -lactone (Fig. 1). LkcE then catalyzes post-assembly macrocyclization via formation of a C-2–C-18 carbon-carbon bond⁸. This reaction is proposed to occur on a C-7 acetylated metabolite LC-KA05 (1), although the timing of the acetylation reaction, and the enzyme responsible, are unknown. A strain of *S. rochei* in which LkcE is inactivated accumulates LC-KA05 (Fig. 1), which can be converted into lankacidin C (3) by a strain blocked early in lankacidin biosynthesis, demonstrating the chemical competence of LC-KA05 as an intermediate⁸. In the final stages of the biosynthesis, the C-24 hydroxyl group is oxidized to a ketone and the C-7 acetyl group is cleaved, yielding lankacidin C, with the

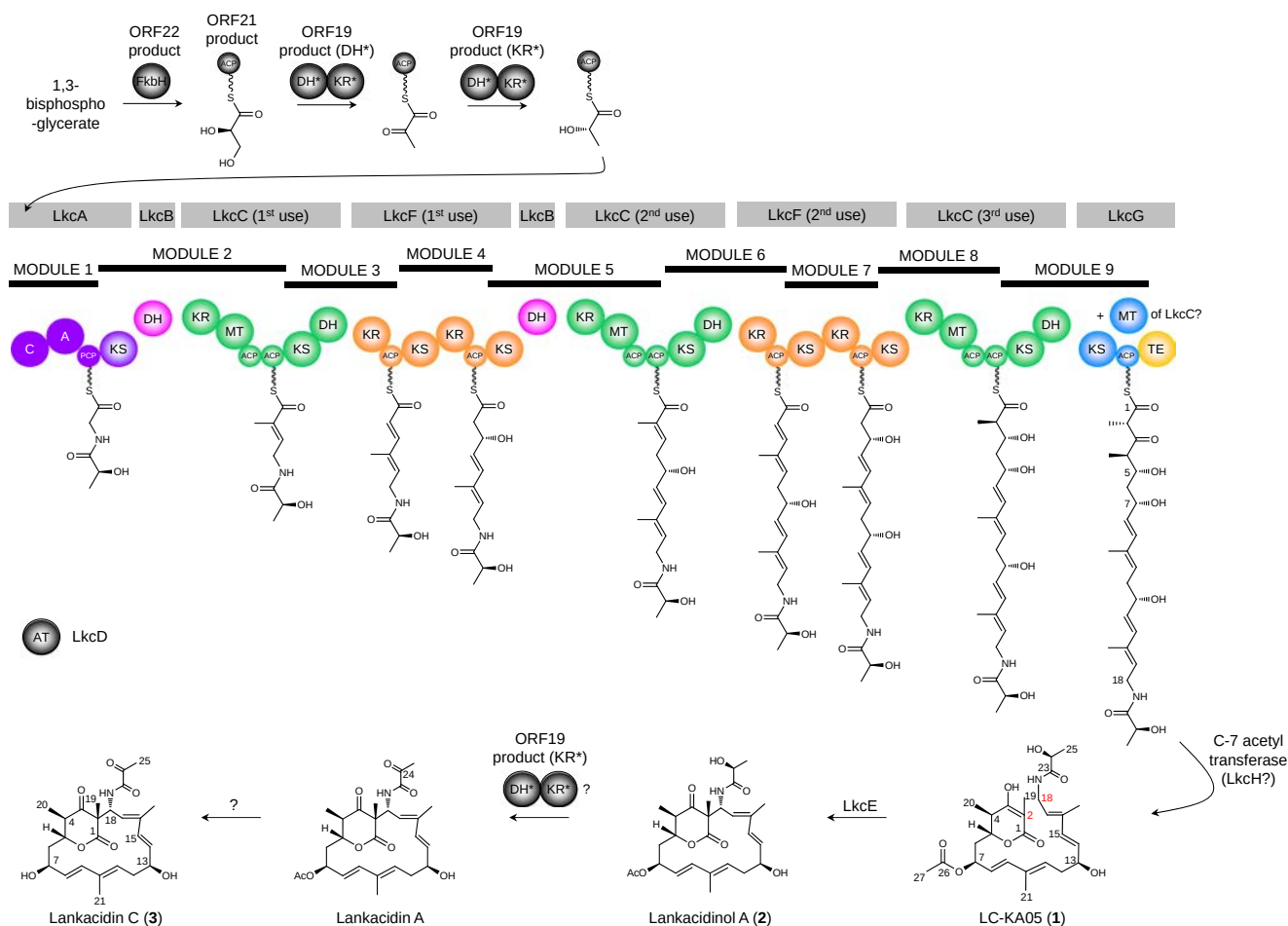


Fig. 1 Proposed biosynthetic pathway to lankacidin C. The gene cluster, which has been shown to be sufficient for lankacidin biosynthesis⁴⁴, encodes five assembly-line proteins, LkcA (a hybrid NRPS/PKS subunit), LkcB (a discrete DH), LkcC, LkcF, and LkcG, together containing a total of only four KS domains, although eight KS-mediated extension cycles are required. One possibility that agrees with phylogenetic analysis of KS substrate specificity²³ is shown here, in which the assembly-line incorporates multiple copies of the proteins LkcB, LkcC and LkcF (see ref. 44 for an alternate view). The starter unit may be either pyruvoyl-ACP or lactoyl-ACP, both derived from 1,3-bisphosphoglycerate⁴⁵. However, we and others⁸ have found that in an LkcE-deleted mutant, only the lactoyl form of the first free intermediate LC-KA05 accumulates (Supplementary Fig. 8), so we propose that lactoyl-ACP serves as the starter unit. The enzyme responsible for acetyl transfer at C-7 likewise remains to be determined, although a candidate is LkcH, which shows homology to isochorismatases⁸. The object of this study, LkcE, catalyzes the critical macrocyclization reaction of LC-KA05 (1), resulting in lankacidinol A (2). The carbons implicated in the ring closure are indicated in red. Domain abbreviations: C, condensation; A, adenylation; PCP, peptidyl carrier protein; KS, ketosynthase; DH, dehydratase; KR, ketoreductase; MT, C-methyltransferase; ACP, acyl carrier protein; TE, thioesterase

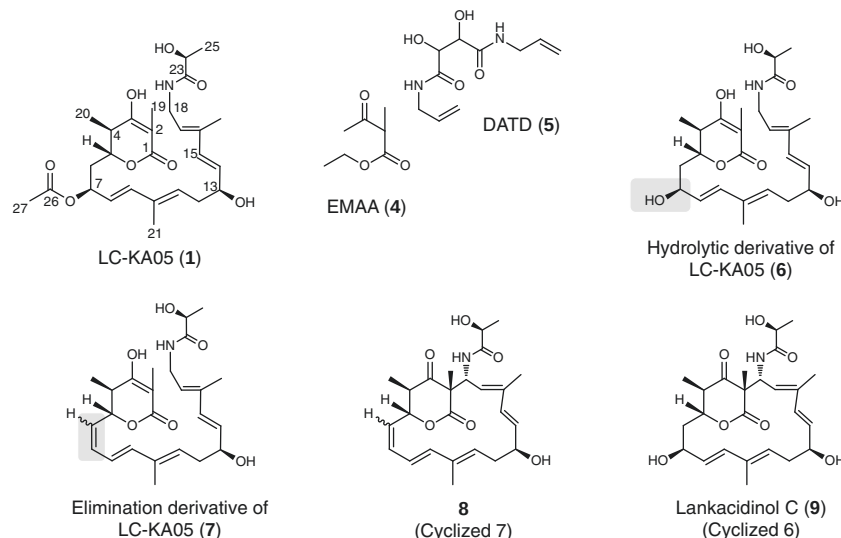


Fig. 2 Structures of selected compounds investigated in this study. Compound LC-KA05 (**1**) was shown by NMR to exist almost exclusively in the enol form, and thus its derivatives **6** and **7** are also represented as enols. The gray boxes indicate where **6** and **7** differ from **1**. The stereochemistry of the C-6–C-7 double bond in **7** and cyclized **7** (**8**) is unknown, as indicated

relative timing of these two transformations also remaining to be established (Fig. 1).

The catalytic mechanism proposed for LkcE involves initial oxidation of the amide function of LC-KA05 to an iminium ion⁸. This is followed by attack of a C-2 enolate anion on the iminium to give the 17-membered lankacidinol A (**2**)⁸—an enzymatic intramolecular Mannich reaction. Although numerous flavin adenine dinucleotide (FAD)-dependent amine oxidases have been characterized to date⁹, to our knowledge LkcE is only the second described amide oxidase¹⁰. This observation, coupled with the fact that no other known enzyme carries out Mannich chemistry except as a promiscuous activity¹¹, motivated our interest in establishing a detailed structure/function relationship for LkcE.

In this study, we have used a combination of X-ray crystallography, structure-guided mutagenesis, and kinetic analysis, to reveal key architectural differences between LkcE and other members of the MAO family, which underpin its second, cyclization activity. We also demonstrate the tolerance of LkcE to certain structural modifications of the substrate. As 1,3-diketones and amide functional groups are common in PK/NRP hybrid metabolites, LkcE represents a potentially valuable addition to a synthetic biology toolbox as a means to access novel macrocyclic structures of various sizes and functionality.

Results

Structural characterization of holo and ligand-bound LkcE. *S. rochei* LkcE was purified to homogeneity as a recombinant protein (Supplementary Tables 1 and 2) from *Escherichia coli* and its redox cofactor identified as FAD by mass spectrometry following chromatographic separation under denaturing conditions (Supplementary Fig. 1); thus in common with certain MAO family members¹² but distinct from the only other reported amide oxidase Af12070 (ref. ¹⁰), the FAD cofactor is non-covalently bound. By UV-Vis, 45% of the protein was estimated to contain FAD (on a par with Af12070 (ref. ¹⁰)) (Methods). We crystallized the enzyme in its holo form, but additionally in the presence of substrate analogs: ethyl 2-methylacetoacetate (EMAA) (**4**) that mimics the δ -lactone of LC-KA05, and *N,N'*-diallyl-L-tartardiamide (DATD) (**5**), which resembles the amide region (Fig. 2). The X-ray crystal structure of the holo protein was solved at 3.15 Å resolution by single-wavelength anomalous dispersion (SAD) using data collected on seleniated protein (Fig. 3a). The structure

of LkcE (Fig. 3b, c) co-crystallized in the presence of the analogs was then solved at 2.80 Å by molecular replacement (the statistics for data collection, refinement, and validation of both structures are presented in Supplementary Table 3). For all four structures, >99.7% of the residues were in allowed regions of the Ramachandran plot. In all but the SeLkcE structure, one Gly (299 or 300) was an outlier, whereas Glu313 was an outlier in both forms of the holo enzyme structure. However, for Glu313, clear electron density is present corresponding to the residue.

Sequence alignment (Supplementary Fig. 2) and comparison with the structures of the five closest structural homologs to LkcE identified by the DALI¹³ server confirmed that LkcE belongs to the MAO family. The respective root mean square deviation (rmsd) of Ca positions was 4.232 Å (212 Ca) for 6-hydroxy-L-nicotine oxidase (6HDNO) from *Paenarthrobacter nicotinovorans* (PDB 3NG7)¹⁴; 4.720 Å (254 Ca) for human monoamine oxidase (hMAO) A (2Z5X)¹⁵; 5 Å (267 Ca) for hMAO B (1GOS)¹⁶; 10.239 Å (271 Ca) for protoporphyrinogen oxidase (PPOX) of *Bacillus subtilis* (3I6D)¹⁷; and 14.660 Å (162 Ca) for a polyunsaturated fatty acid isomerase of *Propionibacterium acnes* (2BAB)¹⁸.

The LkcE monomer contains two domains, a FAD-binding domain and a substrate-binding domain (Fig. 3a), which are joined by a substantial number of loop regions. The cofactor-binding domain (which includes residues 1–45 incorporating a characteristic FAD-binding motif (xhxhGxGxxGxxxhxxh (x)₈hxh(D)¹⁹), 70–83, 206–279, and 368–438), is topologically similar to other proteins with the ‘3-layer (BBA) sandwich’ fold in the CATH database²⁰, and comprises a central four-stranded antiparallel β -sheet flanked on one side by five α -helices, and on the other side by a second three stranded, antiparallel β -sheet and three α -helices. The FAD cofactor is present at 100% occupancy (Supplementary Fig. 3), and as in other members of the MAO family, it adopts an elongated conformation. Although the specific FAD-binding residues differ among LkcE and its closest structural homologs (Supplementary Fig. 2), the types of interactions are similar, with the exception of those to the isoalloxazine ring. In all of the structures except LkcE and PPOX, the isoalloxazine is flanked by two bulky aromatic residues (Tyr, and in 3NG7 only, a Tyr and a Trp). In LkcE, the equivalent residue positions are Gly364 and Leu398, respectively, whereas in PPOX, the analogous amino acids are M413 and G449. Thus, it

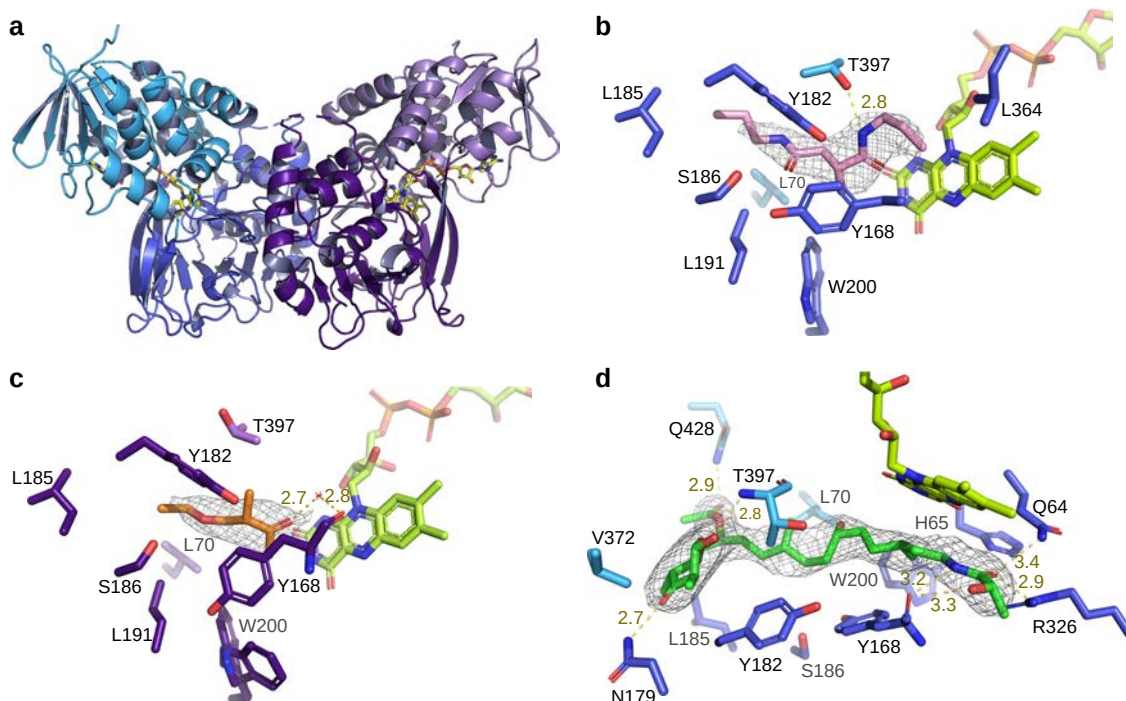


Fig. 3 Crystal structures of LkcE and its mutants. **a** Crystal structure of homodimeric, wild-type LkcE. The FAD-binding domain is shown in light blue and light purple for the monomers A and B, respectively, whereas the substrate-binding domain is shown in dark blue and purple. The FAD is colored in yellow. **b** View of the active site of wild-type LkcE in the presence of bound DATD (pink) with its $2F_o - F_c$ map contoured at 1σ . Binding occurs mainly via hydrophobic interactions and a single hydrogen bond with T397. **c** View of the active site of wild-type LkcE in the presence of bound EMAA (orange) with the $2F_o - F_c$ map contoured at 0.6σ . Comparison with **b** shows the analog to be binding in the same region of the active site. **d** View of the active site of the LkcE E64Q mutant in the presence of bound LC-KA05 (green), which adopts a linear conformation. The $2F_o - F_c$ map surrounding the substrate is contoured at 1σ

appears that the FAD-binding site has been modified in these two enzymes in order to accommodate the large macrocycles of the substrates/products.

The substrate-binding domain contains an orthogonal α -helix bundle that packs against a β -sheet. Visual inspection of the overlaid structures shows that although the cofactor-binding domain is well-conserved, the substrate-binding domain is significantly divergent in terms of both α -helical content and orientation (Supplementary Fig. 4). This is unsurprising given the pronounced structural differences between LC-KA05 and the typical small-molecule amine substrates of the MAO family²¹.

Although the majority of characterized MAO family members are homodimeric, there is some disagreement between crystallographic and solution data, and at least one homolog is monomeric⁹. LkcE is clearly a homodimeric protein in the crystal, with an extensive interface ($4705 \text{ \AA}^2$, representing 15% of the total surface area). It is also homodimeric in solution, as shown by small-angle X-ray scattering (Fig. 4). However, it dimerizes differently to the human enzymes hMAO A¹⁵ and B¹⁶, as well as to its closest structural homolog 6HDNO¹⁴ (Fig. 5), giving it a distinctive quaternary organization. In the hMAO B crystal structure, the dimerization interface is formed by both the cofactor- and substrate-binding domains, whereas in the case of 6HDNO, stabilization of the homodimer is additionally provided by two molecules of diacylglycerophospholipid. In contrast, the dimer interface of LkcE is formed exclusively by the substrate-binding domain of each monomer. The interface residues include E68, K69, V87, L106, F108, E114, D121, Q125, N128, Q129, S188, Y199, R201, H332, and R335. Most of these are well-conserved between LkcE and its nearest homologs, but not all (Supplementary Fig. 2). This distinct mode of dimerization, coupled with the

high content of loops at the interdomain interface within each monomer (Fig. 3a), would favor movement of the domains relative to each other, with potentially important implications for the catalytic mechanism, as discussed below.

Structure-guided site-directed mutagenesis of LkcE. Analysis of the crystal structures with substrate analogs bound revealed clear electron density attributable to EMAA and DATD, but only one of these compounds was present in any given active site, as their positions overlapped (Fig. 3b, c). Inspection of the two binding sites coupled with mechanistic considerations identified E64, H65, Y168, Y182, R326, and T397 as candidate catalytic residues (Supplementary Fig. 2). The presence of amino acids in the active site capable of acting as general acids/bases contrasts with other flavin-dependent amine oxidase homologs which lack such residues^{14,21}.

In particular, E64, Y182, and R326 are strictly conserved in all genes putatively encoding LkcE-type cyclases (Supplementary Fig. 2). Site-specific mutagenesis was used to produce five mutants: E64A, E64Q, Y182F, R326L, and R326Q, which were purified to homogeneity (FAD content, respectively: E64A, 58%; E64Q, 53%; Y182F, 42%; R326L, 20%; R326Q, 60%). Analysis by circular dichroism (CD) confirmed that the mutations had not dramatically altered the structures (Supplementary Fig. 5).

Isolation of the native substrate of LkcE. To obtain the native LkcE substrate LC-KA05, we disrupted gene *lkce* in *S. rochei* by PCR targeting, using a strategy similar to that described previously^{8,22,23} (Supplementary Figs. 6 and 7). Following analysis of the *lkce* mutant by high-performance liquid chromatography-mass spectrometry (HPLC-MS) to verify the presence of LC-KA05

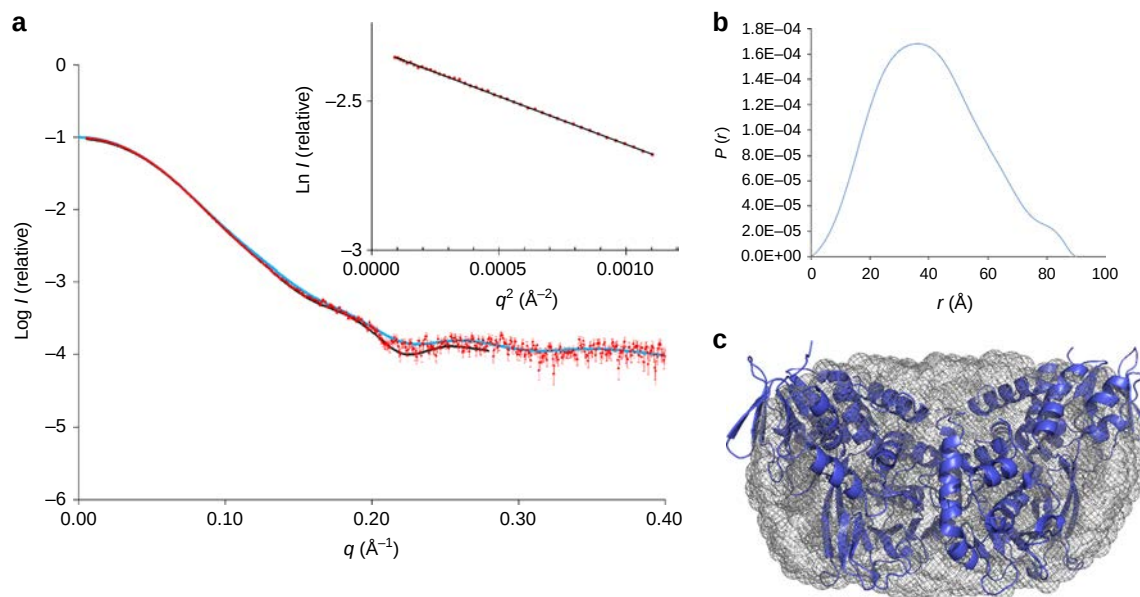


Fig. 4 Characterization of recombinant LkcE by SAXS. **a** Fit between the *ab initio* model computed with GASBOR⁴⁶ (black line) ($\chi^2 = 1.44$), the theoretical scattering curve calculated on the basis of the structure of LkcE_{WT} with CRY SOL⁴⁷ (blue line) ($\chi^2 = 3.8$) and the experimental SAXS data (red dots). Inset is the Guinier plot, which yields an R_g of 31.0 Å. A molecular weight (MW) of 99.3 kDa was calculated using the SAXS MoW⁴⁸ program (homodimer calculated MW = 98.6 kDa). **b** The distance distribution function derived for LkcE calculated with GNOM⁴⁹, yielding a $D_{\max}$ of 91 Å. **c** Averaged *ab initio* envelope of LkcE calculated using GASBOR⁴⁶ (gray mesh) with superimposition of the LkcE_{WT} crystal structure carried out using SUPCOMB⁴⁹ (blue)

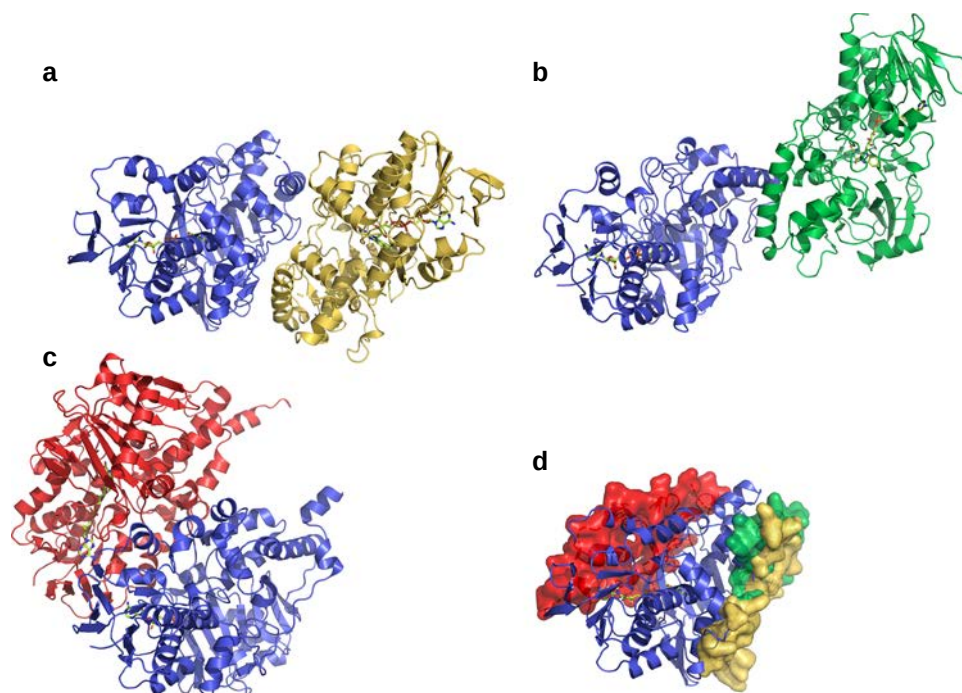


Fig. 5 Alternate homodimerization mode of LkcE relative to hMAO B and 6HDNO. **a** Homodimeric structure of LkcE. **b** Homodimerization mode of 6HDNO. The dimer is superimposed on one monomer of LkcE (in blue in **a**), in the same orientation. **c** Homodimerization mode of hMAO B. Again, the dimer has been superimposed on a monomer of LkcE. **d** Interaction surface between the two monomers in LkcE (yellow), hMAO B (red), and 6HDNO (green), relative to a monomer of LkcE. This analysis clearly shows that the mode of homodimerization of LkcE, as well as its overall quaternary organization, differ from these two homologs

and the absence of all downstream metabolites (Supplementary Fig. 8 and Supplementary Table 4), **1** was obtained in pure form by preparative HPLC (Supplementary Fig. 9). Unfortunately, LC-KA05 was unstable, undergoing degradation via loss of the C-7 acetate, both by hydrolysis (to give 7-OH **6**, Fig. 2) and

elimination (to give **7** (Fig. 2)), which co-eluted with LC-KA05. However, 16 mg of mixed metabolite could be obtained from the extracts of 20–30 L of culture, with **1** and **7** present in an ~ 2:1 ratio (chemical data on all three compounds are provided in the Methods). As HPLC-MS analysis of extracts of both the wild-type strain

and the *lkcE* mutant did not reveal any evidence for **7**, nor, in the case of the wild type, for its cyclized equivalent **8** (Fig. 2; Supplementary Fig. 10 and Supplementary Table 4), the cellular environment must protect against this type of non-productive degradation. The 7-OH compound **6** was observed in extracts of the *lkcE* mutant although not in the wild type (Supplementary Fig. 10 and Supplementary Table 4), presumably because in the presence of active LkcE it can be productively cyclized to give lankacidinol C (**9**) (Fig. 2) (see below), and from there transformed to lankacidin C. Interestingly, NMR analysis of purified LC-KA05 revealed that the δ -lactone was almost exclusively in the enol form (Supplementary Fig. 9) as recently described for a new lankacidin-derived metabolite²⁴, presumably because the six-membered ring forces overlap in the π -system.

Kinetic characterization of LkcE and its mutants. Attempts to determine the kinetics of LkcE-catalyzed conversion of LC-KA05 to lankacidinol A by HPLC-MS were frustrated by the rapid degradation of **1** under the assay conditions (Supplementary Fig. 11). Nonetheless, larger scale incubation of **1** in the presence of LkcE produced authentic lankacidinol A, as judged by HPLC-MS comparison with extracts of wild-type *S. rochei* containing this metabolite, providing conclusive evidence that recombinant LkcE was active (Supplementary Fig. 12 and Supplementary Table 4). The enzyme was also found capable of cyclizing the degradation product **7**, as well as purified, deacetylated substrate **6**, to give lankacidinol C, as established by HPLC-MS (Supplementary Fig. 12 and Supplementary Table 4), evidence for its relaxed specificity toward structural variations at the C-6/C-7 ring positions.

As an alternative and more-sensitive kinetic method, we employed a coupled enzymatic assay to detect the H_2O_2 formed during each catalytic cycle. For this, we utilized NADH peroxidase from *Streptococcus faecalis*, which catalyzes the NADH-dependent reduction of H_2O_2 to water. In this way, LkcE turnover (production of H_2O_2) was detected via the consumption of NADH (loss of absorbance at 340 nm) under conditions where NADH peroxidase was not rate-limiting (see Methods)²⁵. We also confirmed that the presence of dimethyl sulfoxide (used to solubilize the substrates) had no effect on the measured kinetic parameters (Fig. 6 and Table 1). However, as **1** was present at only ca. 66% in the mixture, and degraded spontaneously during the assays, the kinetic parameters must be considered lower estimates. Nonetheless, the possibility to rapidly acquire many data points in the initial stages of the reaction ($A_{340\text{ nm}}$ was measured every 0.4 s) meant that effects of the decomposition of **1** were minimized.

With the ca. 2:1 LC-KA05/eliminated derivative **7** mixture, this system yielded the following steady-state kinetic parameters for the wild type (average of measurements in duplicate, and taking into account the observed proportion of holo LkcE (45%)): $k_{\text{cat}} = 3.4 \pm 0.2 \text{ min}^{-1}$ and $K_{\text{M}} = 5 \pm 2 \mu\text{M}$. This k_{cat} is comparable to both the overall rate of turnover reported for an intact PKS system in vitro (1.1 min^{-1})²⁶, as well as that for the amide oxidase AF12070 (ca. 5 min^{-1} (ref. 10)) (Fig. 6 and Table 1). Analysis of the wild-type enzyme with deacetylated derivative **6** yielded comparable parameters ($k_{\text{cat}} = 2.4 \pm 0.1 \text{ min}^{-1}$, $K_{\text{M}} = 5 \pm 1 \mu\text{M}$) (Fig. 6 and Table 1). Mutants E64A, E64Q, and R326L were completely inactive with the 1/7 mixture (as was boiled protein control), whereas mutants Y182F and R326Q showed good activity towards the deacetylated derivative **6**: R326Q, $k_{\text{cat}} = 0.88 \text{ min}^{-1}$, $K_{\text{M}} = 4 \mu\text{M}$; Y182F, $k_{\text{cat}} = 1.5 \text{ min}^{-1}$, $K_{\text{M}} = 4 \mu\text{M}$ (Fig. 6 and Table 1) (owing to limited quantities, the 1/7 mixture was not tested with these mutants). Taken together, these data are consistent with residues E64 and R326 playing

critical roles in the catalytic mechanism (although R can be substituted by Q without a significant effect on activity), whereas Y182 is not essential. We also confirmed lack of turnover with the substrate analogs EMAA and DATD, both separately and together (using 2 and 10 μM enzyme and up to 400 μM in both substrates), consistent with their overlapping binding modes (Fig. 3b, c).

Crystal structures of LkcE mutants in complex with LC-KA05.

In order to obtain complexes with LC-KA05, mutants E64A, E64Q, Y182F, R326L, and R326Q were soaked with the 1/7 substrate mixture followed by rapid acquisition of the X-ray diffraction data (owing both to the low availability of LC-KA05 and its high instability, it was not possible to carry out co-crystallization experiments). Of the five mutants, only E64Q and R326Q yielded co-complexes with **1**; **7** was not observed in the active sites, as the acetate of **1** was clearly visible. The structure of the LkcE E64Q/LC-KA05 complex was solved at 3.03 Å resolution and that of the LkcE R326Q/LC-KA05 complex at 2.50 Å resolution by molecular replacement, using the structure of the LkcE/EMAA/DATD complex as the search model, and the presence of LC-KA05 confirmed using a weighted $F_o - F_c$ omit map (Supplementary Fig. 3; Supplementary Table 3).

LC-KA05 lies at the interface between the substrate-binding domain and the cofactor-binding domain and adopts an extended conformation. The substrate sits in a deep pocket (~ 20 Å from the protein surface and ~ 6 – 8 Å wide), and is surrounded by hydrophobic residues (L70, Y168, Y182, L185, M189, W200, L324, F345, V396, and G398). In addition to these hydrophobic interactions, binding is likely mediated by seven hydrogen bonds: two to the lactone ring at the entry point of the pocket (by N179 and T397), one to the acetate at C-7 by the side chain of Q428, and four to the lactoyl moiety (two with Y168, and one each with E64 and R326). Notably, in this configuration, the lactoyl portion of LC-KA05 stacks against the FAD isoalloxazine in a position appropriate for hydride transfer (the amide proton is situated at 3.2 Å from the FAD N-5). On the other hand, superposition of the R326Q/LC-KA05 structure and that of the holo wild type shows that if LC-KA05 bound into the wild type in the same orientation as in the LkcE R326Q/LC-KA05 complex, its C-24 hydroxyl would sterically clash with R326 (Supplementary Fig. 13). In the E64Q structure in which R326 is present, the protein is identical, but the substrate adopts an alternative orientation, which allows it to interact with R326 (3.1 Å between the C-23 carbonyl of LC-KA05 and the R326 side chain) (Supplementary Fig. 13).

In both structures obtained in the presence of **1**, the observed linear configuration of the polyketide chain would not permit subsequent cyclization involving the C-2 and C-18 centers (they are separated by 13.2 Å). This observation is consistent with the idea that LC-KA05 bound into the E64Q and R326Q mutants represents an on-pathway, pre-cyclization conformation. Indeed, the fact that the R326Q mutant retains essentially wild-type activity means that we have not simply captured a non-productive complex with the enzyme. To create the space necessary for LC-KA05 to adopt its cyclization-ready state, the enzyme must undergo a conformational change. Based on our structural analysis, we propose that this involves a hinge movement between the cofactor-binding and substrate-binding domains, made possible by the distinctive dimerization mode of LkcE. The new substrate configuration may be stabilized by a second hydrogen bonding interaction with the bifunctional R326. This new configuration would also position the substrate enol in proximity to E64. Such a large-scale conformational change of the enzyme might account for the observation that soaking **1** with the

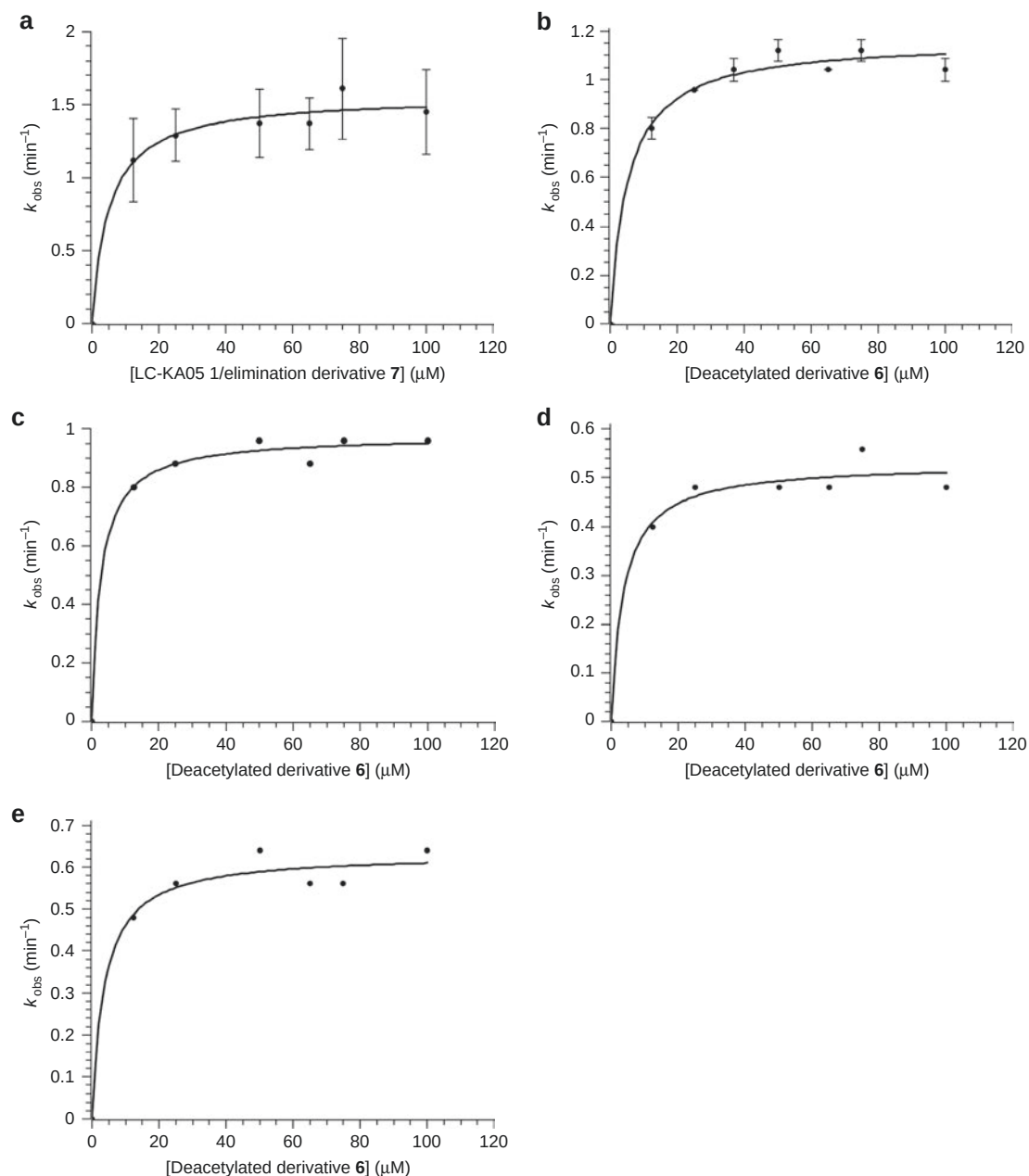


Fig. 6 Kinetic analysis by coupled assay of LkcE and its mutants. Data for which errors are shown were obtained by experiments repeated in duplicate (**a**) (the error bars thus show the two data points) or triplicate (**b**) (the error bars indicate the standard deviation). **a** LkcE acting on its native substrate, LC-KA05 (**1**), as a 2:1 mixture with eliminated product **7**. **b** LkcE acting on deacetylated (7-OH) substrate **6**. **c** LkcE acting on **6** but with each concentration adjusted to contain the maximum amount of DMSO (that present at 100 μM **6** in **b**). As the data in **b** and **c** are essentially identical, the concentration of DMSO in this range had no effect on the rate. **d** LkcE R326Q acting on **6**. **e** LkcE Y182F acting on **6**. It must be noted that as we were limited in these assays for reasons of sensitivity to higher concentrations of substrate, it is possible that we missed an earlier sigmoidal dependence on concentration, indicative of cooperative behavior between the two LkcE monomers

E64Q and R326Q mutants failed to yield a cyclization-compatible conformation.

Based on these data, we propose the following mechanism for LkcE (Fig. 7). LC-KA05 in its enol form binds into the active site via interactions that include H-bonding between the amide portion and R326. The fact that no degradation products of LC-KA05 are observed in the wild type in vivo strongly suggests that binding into LkcE is closely coupled to release of the intermediate from the PKS and its acetylation, perhaps by physical association of LkcE, the last PKS subunit, LkcG, and the as yet unidentified acetyltransferase. Hydride transfer⁹ then occurs to FAD to form

the iminium. Although we initially considered that the conserved R326 in the active site might facilitate substrate oxidation by acting as a catalytic general base to deprotonate the amide²⁷, the near wild-type activity of the R326Q mutant is instead consistent with the residue playing a role in maintaining the cyclization-ready conformation of the substrate established by a subsequent conformational change. The side chain of E64 would then initiate ring closure by acting as a general base catalyst to activate the enol (there is no counterpart of this residue, to our knowledge, in other MAO family members⁹). This key role is supported by the complete loss of activity seen with the E64A and E64Q mutants.

Table 1 Kinetic data obtained for LkcE and its mutants

Enzyme	Substrate(s)	k_{cat} (s^{-1})	K_{M} (μM)
LkcE wt	LC-KA05 (1)/elimination derivative (7) mixture	3.4 ± 0.2	5 ± 2
LkcE wt	Deacetylated derivative (6)	2.4 ± 0.1	5 ± 1
LkcE wt	Deacetylated derivative (6) complemented with DMSO	2.2	3
LkcE wt	EMAA (4)	X	X
LkcE wt	DATD (5)	X	X
LkcE wt	EMAA (4)/DATD (5)	X	X
LkcE E64A	LC-KA05 (1)/elimination derivative (7) mixture	X	X
LkcE E64Q	LC-KA05 (1)/elimination derivative (7) mixture	X	X
LkcE R326L	LC-KA05 (1)/elimination derivative (7) mixture	X	X
LkcE R326Q	Deacetylated derivative (6)	0.88	4
LkcE Y182F	Deacetylated derivative (6)	1.5	4

X = no detected activity

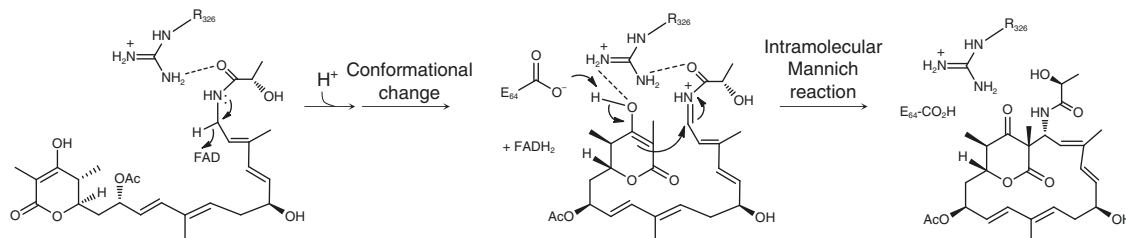


Fig. 7 Proposed mechanism of the LkcE-catalyzed reaction. LC-KA05 binds into the active site, and undergoes FAD-catalyzed oxidation to yield the iminium ion. A conformational change then occurs to bring the δ -lactone in proximity to the iminium, with the new LC-KA05 conformation potentially stabilized by interaction with R326. The final step is an intramolecular Mannich reaction, involving general base catalysis by E64 to generate the nucleophilic C-2 enolate

This mechanism could also explain the advantage of modifying the intermediate with an acetyl protecting group²⁸ as this assures that attack on the iminium occurs only from C-2 and not from the C-7 hydroxyl.

Discussion

Studying how enzymes naturally gain new functions is a valuable source of information for attempts to replicate this process in the laboratory using rational design or combined rational/directed evolution approaches²⁹. Here, we have investigated the bifunctional amide oxidase/macrocyclase LkcE to understand how a subsequent cyclization function might have been acquired by an MAO scaffold.

Existing structure/function data, although insufficient to reliably trace the evolution of the MAO family, clearly demonstrate that the shared di-domain structure of the enzymes can support a common oxidation mechanism of diverse amine substrates, including primary and secondary amines, polyamines, amino acids, and methylated lysine side chains in proteins⁹. Our results strongly suggest that the new catalytic activity of LkcE depends, among other factors, on a change in the mode of dimerization of the enzyme relative to classical members of the MAO family, such as hMAOs A and B. With dimerization limited to the substrate-binding domain in LkcE, the cofactor-binding domain is substantially less constrained. This physical decoupling would appear to permit the two domains to undergo a hinge motion, which we propose underlies a necessary conformational change of the initially-bound substrate between an essentially linear and a cyclization-ready conformation. In addition, the residue at position 326 (LkcE numbering), which in the hMAOs is positioned at the dimer interface, is liberated in LkcE, allowing it to serve a

catalytic role in the active site. E64, a second residue that has been implicated in the reaction mechanism, is located in a position that favors non-deleterious mutation, on a surface loop whose sequence is highly divergent between LkcE and hMAOs A and B (Supplementary Fig. 2).

Evolving specificity for LC-KA05 relative to classical MAO substrates also evidently required a substantial shift in amino-acid sequence within the substrate-binding domain (Supplementary Fig. 2)¹, as well as at least two mutations increasing the space available adjacent to the FAD cofactor. In this context, the inherent promiscuity of the MAOs likely provided a favorable evolutionary starting point for the acquisition of new substrate specificity¹. We have shown here that LkcE is capable of binding two substantially smaller substrate analogs, and shows tolerance toward structural variation at C-6/C-7 in the native substrate. Although further work will be required to define its substrate profile in detail, these data encourage the idea that LkcE may find wider use as a ligation/macrocyclization catalyst in both synthetic biology and organic synthesis applications. In view of the largely hydrophobic nature of the binding pocket, it will be particularly interesting to explore whether minimally functionalized acyl chains are also substrates.

In conclusion, on the basis of our analysis we propose that the bifunctional amide oxidase/macrocyclase LkcE acting on a highly functionalized polyketide chain, arose from an ancestral amine oxidase not only through substantial modification of the substrate-binding region and recruitment of two catalytic residues into the active site, but also by adoption of a different quaternary organization from modern MAOs. This adds to the emerging awareness of the wider structural alterations that may be necessary to introduce new reaction chemistry into existing enzyme active sites³⁰.

Methods

Analysis in silico of LkCE. As a starting point for characterizing LkCE, we identified its closest sequence homologs in the NCBI database using BlastP³¹. For the top 11 hits, we also determined the genomic context of the genes (Supplementary Fig. 2). This analysis revealed nine LkCE homologs located within complete or partial lankacidin (or the closely related chejuenolide) gene clusters (as determined with reference to that described in *S. rochei*) (Supplementary Fig. 2), with sequence identity to the *S. rochei* *lkcE* gene in the range of 54–100%. To aid in defining features that distinguish LkCE from classical monoamine oxidases, we also included in our alignment the two nearest homologs which are not present in lankacidin clusters. We also used two of the closest structural homologs to LkCE from the PDB, which show amine oxidase activity (as identified using the DALI¹³ server). Multiple sequence alignment was carried out using ClustalW (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html)³² and the figures created using ESPript³³.

Materials and DNA manipulation. Biochemicals and media were purchased from Thermo Fisher Scientific (EDTA), VWR (glycerol, NaPi, NaCl), BD (peptone, yeast extract), Euromedex (isopropyl β-D-1-thiogalactopyranoside; IPTG), and Sigma-Aldrich (imidazole). The enzymes for genetic manipulation were purchased from Thermo Fisher Scientific and NADH peroxidase from NZYTech. DNA isolation and manipulation were performed using standard methods^{34,35}. Isolation of DNA fragments from agarose gel and purification of PCR products were carried out using the NucleoSpin Extract II kit (Macherey Nagel, Hoerd, France). Standard PCR reactions were performed with Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific); reactions were carried out on a Mastercycler Pro (Eppendorf). Synthetic oligonucleotides were purchased from Sigma-Aldrich, and DNA sequencing was carried out by GATC Biotech (Mulhouse, France). All organic solvents used were HPLC grade and purchased from Sigma-Aldrich.

Bacterial strains and culture conditions. *E. coli* strain DH5α was used for cloning (the primer sequences are provided in Supplementary Table 1), *E. coli* Rosetta 2 and *E. coli* B834 pRARE2 (DE3) were used respectively for producing unlabeled and seleniated protein, *E. coli* BW25113 for PCR targeting^{22,36}, and *E. coli* ET12567 [pUZ8002] for transforming *S. rochei* var. *volubilis* ATCC 21250. This strain of *S. rochei* was a kind gift of Professor P.F. Leadlay (University of Cambridge). *E. coli* strains were grown on 2TY (16 g L⁻¹ tryptone, 5 g L⁻¹ yeast extract, 5 g L⁻¹ NaCl, adjusted to pH 7.6 with NaOH) for cloning purposes and LB (10 g L⁻¹ tryptone, 10 g L⁻¹ yeast extract, 5 g L⁻¹ NaCl, adjusted to pH 7 with NaOH) for production of unlabeled protein. Seleniated protein was produced in M9 medium (50 mM Na₂HPO₄, 22 mM KH₂PO₄, 10 mM NaCl, 20 mM NH₄Cl, adjusted to pH 7.2 with NaOH). After autoclaving, sterile-filtered ingredients were added as follows: 50 mg L⁻¹ of thiamine and riboflavin, 4 g L⁻¹ glucose, 100 μM CaCl₂, 2 mM MgSO₄, 40 mg L⁻¹ selenomethionine, and 40 mg L⁻¹ of the remaining 19 amino acids). The *E. coli* cultures also contained the appropriate concentration of antibiotics (50 mg mL⁻¹ kanamycin and 25 mg mL⁻¹ chloramphenicol in pre-cultures; 5 and 2.5 mg mL⁻¹, respectively, in protein production cultures). *S. rochei* was grown on SFM (20 g L⁻¹ mannitol, 20 g L⁻¹ soy flour, 20 g L⁻¹ agar) in order to produce a spore suspension for conjugation with *E. coli*, and yeast extract-malt extract-glucose (YMG) medium (4 g L⁻¹ glucose, 4 g L⁻¹ yeast extract, 10 g L⁻¹ malt extract, adjusted to pH 7.2 with KOH) for metabolite production. For conjugation experiments, a spore suspension was mixed with the *E. coli* 12567[pUZ8002] strain and plated on SFM containing 10 mM MgCl₂. For metabolite production, *S. rochei* was grown at 30 °C with shaking at 180 rpm in a rotary incubator for 3 days.

Sub-cloning and site-directed mutagenesis of *lkcE*. Plasmid pGEX-Lcs [a kind gift of Professor P.F. Leadlay; *lcsf* is equivalent to *lkcE*²³] was digested with *Bam*HI and *Xho*I, and the resulting fragment cloned into the equivalent sites of plasmid pBG-102 (Center for Structural Biology, Vanderbilt University), which codes for a His₆-SUMO tag, to yield pBG-102-LkCE (the full list of plasmids used in this study is provided in Supplementary Table 2). The sequence of *lkcE* was re-verified by sequencing. Site-directed mutations were introduced into *lkcE* by PCR using mutagenic oligonucleotides (Supplementary Table 1) and Phusion high-fidelity polymerase, followed by digestion of the parental DNA by 1 μL of *Dpn*I Fast digest (Thermo Fischer Scientific). The presence of the correct mutations was confirmed by sequencing.

Expression and purification of recombinant LkCE and mutants. After an overnight pre-culture at 37 °C, *E. coli* Rosetta 2 containing pBG-102-LkCE was grown in LB medium supplemented with riboflavin (10–50 mg L⁻¹) in order to support FAD biosynthesis. When the cultures reached an A₆₀₀ of 0.6, the cultures were subjected to a combined chemical and thermal shock (addition of 3% ethanol, followed by 2 h at 4 °C), and then protein expression induced by the addition of IPTG (final concentration of 0.1 mM). Incubation was then continued at 15 °C overnight. The culture was then centrifuged for 30 min at 3500 g, the resulting cell pellet resuspended in 30 mL phosphate buffer (100 mM NaPi, 10% glycerol, 10 mM EDTA, pH 7.4), and the cells re-centrifuged, before storage at –20 °C.

The cells resulting from 1 L of culture were resuspended in lysis buffer (30 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 500 mM NaCl, pH

7.5) (10 mL of buffer were used for each A₆₀₀ unit at the end of culturing). In total, 400 units of benzonase were added to each 100 mL of culture, as well as 6 mM MgSO₄, in order to eliminate nucleic acids. The cells were then lysed using a cell disruptor (Basic Z, Constant Systems Ltd.) at 15 kPsi (1000 bars) at 4 °C, and the cellular debris removed by centrifugation (35,000 g for 40 min). After sterile filtration (0.22 μm filter) and addition of 70 mM imidazole, the supernatant was injected at 3 mL min⁻¹ onto a Ni-Sepharose column (GE Healthcare Life Sciences) pre-equilibrated in lysis buffer containing 70 mM imidazole, using an Äkta Avant system (GE Healthcare Life Sciences). This was followed by a wash step using equilibration buffer until the OD stabilized. The protein was then eluted using the same buffer but containing 250 mM imidazole, at 5 mL min⁻¹. The LkCE-containing fractions were identified by 12.5% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), and pooled. The His₆-SUMO tag was removed by incubation with 150 μg human rhinovirus 3 C protease (HRV3C protease), and dialyzed against lysis buffer overnight to eliminate the imidazole. The sample was then reinjected onto the Ni-Sepharose column (2 mL min⁻¹), which had been pre-equilibrated with lysis buffer containing 20 mM imidazole, and the flow-through containing LkCE collected. After this reverse nickel step, the protein was diluted three times into reduced ionic strength buffer (30 mM HEPES, 1 mM EDTA, pH 7.5), in order to allow for purification by anion exchange.

For this, the sample was injected (5 mL min⁻¹) onto a Q-sepharose column (trimethylammonium on 6% agarose) equilibrated in buffer (30 mM HEPES, 100 mM NaCl, 1 mM EDTA, pH 7.5). LkCE was then eluted using a NaCl gradient (from 100 mM to 1 M) at 5 mL min⁻¹. The LkCE-containing fractions were identified by 12.5% SDS-PAGE, and pooled. In order to eliminate the remaining contaminants and any aggregates, the protein was subjected to a final gel filtration step. For this, it was concentrated using an Amicon Ultracel-30 (Merck Millipore) by centrifugation at 4000 g, to obtain a volume of less than 8 mL. This was then injected via a 10 mL loop onto a Superdex 200 26/60 prep grade column (GE), which had been pre-equilibrated in buffer (30 mM HEPES, 150 mM NaCl, 1 mM EDTA, pH 7.5). The protein was eluted in the same buffer at 2 mL min⁻¹. Fractions containing pure LkCE were identified by 12.5% SDS-PAGE, and pooled. The protein was then concentrated to 20 mg mL⁻¹ by centrifugation at 4000 g using an Amicon Ultracel-30, and 50 μL aliquots frozen in liquid nitrogen prior to storage at –80 °C.

Analysis by CD of LkCE and its mutants. CD was carried out on a Chirascan CD (Applied Photophysics). Data were collected at 0.5 nm intervals in the wavelength range of 200–260 nm at 20 °C, using a temperature-controlled chamber. A 0.01 cm cuvette containing 30 μL of protein sample at 50 μM was used for all the measurements. Each spectrum represents the average of three scans, and sample spectra were corrected for buffer background by subtracting the average spectrum of buffer alone.

Identification of LkCE FAD cofactor. The LkCE cofactor was identified as FAD by passing the enzyme over a reverse phase C8 column (Grace) using an HPLC (Äkta Explorer, GE Healthcare) in a gradient of 0–80% acetonitrile containing 0.1% trifluoroacetic acid. The peaks corresponding to the protein and to the released cofactor were then analyzed by mass spectrometry (F. Dupire, Mass Spectrometry Service of the Faculty of Sciences and Technologies, Université de Lorraine). The FAD content of LkCE was then estimated by UV-Vis by measuring its absorbance (450 nm, ε = 11500) using a SAFAS spectrometer (UVMc2), as well as that of the protein (280 nm, ε = 56840). For this, LkCE was prepared at three different concentrations, and the OD measured at 280 and 540 nm. The value at 280 nm was divided by 56,840 to obtain the protein concentration and that at 540 nm by 11,500 to obtain the FAD concentration, and then the ratio of the two values determined. The % FAD values reported represent the average of the three calculated concentrations.

Creation of the *lkcE* knockout strain of *S. rochei*. The gene *lkcE* was inactivated in *S. rochei* var. *volubilis* ATCC 21250 by PCR targeting, using a method similar to that described previously⁸. For this, oligonucleotides (Supplementary Table 2) were designed to amplify an apramycin resistance cassette, flanked on both sides with 39 bp of homology to the genomic regions up and downstream of *lkcE*. These primers were used in a PCR reaction with plasmid pIJ773 (ref. 22) encoding for Apra^R. The resulting PCR product was then used to replace the *lkcE* region in cosmid Lc2B12 (kind gift of Professor P.F. Leadlay), a derivative of SuperCos1 (ref. 23). For this, *E. coli* BW25113 was co-transformed with cosmid Lc2B12, plasmid pIJ790 (temperature-sensitive λRed recombination helper plasmid²²), and the PCR fragment. Recombinants were selected by growth on LB supplemented with apramycin, and the presence and correct location of the deletion confirmed by PCR screening and sequencing. *E. coli* ET12567[pUZ8002] was then transformed with the mutant cosmid, and used for conjugation with spores of *S. rochei*. Recombinants were selected for apramycin resistance, and the inactivation of *lkcE* on plasmid pSRV (this plasmid contains the lankacidin cluster⁸) was confirmed by sequencing. The presence of the intermediate (and the absence of later-stage metabolites) in the *lkcE* mutant was confirmed by HPLC-MS analysis, by comparison with the wild-type strain (Supplementary Fig. 8 and Supplementary Table 4).

Isolation of LC-KA05. The *lkcE* mutant of *S. rochei* was grown in YMG medium (3×10 L of culture). After removal of the cells by centrifugation (4550 g, 30 min), the supernatant was extracted with ethyl acetate ($3 \times$ equivalent volume). The organic phase was then evaporated, yielding 100–200 mg of material. Approx. 200 mg crude extract was dissolved in 2 mL methanol and cleared by centrifugation (2 min, 9610 g). The clear supernatant was purified by preparative HPLC on a Nucleodur C₁₈ Isis column (5 μ m, 250×21 mm; Macherey Nagel) using a linear gradient (0 min: 80% A, 20% B; 100 min: 55% A, 45% B; flow = 15 mL min^{-1}) of water (A) and acetonitrile (B). The chromatography was monitored by electrospray ionization (ESI)-MS using a 1:500 static splitter (methanol was used to achieve splitting at a flow rate of $500 \mu\text{L min}^{-1}$) coupled to a ZQ mass spectrometer (capillary voltage 3 kV, 650 L h^{-1} nitrogen, 250°C desolvation temperature; Waters). LC-KA05 showed a typical retention time of 39–48 min. The fractions were pooled and freeze dried yielding 20 mg as a 2:1 mixture with an elimination product ($\text{C}_{25}\text{H}_{35}\text{NO}_6$) **7**, corresponding to loss of the C-7 acetate group. It should be noted that traces of formic acid in the chromatography solvents lead to quantitative formation of the elimination product upon freeze drying. An additional hydrolysis product **6** was detected (free C-7-OH group) at a retention time of 23–27 min. Assignment of the identities of **6** and **7** was based on HR-ESI-MS data (QToF Premier (Waters)), as well as their fragmentation patterns (ESI positive), which show characteristic peaks in common with each other and with **1** (Supplementary Table 4). The ^1H -NMR spectrum of **6** showed the absence of H-27, while the difference NMR spectrum of **7** also revealed the absence of signals corresponding to H-6, H-6' and the C-7 acetate (H-27), as expected (Supplementary Fig. 9).

Spectral data for LC-KA05 (**1**) (and see Supplementary Fig. 9): ^1H -NMR (400 MHz, CD_3OD) δ : 6.35 (d, $J = 15.4 \text{ Hz}$, 1H, H-9), 6.25 (d, $J = 15.6 \text{ Hz}$, 1H, H-15), 5.71 (dd, $J = 6.6$ and 15.6 Hz , 1H, H-14), 5.64–5.45 (m, 4H, H-7, H-8, H-11 and H-17), 4.21–4.16 (m, 2H, H-5 and H-13), 4.12 (q, $J = 7.0 \text{ Hz}$, 1H, H-24), 4.02–3.92 (m, 2H, H-18 and H-18'), 2.50–2.36 (m, 3H, H-4, H-12 and H-12'), 2.22 (ddd, $J = 6.2$, 8.9 and 14.5 Hz , 1H, H-6), 2.05 (s, 3H, H-27), 1.91 (ddd, $J = 4.4$, 6.6 and 14.5 Hz , 1H, H-6'), 1.83 (s, 3H, H-22), 1.76 (s, 3H, H-21), 1.73 (s, 3H, H-19), 1.35 (d, $J = 7.0 \text{ Hz}$, 3H, H-25), 1.27 (d, $J = 7.2 \text{ Hz}$, 3H, H-20) ppm. ^{13}C -NMR (100 MHz, CD_3OD) δ : 176.2 (C-23), 170.7 (C-26), 169.6 (C-3), 168.8 (C-1), 138.3 (C-9), 135.5 (C-16), 134.3 (C-10), 133.9 (C-15), 130.8 (C-14), 129.7 (C-11), 127.2 (C-17), 123.6 (C-8), 96.7 (C-2), 77.7 (C-5), 72.4 (C-7), 71.8 (C-13), 67.7 (C-24), 37.9 (C-18), 36.8 (C-4), 36.5 (C-6), 36.2 (C-12), 19.9* (C-25/C-27), 19.8* (C-25/C-27), 15.4 (C-20), 11.3 (C-22, C-21), 7.2 (C-19). *: These signals cannot conclusively be assigned. HR-ESI-MS: 528.2575 (calc. for $\text{C}_{27}\text{H}_{39}\text{NO}_8\text{Na}^+$: 528.2573).

Spectral data for the elimination product **7**: HR-ESI-MS: 468.2356 (calc. for $\text{C}_{25}\text{H}_{35}\text{NO}_6\text{Na}^+$: 468.2362) and see Supplementary Fig. 9 and Supplementary Table 4.

Spectral data for the purified C-7 hydrolysis product **6**: HR-ESI-MS: 486.2462 (calc. for $\text{C}_{25}\text{H}_{37}\text{NO}_7\text{Na}^+$: 486.2468) and see Supplementary Fig. 9 and Supplementary Table 4.

Crystallization and X-ray data collection. For protein crystallization, LkcE was purified and stored in buffer (30 mM HEPES, 100 mM NaCl, 1 mM EDTA, pH 7.5) at a final concentration of 20 mg mL^{-1} . Homogeneity was checked by dynamic light scattering using a Zetasizer NanoS (Malvern). Native LkcE crystals were produced with the JCSG + kit (Molecular Dimensions), under conditions of 15% PEG 8000, 160 mM calcium acetate, 20% glycerol, 80 mM sodium cacodylate, pH 6.5. The 3 μL drops contained a 2:1 mixture of protein solution (5 mg mL^{-1} , 10 mM DATD, 10 mM EMAA, and 1 mM TCEP). Crystals of SeLkcE were obtained using the hanging drop method in Linbro® plates under conditions of 28% PEG 3350, 200 mM ammonium acetate, 100 mM Bis Tris, pH 6.5. This condition was identified using the INDEX screen (Hampton Research). Drops were formed by mixing 2 μL of a protein solution (8 mg mL^{-1} SeLkcE, DATD 5 mM) with 1 μL of crystallization buffer. Crystals grew in 2–3 days. Crystals of mutants E64Q and R326Q were obtained by microseeding using a Seed Bead kit (Hampton Research). Drops contained 1 μL of crystal shred in 16% PEG 8000, 160 mM calcium acetate, 20% glycerol, 80 mM sodium cacodylate, pH 6.6, and 2 μL of protein solution (6 mg mL^{-1}). Crystals were soaked in crystallization buffer with 30% ethylene glycol before being frozen in liquid nitrogen. The crystals of mutants were additionally soaked in cryoprotection buffer containing 30 mM LC-KA05, before cryo-cooling. X-ray diffraction data on SeLkcE and native/mutant LkcE were collected at the SOLEIL synchrotron on the Proxima2 and Proxima1 beamlines, respectively. The images were integrated using X-ray diffraction spectroscopy in the space group $P4_12_12$.

Structure determination and refinement. As LkcE shares only 25% sequence identity with the closest homolog in the PDB (6HDNO (PDB 3NG7)¹⁴) we acquired a complete multiple wavelength anomalous diffraction data set on SeLkcE. Initial phases were ultimately generated using SAD using the peak wavelength ($\lambda = 0.9793 \text{ nm}$). Sixteen selenium positions (out of a possible 18) were located with the SHELX³⁷ package. An initial model was built automatically using BUCANEER^{38,39} with several cycles of manual rebuilding in Coot⁴⁰ and refinement with Buster⁴¹. The structures of native LkcE and mutants were solved by molecular replacement with Phaser MR⁴² using the structure of SeLkcE as the search model, manually

rebuilt in Coot and refined with Buster. All the structures were validated using the program MolProbity⁴³. All protein structure figures were prepared using the program PyMol (Schrödinger, LLC). Data collection, refinement, and validation statistics are presented in Supplementary Table 3.

LkcE activity tests in vitro with substrate analogs. LkcE ($50 \mu\text{M}$) was incubated with 1, 5, or 50 mM of the two substrate analogs (in combination or separately) at 25°C for 1 h in buffer (30 mM HEPES, 150 mM NaCl, 1 mM EDTA, pH 7.5), and the reaction was stopped by adding $2 \times$ the equivalent volume of ethyl acetate. The organic phase was evaporated overnight at room temperature and then the residue analyzed by HPLC-MS (see section 2.1.3), after resuspension in H_2O /acetonitrile (80:20 v/v).

LkcE activity tests in vitro with the native substrate. Steady-state kinetics parameters were determined at 25°C in gel filtration buffer (30 mM HEPES, 150 mM NaCl, 1 mM EDTA, pH 7.5) with variable concentrations of substrate, either the LC-KA05 (**1**)/elimination derivative product **7** mixture, the deacetylated derivative **6**, or the substrate analogs EMAA (**4**) and DATD (**5**) (in combination and separately). Reactions were carried out with $2 \mu\text{M}$ LkcE (and additionally with $4 \mu\text{M}$ enzyme in experiments with the substrate analogs **4** and **5**) in the presence of NADH peroxidase from *S. faecalis* (0.5 U mL^{-1}) and NADH (0.3 mM). Initial rate measurements were obtained on a SAFAS UVmc² spectrophotometer by following the oxidation of NADH at 340 nm. Where appropriate, data were fitted to the Michaelis–Menten equation using least-squares regression analysis to determine k_{cat} and K_{M} . We confirmed that the NADH peroxidase was not rate-limiting in the reaction by showing that doubling its concentration (from 3 to 6 U mL^{-1}) had no effect on the rate, whereas increasing the concentration of LkcE by a factor of five (from 5 to $250 \mu\text{M}$) produced the expected fivefold increase in velocity.

Large-scale assays for analysis by HPLC-MS were carried out with $20 \mu\text{M}$ active LkcE or denatured enzyme (obtained via incubation for 10 min at 95°C) in buffer (30 mM HEPES, 150 mM NaCl, 1 mM EDTA, pH 7.5). In total, $100 \mu\text{M}$ substrate (LC-KA05 **1**/elimination derivative **7** mixture or deacetylated derivative **6**) was added every 10 min, until a final concentration of 1 mM substrate was reached. Enzymatic reactions were carried out overnight at 25°C . One milliliter of ethyl acetate was then added and the mixture was thoroughly vortexed. The organic phase was separated from the aqueous phase, and the extraction repeated twice. The combined organic layers were evaporated to dryness on a SpeedVac concentrator.

HPLC-MS analysis. HPLC-MS analysis was performed using an HPLC (Dionex, Ultimate 3000) coupled to a LTQ Orbitrap XL hybrid mass spectrometer (Thermo Scientific) fitted with an ESI source. HPLC-MS data were processed using Xcalibur (v. 2.1) (Thermo Scientific). For analysis of the compounds of interest, the HPLC was fitted with an Alltima $5 \mu\text{m}$ C18 column ($150 \text{ mm} \times 2.1 \text{ mm}$, Grace Alltech) column. A solvent system of acetonitrile and water both containing 0.1% formic acid (v/v) was used. Samples were first eluted with a linear gradient from 20–35% acetonitrile over 20 min, followed by a second linear gradient from 35–90% acetonitrile over 15 min, at a flow rate 0.2 mL min^{-1} . The mass spectrometer was run in either positive or negative ionization modes scanning from m/z 100–1000. Mass spectrometric conditions were as follows for ESI⁺ mode: spray voltage was set at +4.5 kV; source gases were set (in arbitrary units min^{-1}) for sheath gas, auxiliary gas, and sweep gas at 30, 10, and 10 respectively; capillary temperature was set at 275°C ; capillary voltage at +4 V; and tube lens, split lens, and front lens voltages at +155 V, –28 V, and –6 V, respectively. The MS parameters for the ESI[–] mode were directly transposed from the ESI⁺ mode. To confirm some structures, MS² fragmentations in the ion trap and exact mass scans using the Orbitrap analyzer were carried out.

Data availability

Protein Data Bank coordinates for seleniated *holo* LkcE wild type, LkcE wild type in complex with EMAA (**4**) and DATD (**5**), and the LkcE E64Q and R326Q mutants in complex with LC-KA05 (**1**), have been deposited under accession codes 6FJH, 6F32, 6F7V, and 6F7L, respectively. All data presented in the manuscript are available upon reasonable request from the corresponding authors.

Received: 26 March 2018 Accepted: 20 August 2018

Published online: 28 September 2018

References

1. Tyzack, J. D., Furnham, N., Sillitoe, I., Orenco, C. M. & Thornton, J. M. Understanding enzyme function evolution from a computational perspective. *Curr. Opin. Struct. Biol.* **47**, 131–139 (2017).

2. Furnham, N., Dawson, N. L., Rahman, S. A., Thornton, J. M. & Orengo, C. A. Large-scale analysis exploring evolution of catalytic machineries and mechanisms in enzyme superfamilies. *J. Mol. Biol.* **428**, 253–267 (2016).
3. Eberhardt, R. Y. et al. Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics* **14**, 327 (2013).
4. Weiss, A. K. H., Loeffler, J. R., Liedl, K. R., Gstach, H. & Jansen-Dürr, P. The fumarylacetoacetate hydrolase (FAH) superfamily of enzymes: multifunctional enzymes from microbes to mitochondria. *Biochem. Soc. Trans.* **46**, 295–309 (2018).
5. Gerlt, J. A., Babbitt, P. C., Jacobson, M. P. & Almo, S. C. Divergent evolution in enolase superfamily: strategies for assigning functions. *J. Biol. Chem.* **287**, 29–34 (2012).
6. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).
7. Horsman, M. E., Hari, T. P. A. & Boddy, C. N. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? *Nat. Prod. Rep.* **33**, 183–202 (2016).
8. Arakawa, K., Sugino, F., Kodama, K., Ishii, T. & Kinashi, H. Cyclization mechanism for the synthesis of macrocyclic antibiotic lankacidin in *Streptomyces rochei*. *Chem. Biol.* **12**, 249–256 (2005).
9. Gaweska, H. & Fitzpatrick, P. F. Structures and mechanism of the monoamine oxidase family. *Biomol. Concepts* **2**, 365–377 (2011).
10. Ames, B. D. et al. Complexity generation in fungal peptidyl alkaloid biosynthesis: oxidation of fumiquinazoline A to the heptacyclic hemiaminal fumiquinazoline C by the flavoenzyme Af12070 from *Aspergillus fumigatus*. *Biochemistry* **50**, 8756–8769 (2011).
11. Wu, L.-L., Xiang, Y., Yang, D.-C., Guan, Z. & He, Y.-H. Biocatalytic asymmetric Mannich reaction of ketimines using wheat germ lipase. *Catal. Sci. Technol.* **6**, 3963–3970 (2016).
12. Dym, O. & Eisenberg, D. Sequence-structure analysis of FAD-containing proteins. *Protein Sci.* **10**, 1712–1728 (2001).
13. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
14. 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. USA* **108**, 4800–4805 (2011).
15. Son, S.-Y. et al. Structure of human monoamine oxidase A at 2.2-Å resolution: the control of opening the entry for substrates/inhibitors. *Proc. Natl. Acad. Sci. USA* **105**, 5739–5744 (2008).
16. Binda, C., Newton-Vinson, P., Hubálek, F., Edmondson, D. E. & Mattevi, A. Structure of human monoamine oxidase B, a drug target for the treatment of neurological disorders. *Nat. Struct. Biol.* **9**, 22–26 (2002).
17. Qin, X. et al. Structural insight into unique properties of protoporphyrinogen oxidase from *Bacillus subtilis*. *J. Struct. Biol.* **170**, 76–82 (2010).
18. 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).
19. Sillitoe, I. et al. CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* **43**, D376–D381 (2015).
20. Fagan, R. L. & Palfrey, B. A. In *Comprehensive Natural Products II: Chemistry and Biology* (eds Liu, H.-W. & Mander, L.) 37–113 (Elsevier, 2010).
21. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. USA* **100**, 1541–1546 (2003).
22. Dickschat, J. S. et al. An additional dehydratase-like activity is required for lankacidin antibiotic biosynthesis. *Chembiochem* **12**, 2408–2412 (2011).
23. Lu, C. et al. Two new lankacidin-related metabolites from *Streptomyces* sp. HS-NF-1178. *J. Antibiot.* **71**, 397–401 (2018).
24. Parsonage, D., Miller, H., Ross, R. P. & Claiborne, A. Purification and analysis of streptococcal NADH peroxidase expressed in *Escherichia coli*. *J. Biol. Chem.* **268**, 3161–3167 (1993).
25. Lowry, B. et al. In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc.* **135**, 16809–16812 (2013).
26. Guillén Schlippe, Y. V. & Hedstrom, L. A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Arch. Biochem. Biophys.* **433**, 266–278 (2005).
27. Reimer, D. & Bode, H. B. A natural prodrug activation mechanism in the biosynthesis of nonribosomal peptides. *Nat. Prod. Rep.* **31**, 154–159 (2014).
28. Chen, R. Enzyme engineering: rational redesign versus directed evolution. *Trends Biotechnol.* **19**, 13–14 (2001).
29. Steiner, K. & Schwab, H. Recent advances in rational approaches for enzyme engineering. *Comput. Struct. Biotechnol. J.* **2**, e201209010 (2012).
30. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
31. Sambrook, J. & Russell, D. *Molecular Cloning: A Laboratory Manual* 3rd edn (Cold Spring Harbor Laboratory Press, New York, 2000).
32. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Streptomyces Genetics* (The John Innes Foundation, Norwich, UK, 2000).
33. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645 (2000).
34. Thorn, A. Experimental phasing: substructure solution and density modification as implemented in SHELX. *Methods Mol. Biol.* **1607**, 357–376 (2017).
35. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 1002–1011 (2006).
36. Cowtan, K. Fitting molecular fragments into electron density. *Acta Crystallogr. D. Biol. Crystallogr.* **64**, 83–89 (2008).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
38. Bricogne, G. et al. *BUSTER version 2.10.2* (Global Phasing Ltd., Cambridge, United Kingdom, 2017).
39. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
40. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).
41. Tatsuno, S., Arakawa, K. & Kinashi, H. Analysis of modular-iterative mixed biosynthesis of lankacidin by heterologous expression and gene fusion. *J. Antibiot.* **60**, 700–708 (2007).
42. He, H.-Y., Yuan, H., Tang, M.-C. & Tang, G.-L. An unusual dehydratase acting on glycerate and a ketoreductase stereoselectively reducing α -ketone in polyketide starter unit biosynthesis. *Angew. Chem. Int. Ed. Engl.* **53**, 11315–11319 (2014).
43. 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).
44. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
45. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003).
46. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOLE – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
47. Fischer, H., de Oliveira Neto, M., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Cryst.* **43**, 101–109 (2010).
48. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
49. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2001).

Acknowledgements

Funding for this work was provided by the Agence Nationale de la Recherche (ANR-11-JSV8-003-01 and ANR-16-CE92-0006-01 to K.J.W.), the Centre National de la Recherche Scientifique (CNRS), the 'IMPACT Biomolecules' project of the Lorraine Université d'Excellence (Investissements d'avenir—ANR 15-004), the Lorraine Region (Bonus Qualité Recherche (BQR) grants to K.J.W. and B.C.) and the Deutsche Forschungsgemeinschaft (Cluster of Excellence REBIRTH, "From Regenerative Biology to Reconstructive Therapy" EXC 62 (to G.D. and A.K.), and grant Ki 397/20-1 (to A.K.)). Jean-Christophe Lec is thanked for invaluable assistance with the kinetics analysis, Russell Cox for helpful discussions, and Peter Leadlay for suggesting the use of EMDA and DATD and for editing of the manuscript. We also acknowledge the following for help in data acquisition: Pierre Legrand and Andrew Thompson (Soleil Synchrotron, Beamline Proxima1), William Shepard and Martin Savko (Soleil Synchrotron, Proxima2), and Javier Perez and Aurelien Thureau (Soleil Synchrotron, Swing). A portion of the NMR data was recorded on the NMR spectrometer of the Plateforme de Biophysique et Biologie Structurale (B2S) (IBSLor, UMS2008, CNRS-UL-INSERM).

Author contributions

J.D. cloned, expressed and purified LkCE, crystallized LkCE, solved the structures, and analyzed data (along with A.G.). F.R. carried out the kinetic analysis of wild-type LkCE and mutants. C.J. generated the LkCE knockout and cultured and extracted the resulting mutant. S.C. contributed to isolation of LC-KA05 (along with B.C., J.D., and A.G.) and the crystallography. G.D. and A.K. purified and structurally characterized LC-KA05. C.P. carried out analysis of the lankacidin metabolites by HPLC-MS, and helped with data interpretation by K.J.W. A.G. and K.J.W. designed the study,

supervised the research, and analyzed the data. K.J.W. wrote the paper, with input from all other authors.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-06323-w>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

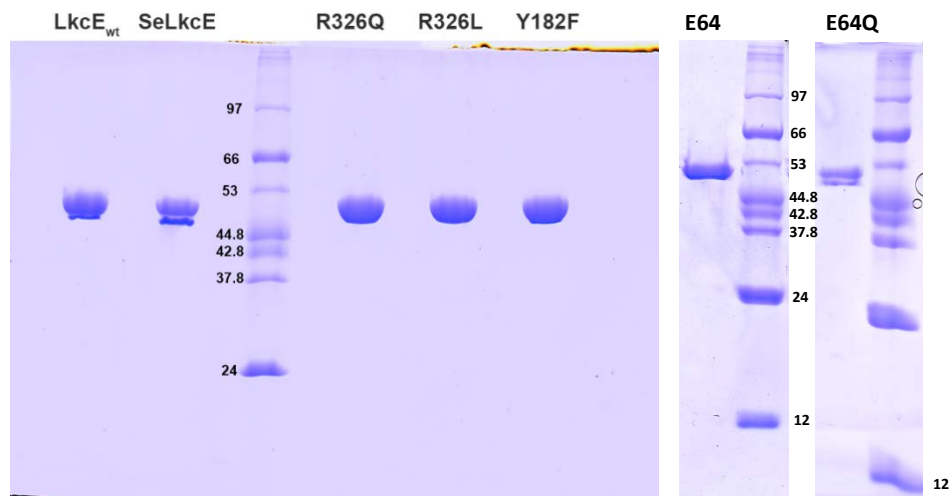
Supplementary Information

Insights into a dual function amide oxidase/macrocyclase from lankacidin biosynthesis

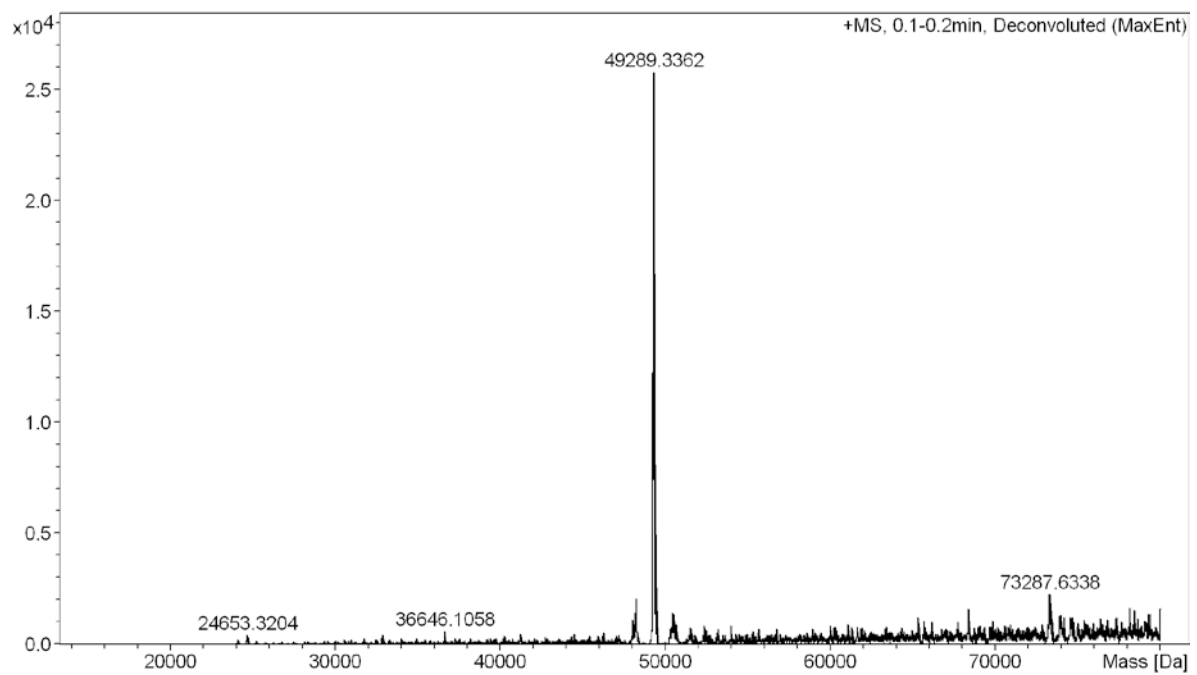
Dorival, *et al.*

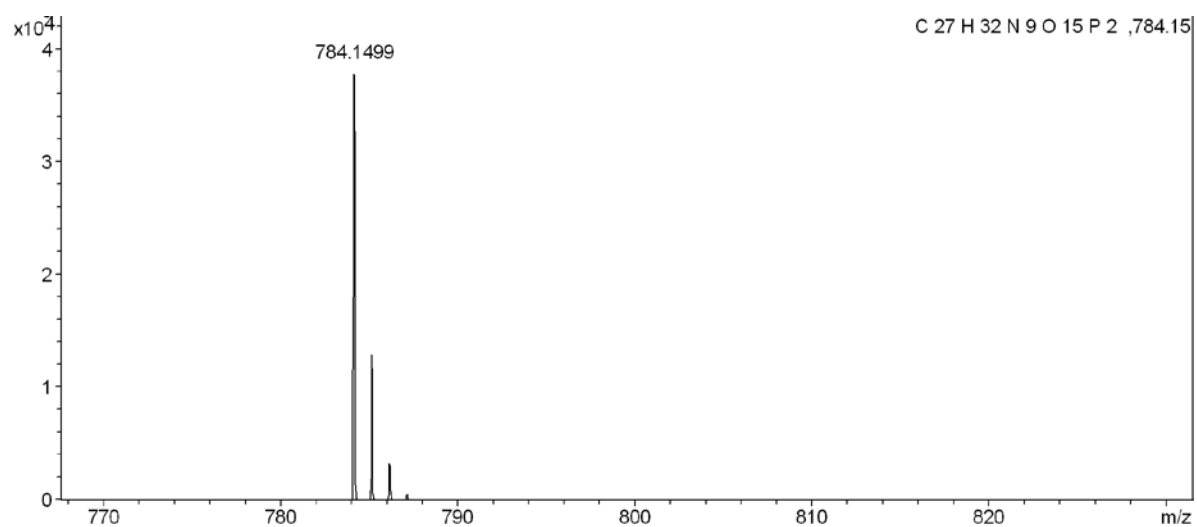
Supplementary Figures

a

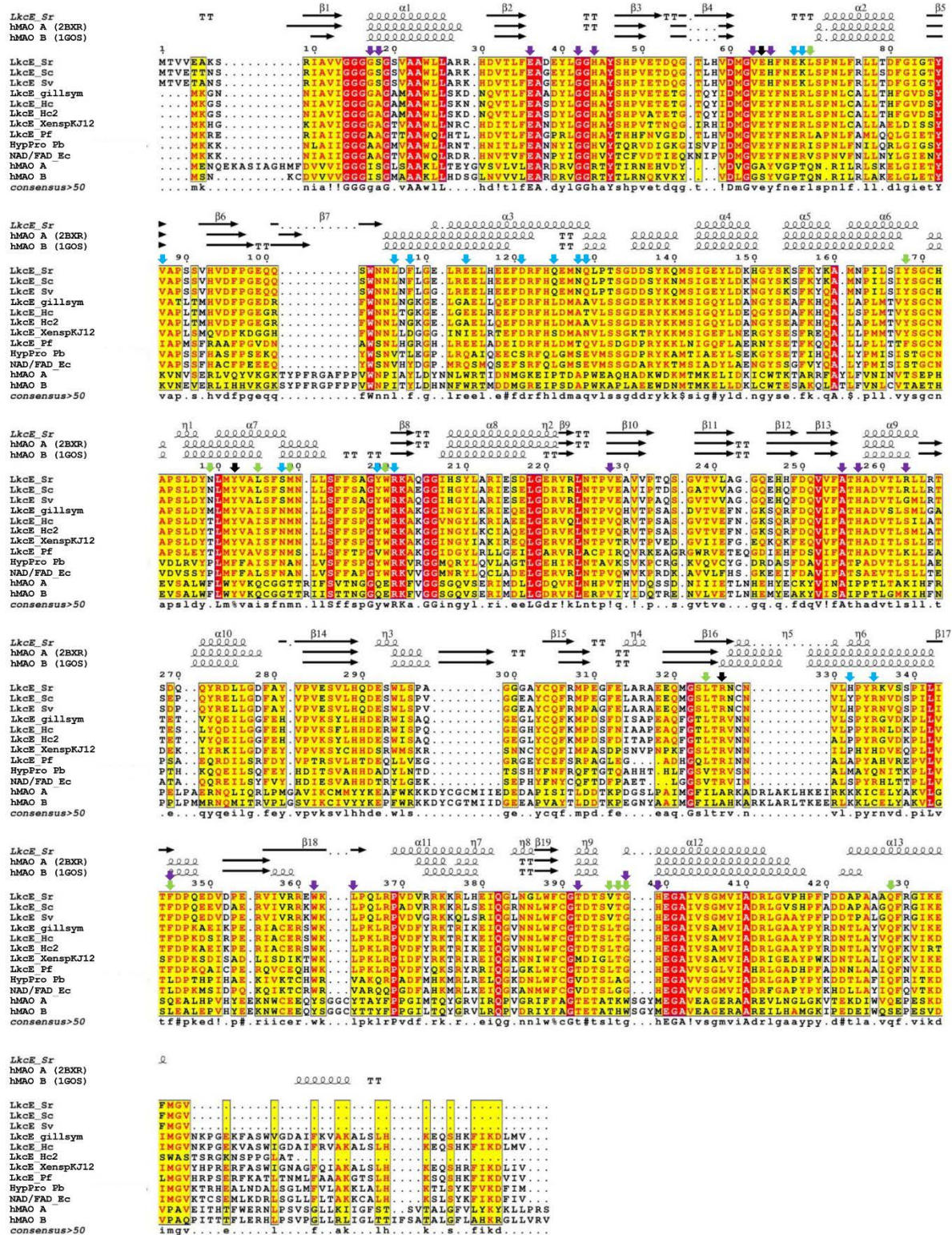


b



c

Supplementary Figure 1 | Characterization of recombinant LkcE. (a) SDS-PAGE (12.5%) gels of purified LkcE (seleniated and non) and its 5 mutants. The molecular weights of the markers are indicated. (b) Deconvoluted mass spectrum of non-seleniated LkcE, giving a mass of 49,289.3 Da (vs. the theoretical mass of 49,289.4 Da). (c) Analysis by mass spectrometry of the liberated cofactor. The measured mass ($m/z = 784.15$ $[M-H]^-$) corresponds to that of FAD (785.5 Da).



b

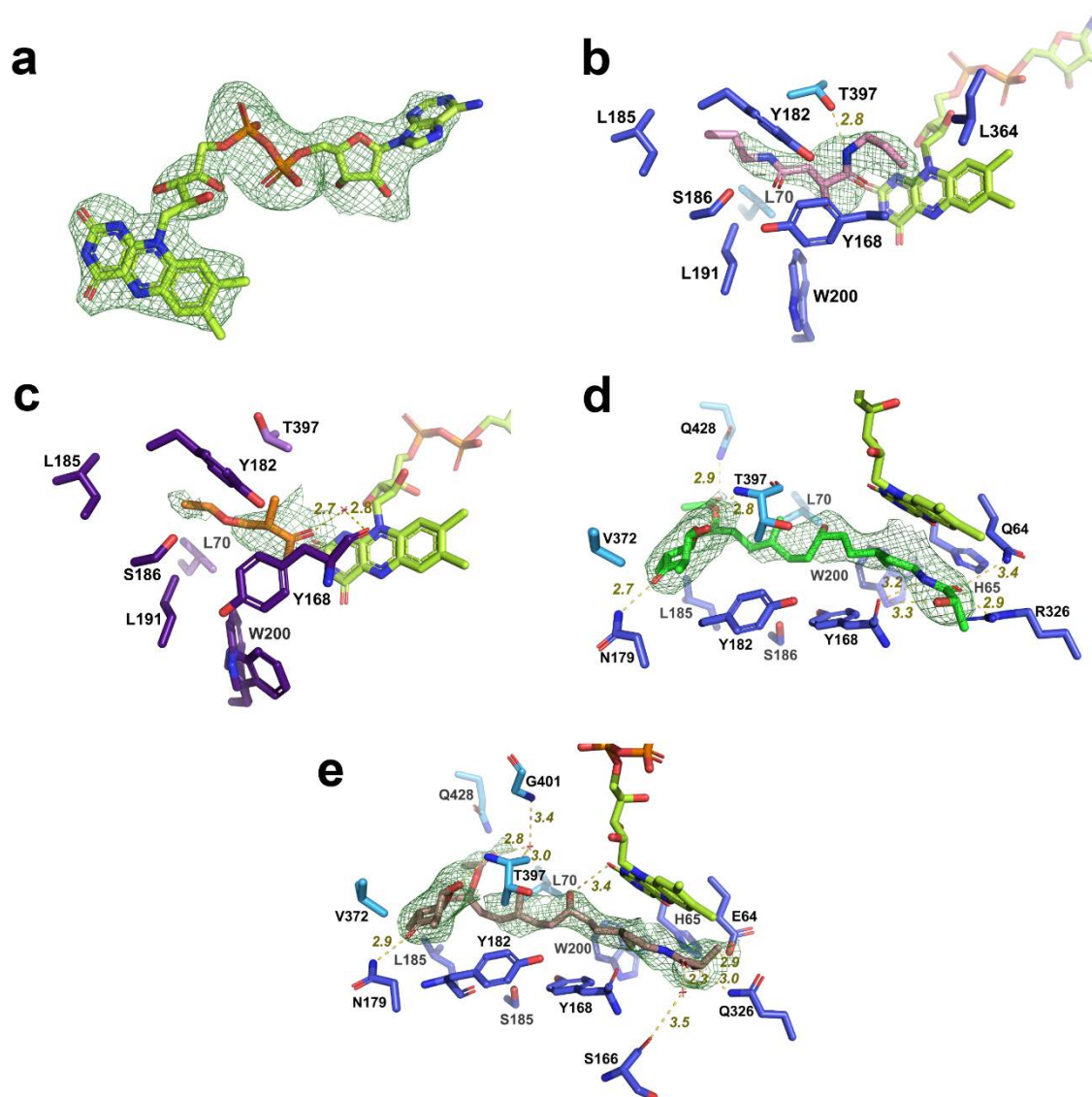
Analysis for the genomic context of the LkcE homologs

LkcE homolog (designation in (a))	NCBI gene sequence ID	Genomic context (translated protein products)
<i>S. rochei</i> var. <i>volubilis</i> (LkcE_Sr)	Not deposited, but differs by 1 amino acid (P424L) from WP_011113095 (ref. 3) and has the same genomic context	PqqB PqqE PqqD PqqC PqqA ABC transporter ABC transporter Isochorismatase LkcG LkcF LkcE LkcD LkcC LkcB LkcA MaoC-like dehydrogenase Efflux transporter PCP PKS FkbH-like PQQ-dependent dehydrogenase Prx Anti-sigma factor PKS protein (AT-KR-ACP-TE) PQQ-dependent DH LkcA LkcB homolog? (KS-DH?) LkcE (2 proteins which have been removed from the database) MerR family transcriptional regulator Hypothetical protein PQQ-dependent dehydrogenase PKS FkbH-like Carrier protein homolog SDR homolog LkcA LkcB LkcC LkcD LkcE LkcF LkcG Nicotinamidase ABC transporter ABC transporter PqqC PqqD PqqE PqqB LkcG LkcF LkcE LkcD Nicotinamidase ^a LkcG ACP-KS? KS? KS-KR-ACP-KS (portion of LkcF?) KR-ACP (portion of LkcF?) Nicotinamidase LkcE LkcD LkcC
<i>S. griseofuscus</i>	WP_051850692	
<i>S. cellostaticus</i> (LkcE_Sc)	KUM93544	
<i>S. yokosukanensis</i> (LkcE_Sy)	WP_067137025	
<i>Solemya velum</i> gill symbiont (LkcE_gillsym)	WP_078462125	

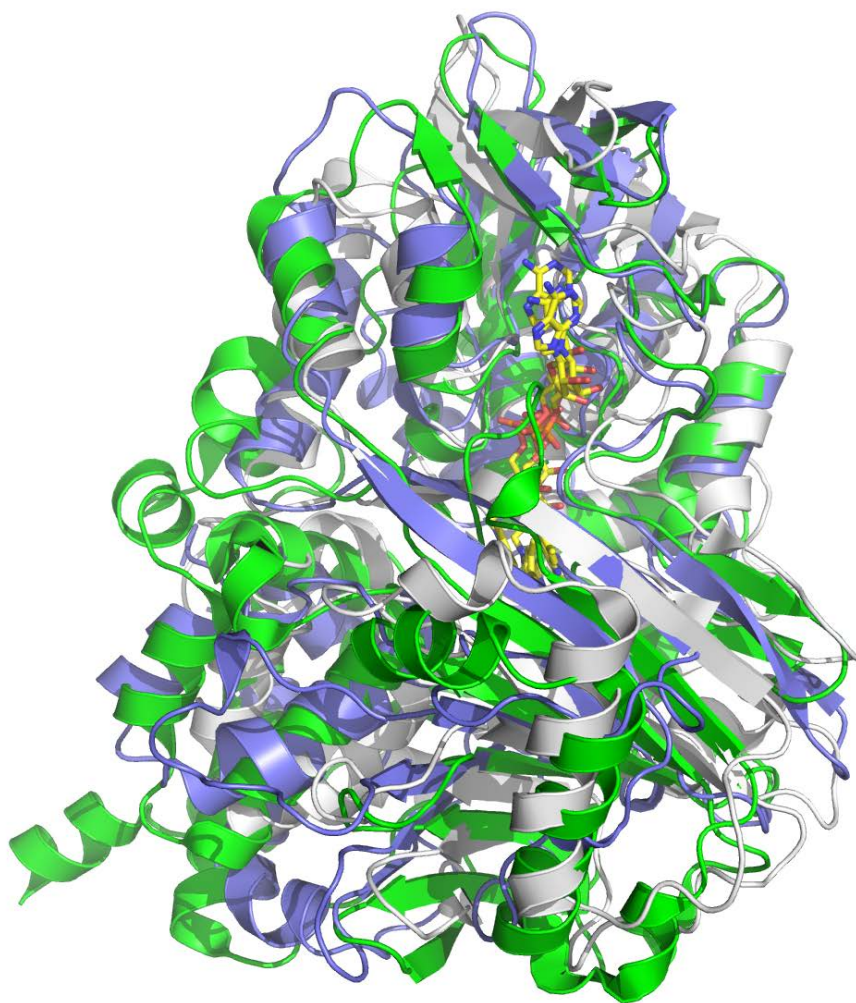
<i>H. chejuensis</i> (LkcE_Hc)	CCG06111	LkcB LkcA Sfp-type phosphopantetheinyl transferase (PPTase) ACP? LkcG LkcF Isochorismatase LkcE LkcD LkcC LkcB LkcA Transcriptional regulator Carrier protein homolog MFS transporter LkcG LkcF Nicotinamidase LkcE LkcD LkcC LkcB LkcA MFS transporter LkcG LkcF Nicotinamidase LkcE LkcD LkcC LkcB LkcA Carrier protein homolog ABC transporter Sfp-type PPTase LkcG LkcF Nicotinamidase LkcE LkcD LkcC LkcB LkcA MFS transporter Hypothetical protein Record removed LkcE homolog Hypothetical protein Paraquat-inducible protein A Hypothetical protein PhzD PhzE PhzF PhzG AMP-dependent ligase Discrete AT A-PCP Hypothetical protein LkcE homolog 3-deoxy-7-phosphoheptulonate synthase PPTase
<i>H. chejuensis</i> (LkcE_Hc2)	WP_011397301	
<i>Xenorhabdus</i> sp. KJ12.1 (LkcE_XenspKJ12)	WP_099111721	
<i>Pseudomonas fluorescens</i> (LkcE_Pf)	WP_096820043	
<i>Pectobacterium betavasculorum</i> (HypPro_Pb)	WP_072012528	
<i>E. coli</i> (NAD/FAD_Ec)	WP_096961323	

^aOn either side of these genes, the NCBI numbers correspond to genes present in *S. celesticus*.

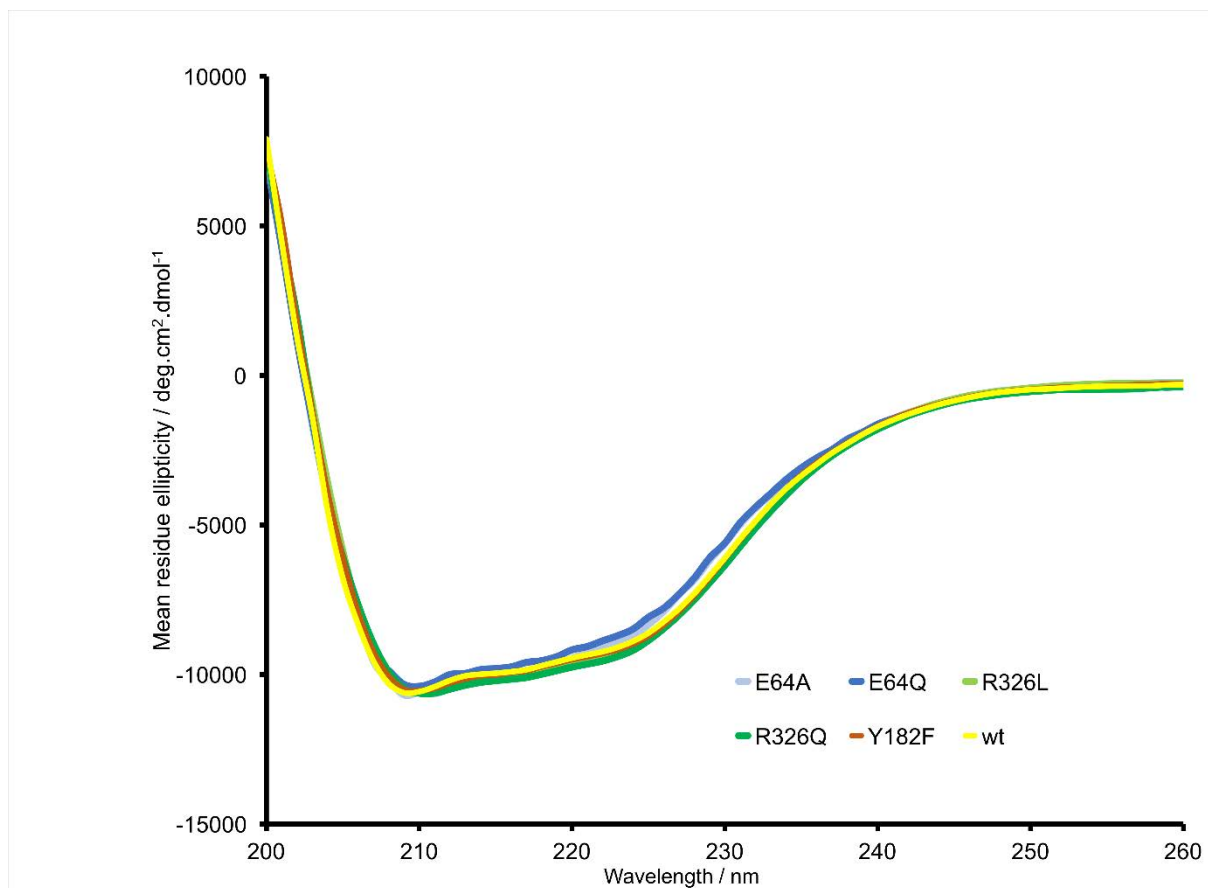
Supplementary Figure 2 | Analysis of LkcE homologs. (a) Multiple sequence alignment of LkcE with various homologs. The sequences include LkcE from *Streptomyces rochei* and its 9 closest homologs identified using BlastP⁴, as well as human amine oxidases A and B, the closest amine oxidase homologs for which structures are available (PDB accession codes 2BXR and 1GOS, respectively) (the secondary structure elements present in each structure are indicated). The sequences include: LkcE *Streptomyces rochei* (Sr) (100% identity with WP_051850692 in *Streptomyces griseofuscus*, which lies in a partial lankacidin cluster (see part (b)); LkcE *Streptomyces cellostaticus* (Sc) (KUM93544): 91% I; 94% S, almost complete lankacidin cluster; LkcE *Streptomyces yokosukanensis* (Sy) (KUM94155): 91% I; 94% S, only partial cluster information available; LkcE (NAD/FAD-binding protein *Solemya velum* gill symbiont) (gillsym) (WP_078462125) 61% I; 79% S, partial lankacidin cluster; LkcE *Hahella chejuensis* (Hc) (CCG061111): 63% I; 79% S, lies in the chejuenolide cluster, which is responsible for biosynthesis of a close relative of lankacidin C⁵; LkcE *Hahella chejuensis* (Hc2) (WP_011397301): 61% I; 79% S, partial lankacidin/chejuenolide cluster; LkcE (NAD/FAD-binding protein *Xenorhabdus* species KJ12.1) (WP_099111721) 56% I; 71% S, partial lankacidin cluster; LkcE *Pseudomonas fluorescens* (Pf) (WP_096820043): 54% I; 66% S, partial lankacidin cluster; hypothetical protein *Pectobacterium betavascuorum* (HyPro_Pb) (WP_072012528): 46% I; 65% S; does not lie in a lankacidin cluster; NAD/FAD binding protein *Escherichia coli* (NAD/FAD_Ec) (WP_096961323): 46% I; 64% S, does not lie in a lankacidin cluster; Human amine oxidase A, 25% I; Human amine oxidase B, 25% I. Key to color-coding of residues indicated by arrows: **purple** = residues implicated directly in FAD binding; **black** = residues mutated in this study; **green** = residues observed to interact with bound LC-KA05; **blue** = residues contributing to LkcE homodimerization. (b) Genomic context supporting assignment of 7 *S. rochei* LkcE homologs (among the 9 in (a)) as lankacidin cyclases.



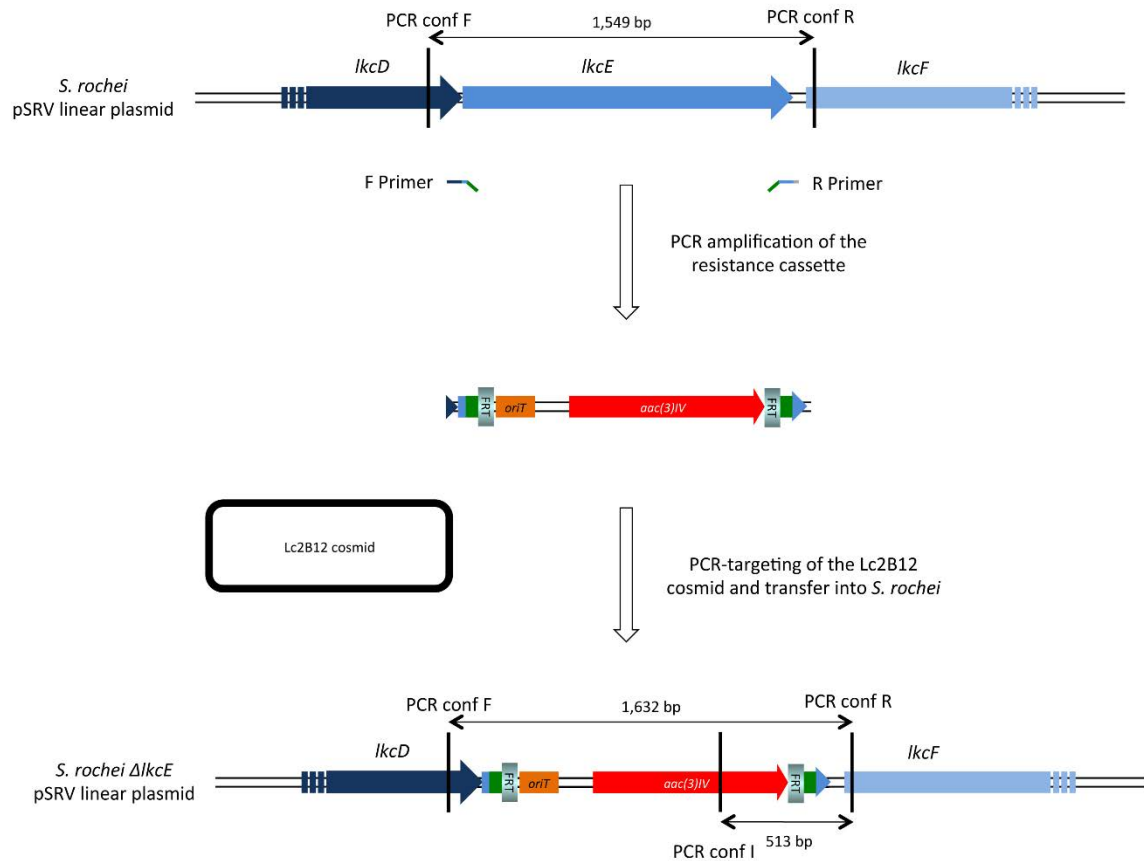
Supplementary Figure 3 | Omit maps for FAD and ligands bound into LkcE. (a) Omit-map calculated for the FAD (yellow) in the structure of the wild type LkcE. The $F_o - F_c$ map is contoured at 2.5σ . (b) Omit-map calculated for the structure of the wild type LkcE. The $F_o - F_c$ electron density corresponding to bound DATD (pink) is contoured at 2.5σ . (c) View of the active site of wild type LkcE in the presence of bound EMAA (orange) with the omit $F_o - F_c$ map contoured at 1.5σ . (d) Omit-map calculated for the structure of the LkcE E64Q mutant in the presence of the bound LC-KA05 (green). The $F_o - F_c$ map contoured at 2.5σ indicates the position of the bound substrate. (e) View of the active site of R326Q mutant in the presence of bound LC-KA05 (brown) with the omit $F_o - F_c$ map contoured at 2.5σ .



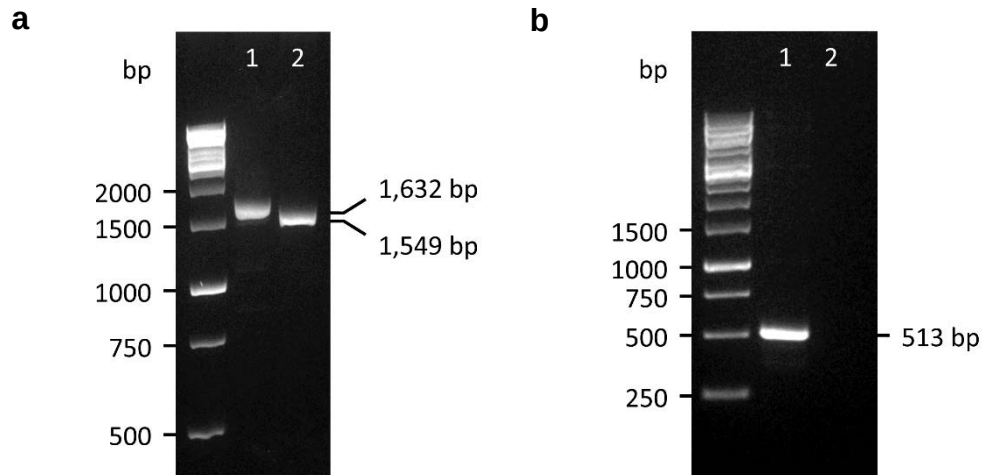
Supplementary Figure 4 | Superposition of a monomer of LkcE with the monomers of two structural homologs. The LkcE monomer is shown in violet, while those of 6HDNO (PDB 3NG7)⁶ and hMAO B (PDB 1GOS)⁷ are represented in grey and green, respectively. The FAD cofactor present in each structure is shown in stick form, with conventional atom coloring. This superposition illustrates in a qualitative manner that the cofactor binding domains are more highly conserved than the substrate-binding domains.



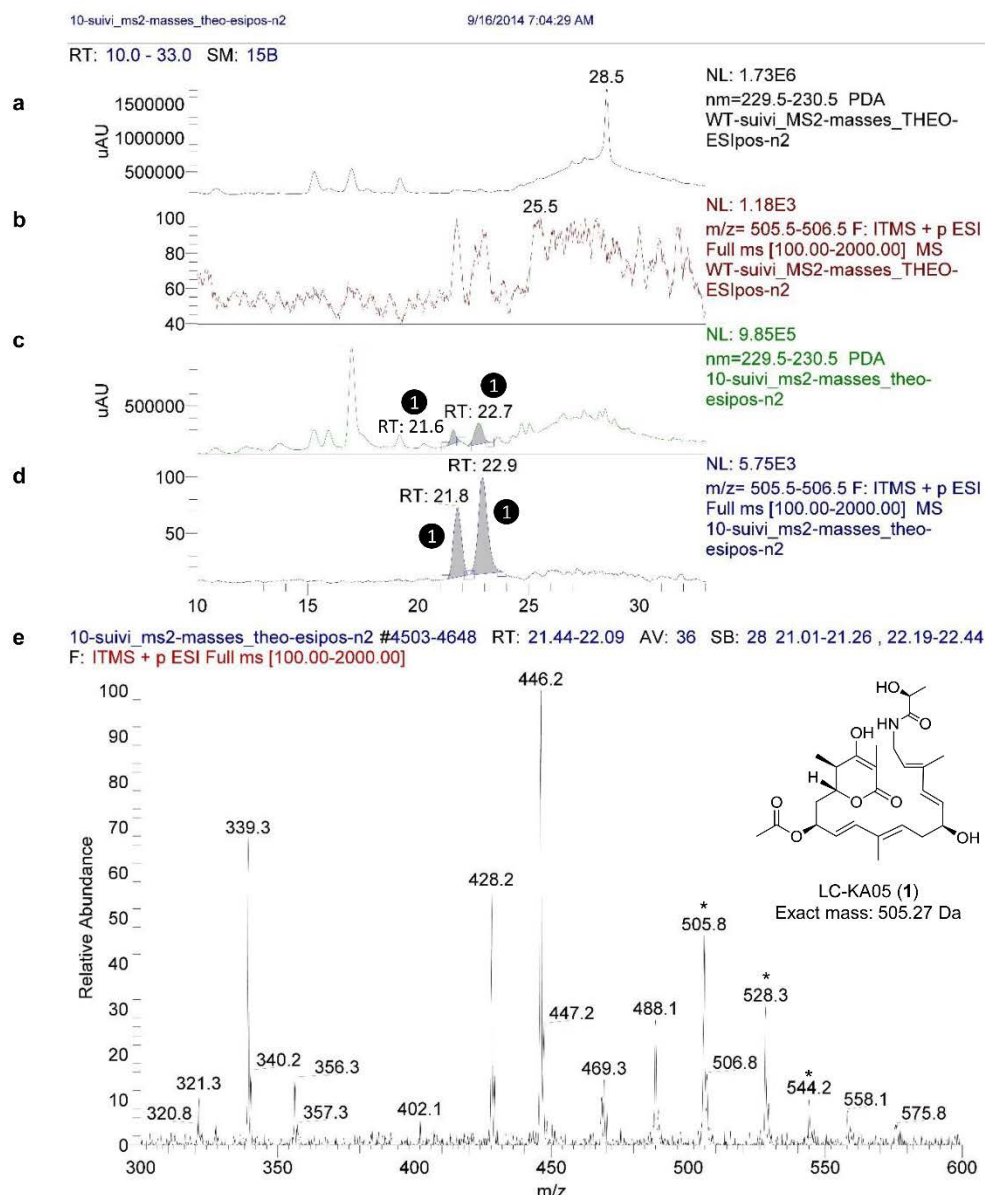
Supplementary Figure 5 | Analysis by circular dichroism of LkcE wild type and its five active site mutants. As the spectra are essentially superimposable, this analysis showed that the mutations did not significantly alter the structure of the enzyme.



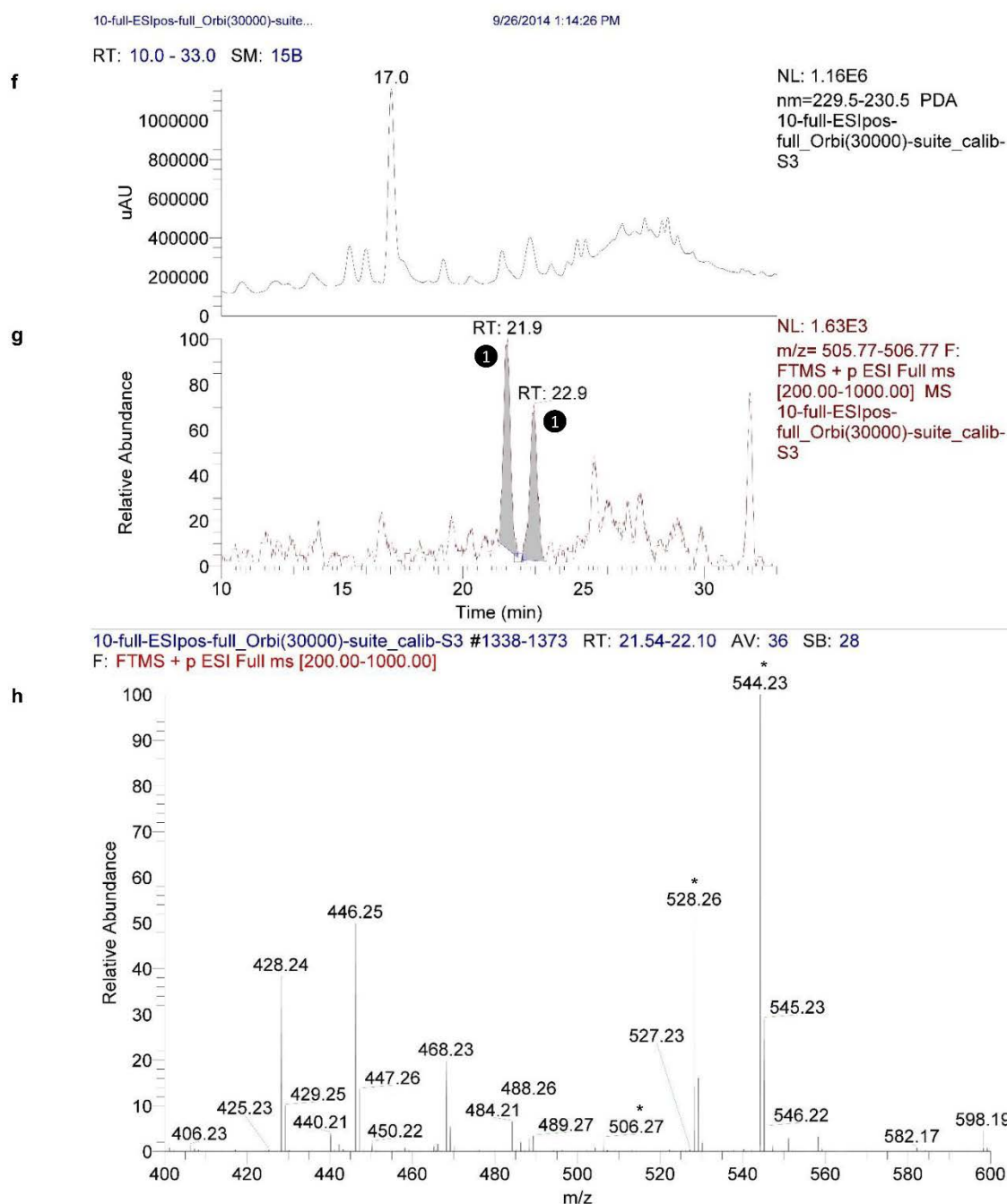
Supplementary Figure 6 | Schematic illustration of the *lkcE* gene disruption in *Streptomyces rochei* by PCR-targeting¹. The resistance cassette amplified from the pIJ773 plasmid using the F and R primers (Supplementary Table 2) was used to replace the *lkcE* gene in the cosmid Lc2B12 (Kan^R). The mutant cosmid was then transformed into *S. rochei* and clones (Kan^S Apr^R) were selected and analyzed by PCR (the binding locations of the primers use to confirm the deletion (PCR conf F, PCR conf R and PCR conf I (Supplementary Table 2)) are indicated).



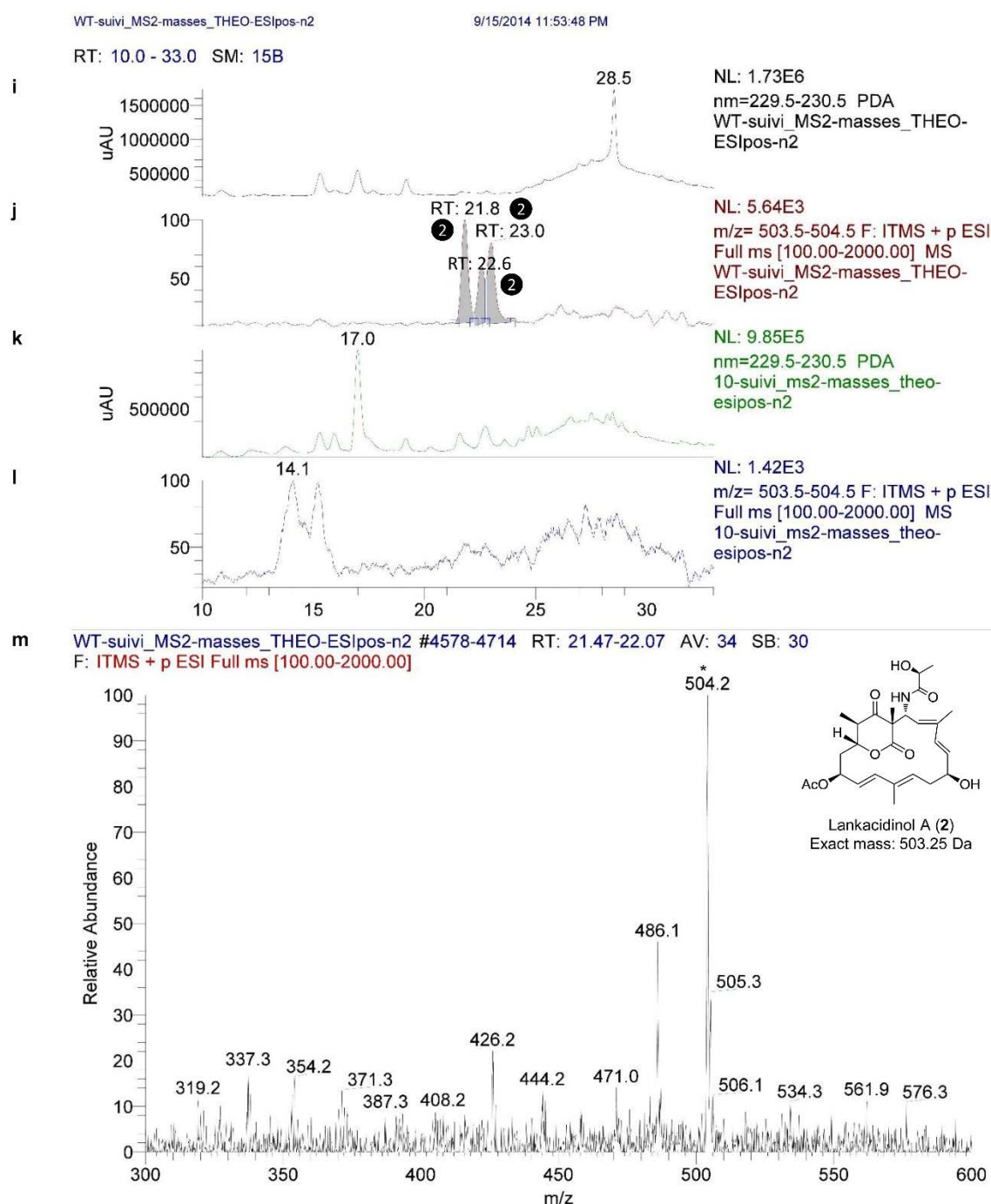
Supplementary Figure 7 | PCR analysis of the lankacidin biosynthetic gene cluster using genomic DNA from *S. rochei* Δ *lkcE* (lanes 1 in panels (a) and (b)) and *S. rochei* WT (lanes 2 in panels (a) and (b)) strains. (a) The product of amplification of the wild type *lkcE* locus using primers conf F/R has the expected size of 1549 bp (as seen in lane 2), while replacement of *lkcE* by the apramycin resistance cassette gives a product of 1632 bp (as seen in lane 1). (b) No product is expected with the wild type strain and primers conf I/R (as with lane 2), while when the apramycin resistance cassette is present, this primer pair yields a product of 513 bp (as in lane 1).



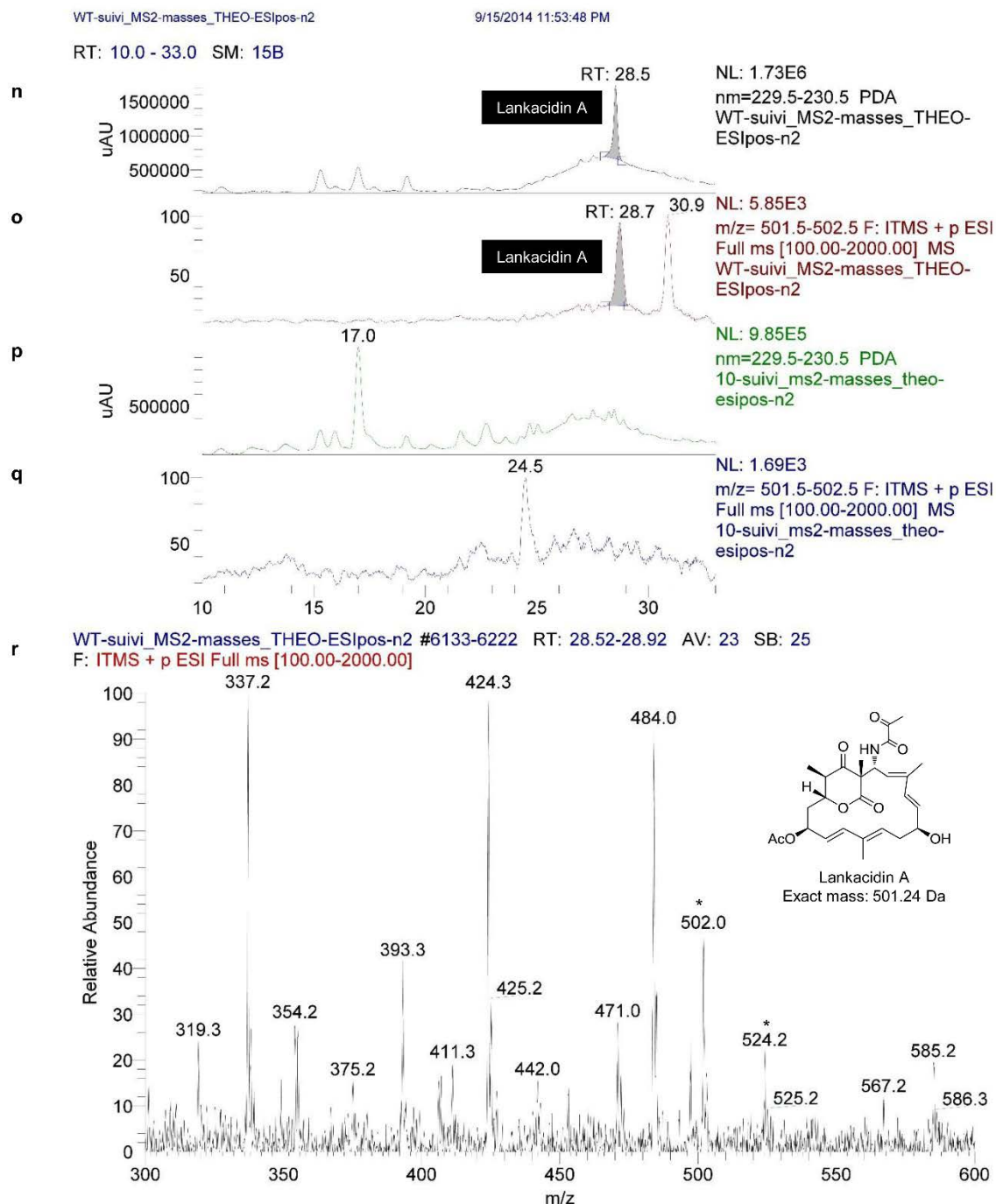
Supplementary Figure 8 | Analysis by HPLC-MS of extracts of wild type *S. rochei* vs. the $\Delta lkcE$ inactivation mutant for LC-KA05 (1) and downstream products. m/z values obtained using the ion trap are reported to one significant figure, while those acquired in the Orbitrap mode are reported to two significant figures. **(a)–(i)** Analysis for the presence of LC-KA05 (1) (exact mass = 505.27 Da). **(a)** Trace of the absorbance at 230 nm of extracts of the wild type. **(b)** SIM of $m/z = 506$ (ESI positive) of extracts of the wild type. **(c)** Trace of the absorbance at 230 nm of extracts of the $\Delta lkcE$ inactivation mutant. The two peaks corresponding to 1 (as indicated), which are not observed in wild type extracts, are greyed. **(d)** SIM of $m/z = 506$ (ESI positive) of extracts of the $\Delta lkcE$ inactivation mutant. The two peaks corresponding to 1 are greyed. **(e)** Mass spectrum of the peak at RT = 21.8 min (which is representative of the second peak at 22.9 min), showing ions consistent with 1 ($[M+H]^+ = 505.8$, $[M+Na]^+ = 528.3$, $[M+K]^+ = 544.2$, starred).



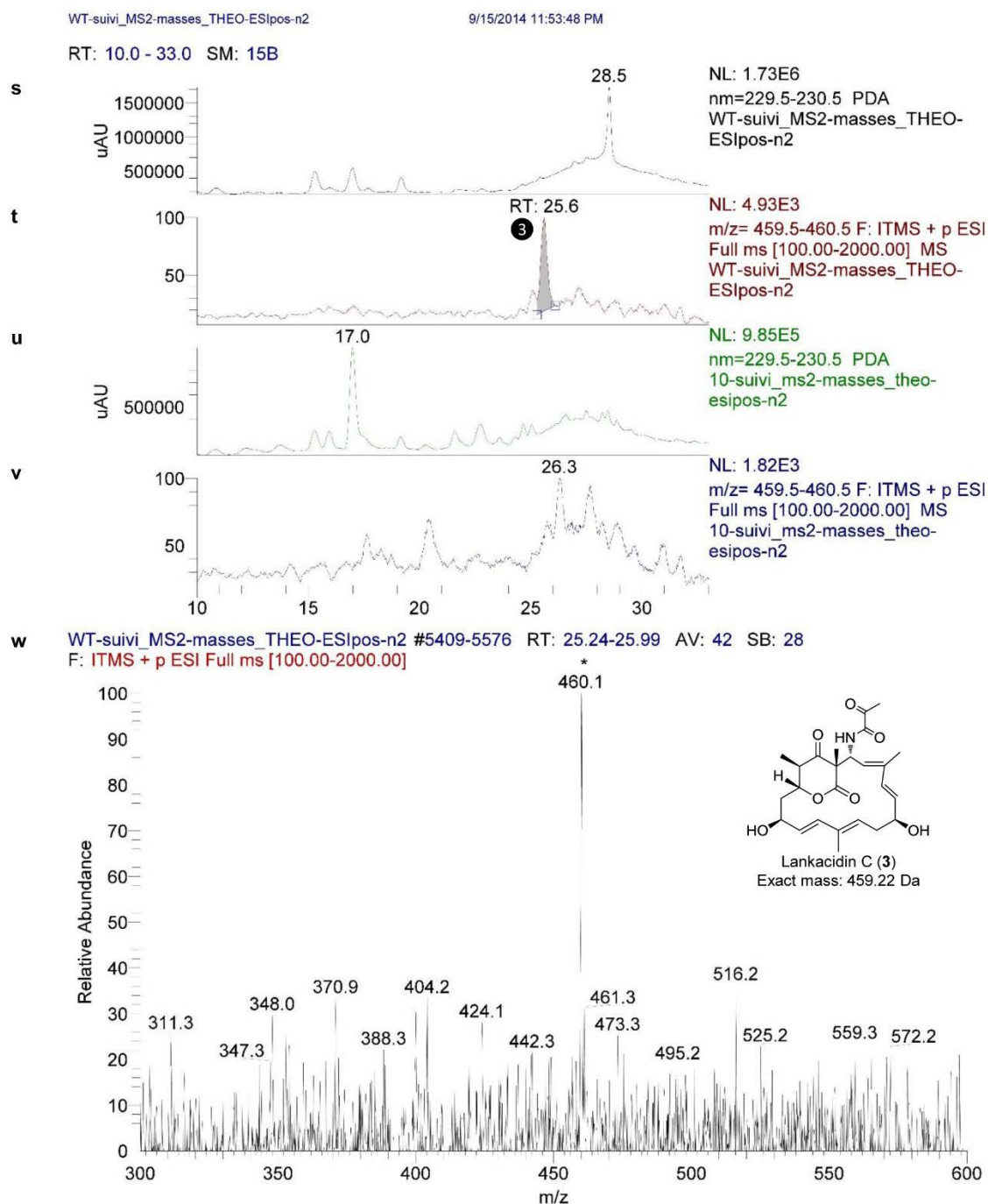
Supplementary Figure 8 | Analysis by HPLC-MS of extracts of wild type *S. rochei* vs. the $\Delta lkcE$ inactivation mutant for LC-KA05 (1) and downstream products, cont. Nonetheless, as the m/z ratio of 505.8 was lower than expected for LC-KA05 (i.e. $m/z = 506.3$ Da), we reanalyzed the extracts at high resolution (Orbitrap mode (panels (f)–(h)). (f) Trace of the absorbance at 230 nm of extracts of the $\Delta lkcE$ inactivation mutant. (g) SIM of $m/z = 506.27$ of the $\Delta lkcE$ inactivation mutant. This analysis yielded peaks of the same retention time as in (d). (h) Mass spectrum of the peak at RT = 21.9 min, confirming the expected exact mass for LC-KA05 (1) (starred – $[M+H]^+$, $[M+Na]^+$ and $[M+K]^+$)).



Supplementary Figure 8 | Analysis by HPLC-MS of extracts of wild type *S. rochei* vs. the $\Delta lkcE$ inactivation mutant for LC-KA05 (1) and downstream products, cont. (i)–(w) Analysis for metabolites formed downstream of LC-KA05. (i)–(m) Analysis for lankacidinol A (2) (exact mass = 503.25 Da). (i) Trace of the absorbance at 230 nm of extracts of the wild type. (j) SIM of $m/z = 504$ (ESI positive) of extracts of the wild type. The three peaks corresponding to lankacidinol A are greyed. (k) Trace of the absorbance at 230 nm of extracts of the $\Delta lkcE$ inactivation mutant. (l) SIM of $m/z = 504$ (ESI positive) of extracts of the $\Delta lkcE$ inactivation mutant. As expected for the absence of a functional LkcE, no 2 is detected. (m) Mass spectrum of the peak at RT = 21.8 min in (j), which is representative of the other two peaks. The parent ion is starred.



Supplementary Figure 8 | Analysis by HPLC-MS of extracts of wild type *S. rochei* vs. the $\Delta lkcE$ inactivation mutant for LC-KA05 (1) and downstream products, cont. (n) Trace of the absorbance at 230 nm of extracts of the wild type. The peak corresponding to lankacidin A is greyed. (o) SIM of $m/z = 502$ of extracts of the wild type. The peak corresponding to lankacidin A is greyed. (p) Trace of the absorbance at 230 nm of extracts of the $\Delta lkcE$ inactivation mutant. No obvious peak with RT = 28.5 min is present, as expected. (q) SIM of $m/z = 502$ of extracts of the $\Delta lkcE$ inactivation mutant. No peak with RT = 28.7 min is present, as expected. (r) Mass spectrum of the peak at 28.7 min in extracts of the wild type (o). Ions consistent with lankacidin A ($[M+H]^+ = 502.0$, $[M+Na]^+ = 524.1$) are starred.

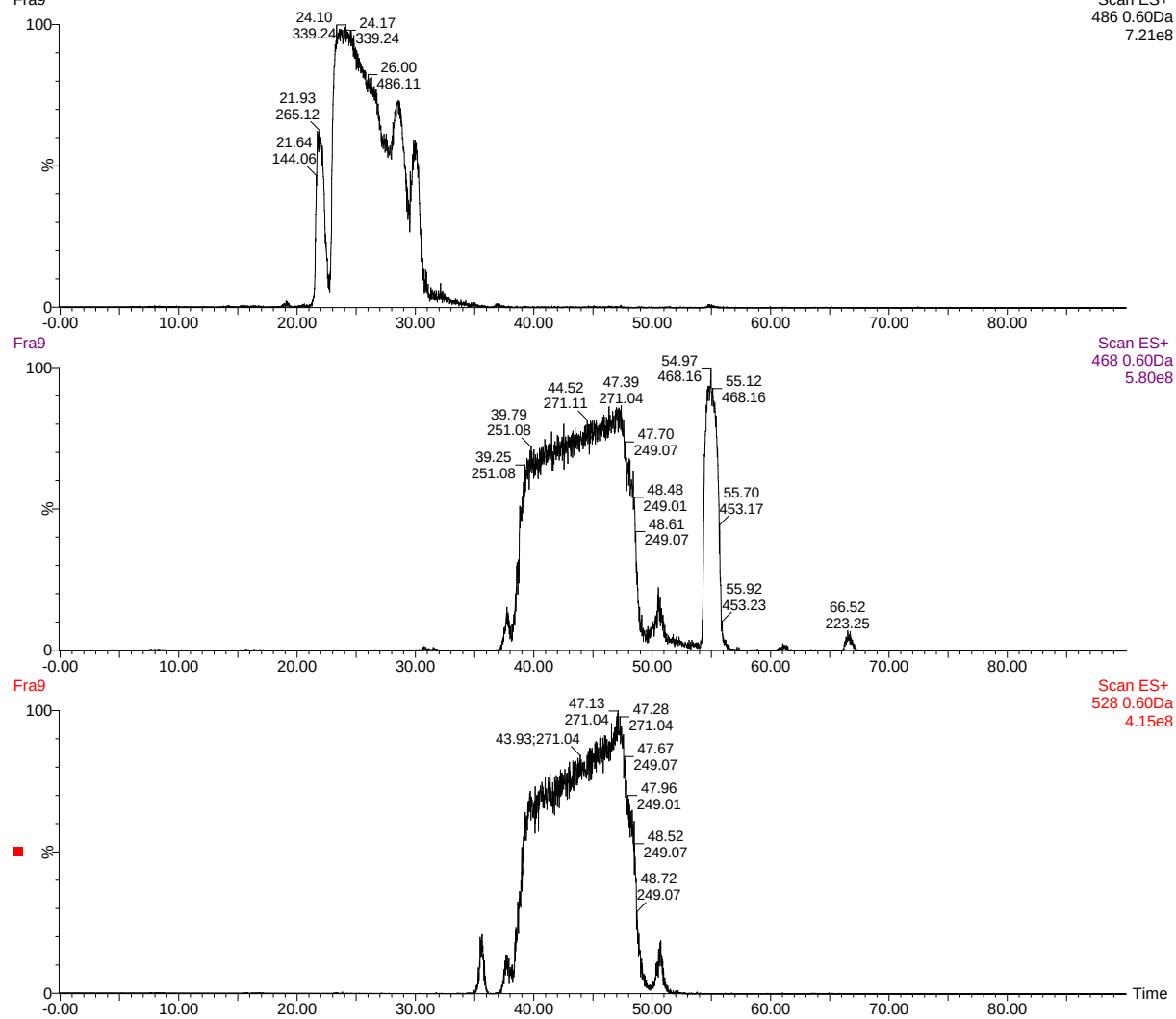


Supplementary Figure 8 | Analysis by HPLC-MS of extracts of wild type *S. rochei* vs. the $\Delta lkcE$ inactivation mutant for LC-KA05 (1) and downstream products, cont. (s)–(w) Analysis for lankacidin C (3) (exact mass = 459.22 Da). (s) Trace of the absorbance at 230 nm of extracts of the wild type. (t) SIM of $m/z = 460$ of extracts of the wild type. The peak corresponding to 3 at RT = 25.6 min is greyed. (u) Trace of the absorbance at 230 nm of extracts of the $\Delta lkcE$ inactivation mutant. (v) SIM of $m/z = 460$ of extracts of the $\Delta lkcE$ inactivation mutant. No peak with RT = 25.6 min is detected, as expected. (w) Mass spectrum of the peak at RT = 25.7 min in extracts of the wild type (t), showing the expected mass ($[M+H]^+ = 460.1$) is present (starred).

a

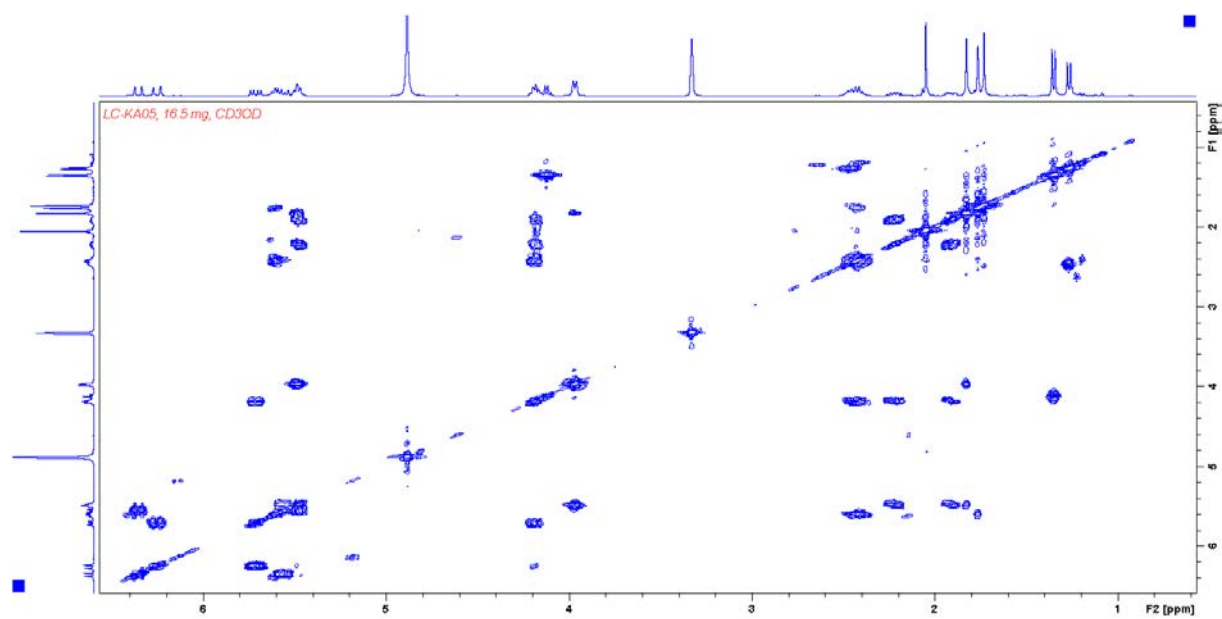
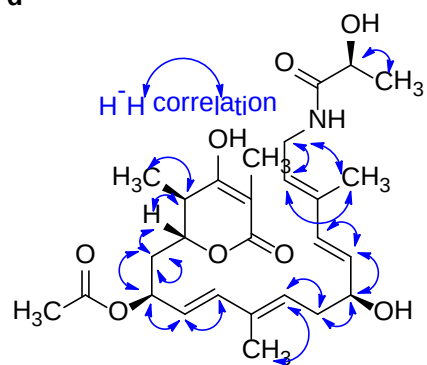
20-45 ACN 100 min

Fra9

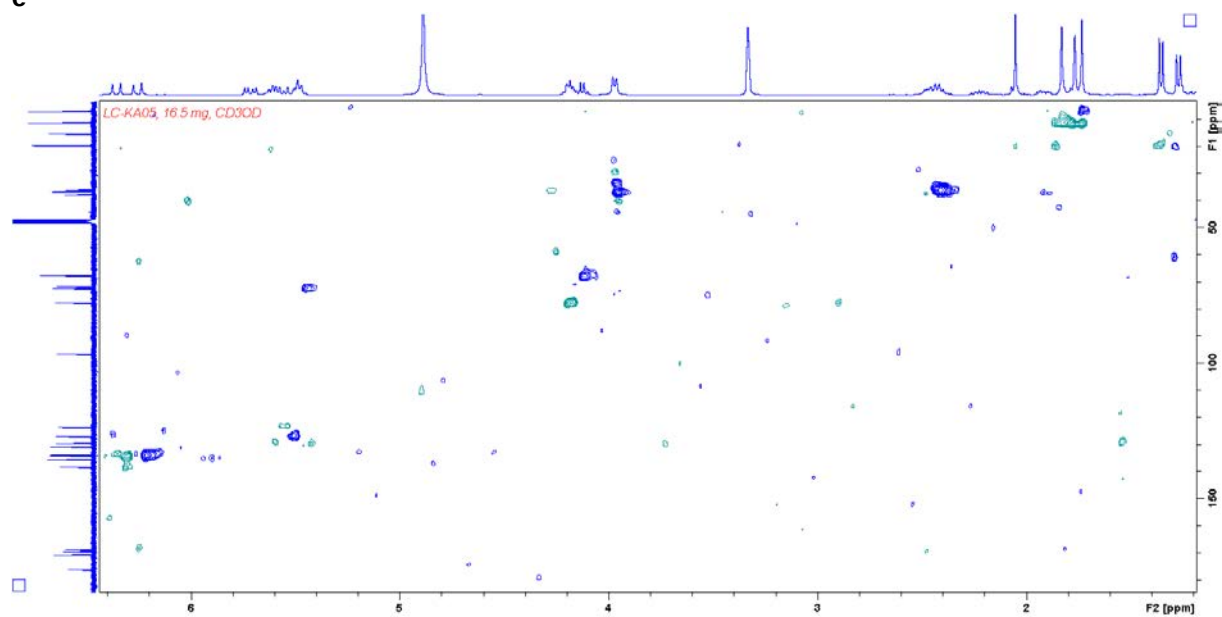


[illegible]

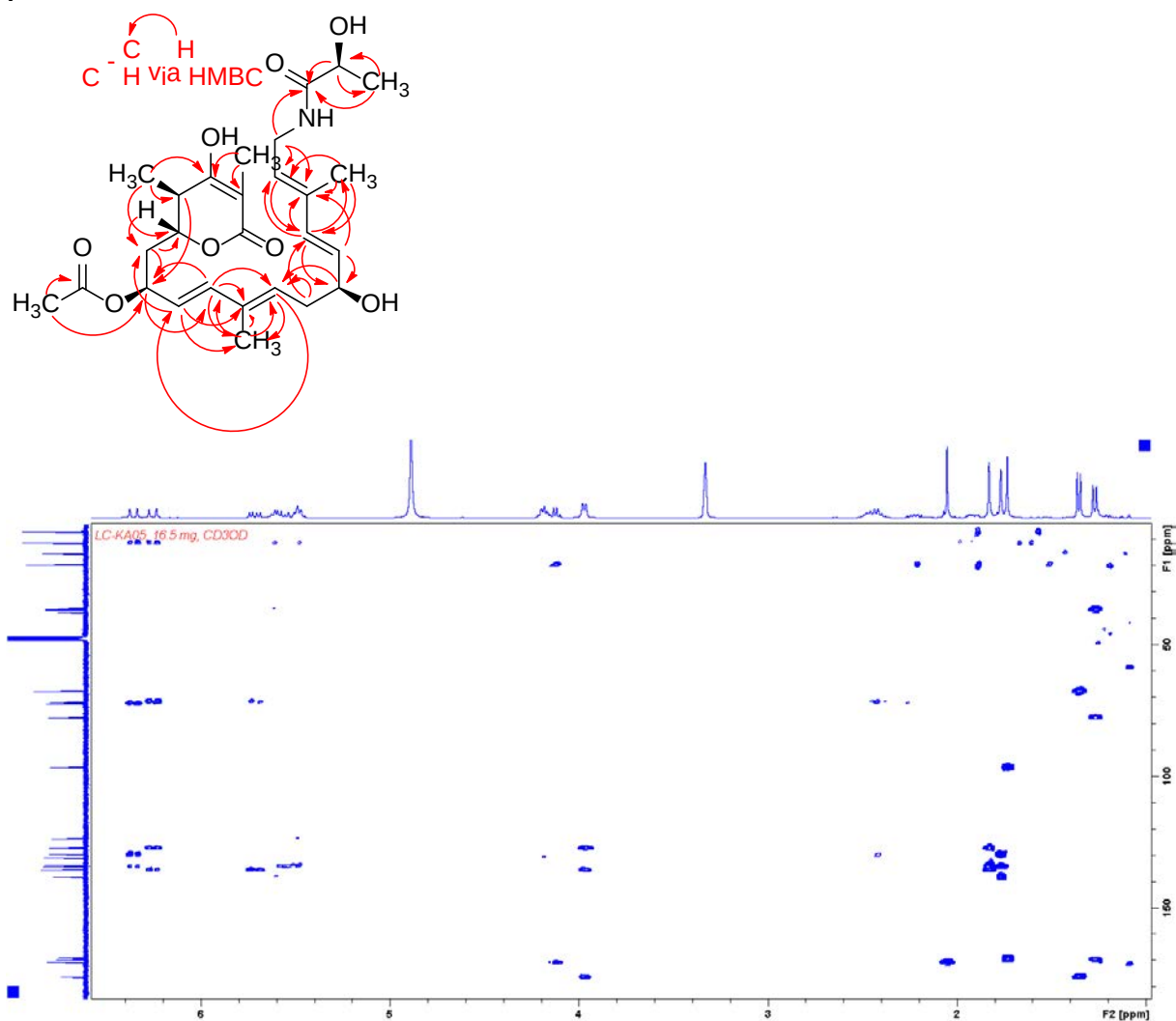
d



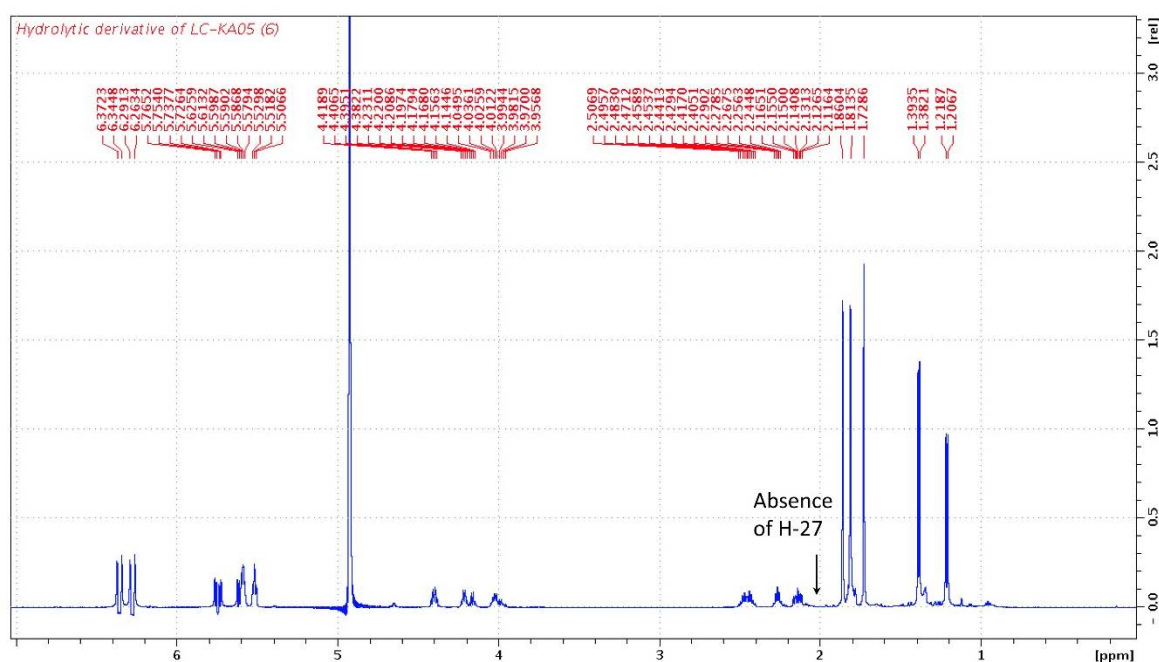
e



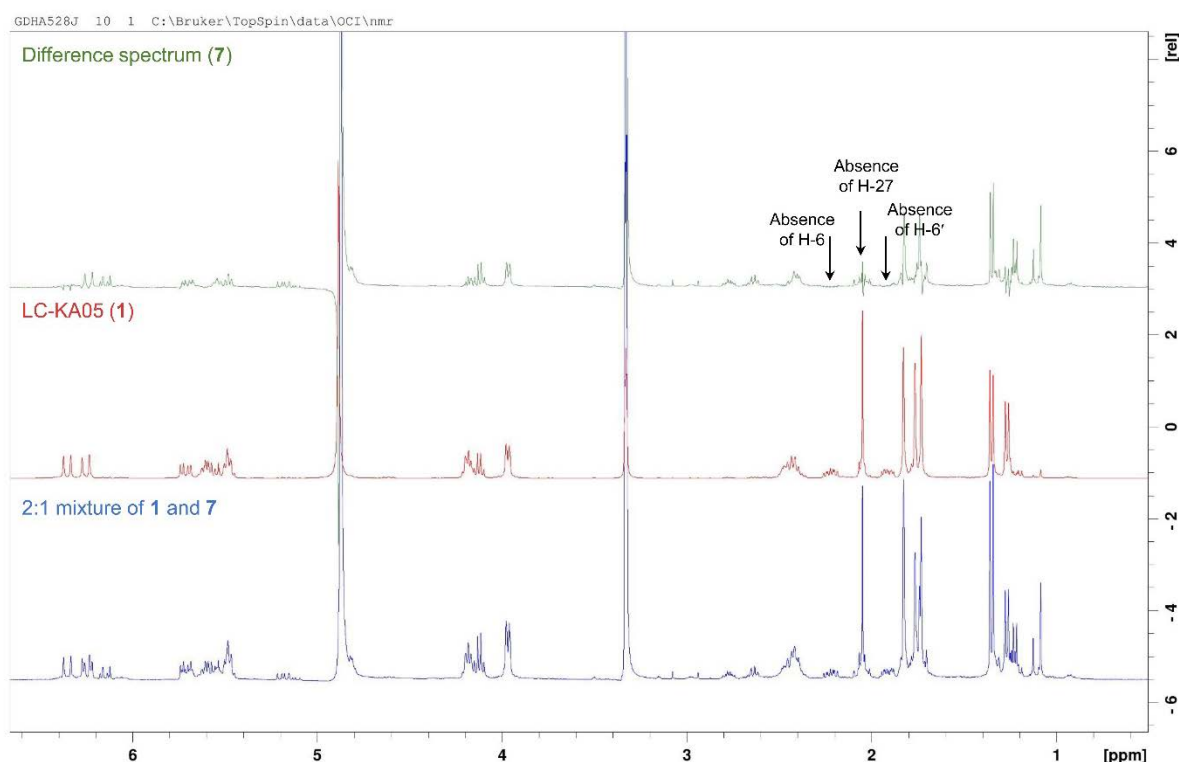
f



g

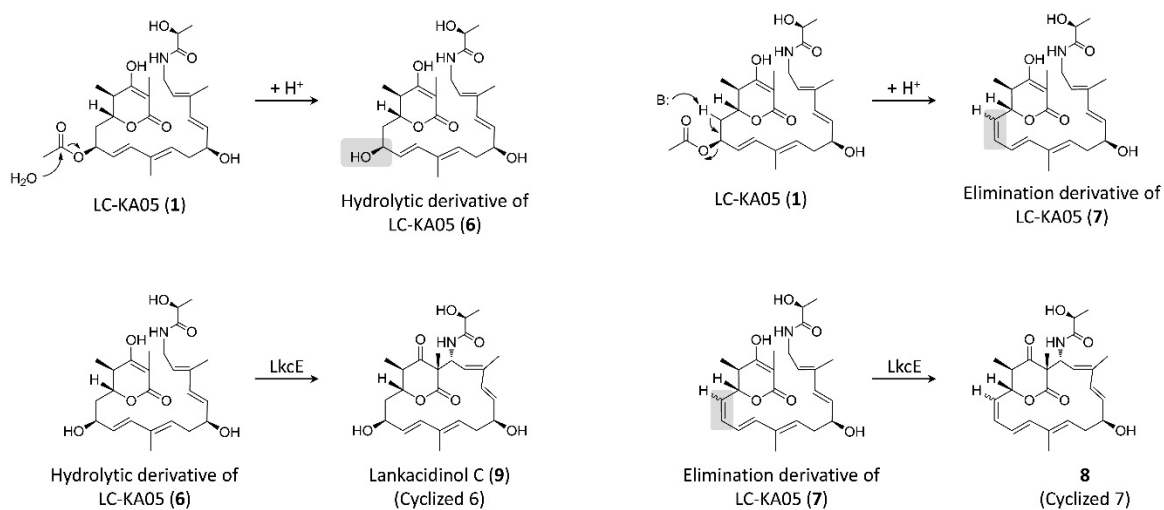


h

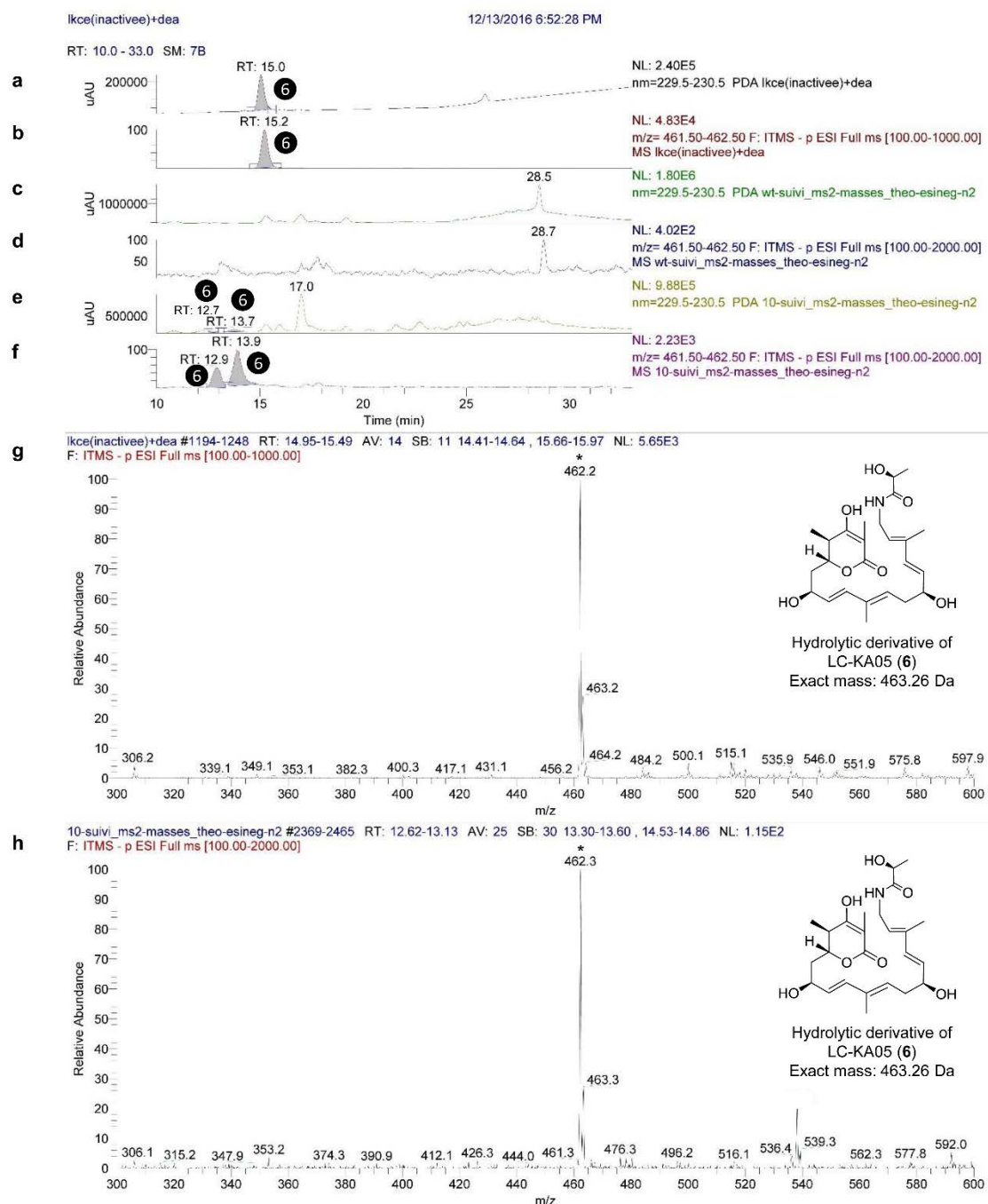


Supplementary Figure 9 | Purification and NMR analysis of LC-KA05 (1) and its elimination product (7). (a) Preparative HPLC-MS of crude extract of *S. rochei* containing LC-KA05. Isolated fractions: RT = 23–27 min (7-OH form 6; $[M+Na]^+$ m/z = 486) and 39–48 min (LC-KA05, $[M+Na]^+$ m/z = 528). The mass of 468 $[M+Na]^+$ corresponds to the elimination product 7 (loss of acetic acid from LC-KA05). This mass is visible as an ionization-induced fragment of LC-KA05, and additionally as a discrete compound eluting at 55 min. (b) 1H NMR spectrum of LC-KA05 (in CD_3OD). The * indicates the signal arising from C-19 methyl group, which being a singlet, is diagnostic for the enol form of the compound (the alternative keto form would have given rise to a doublet). In addition, the proton H-2 is

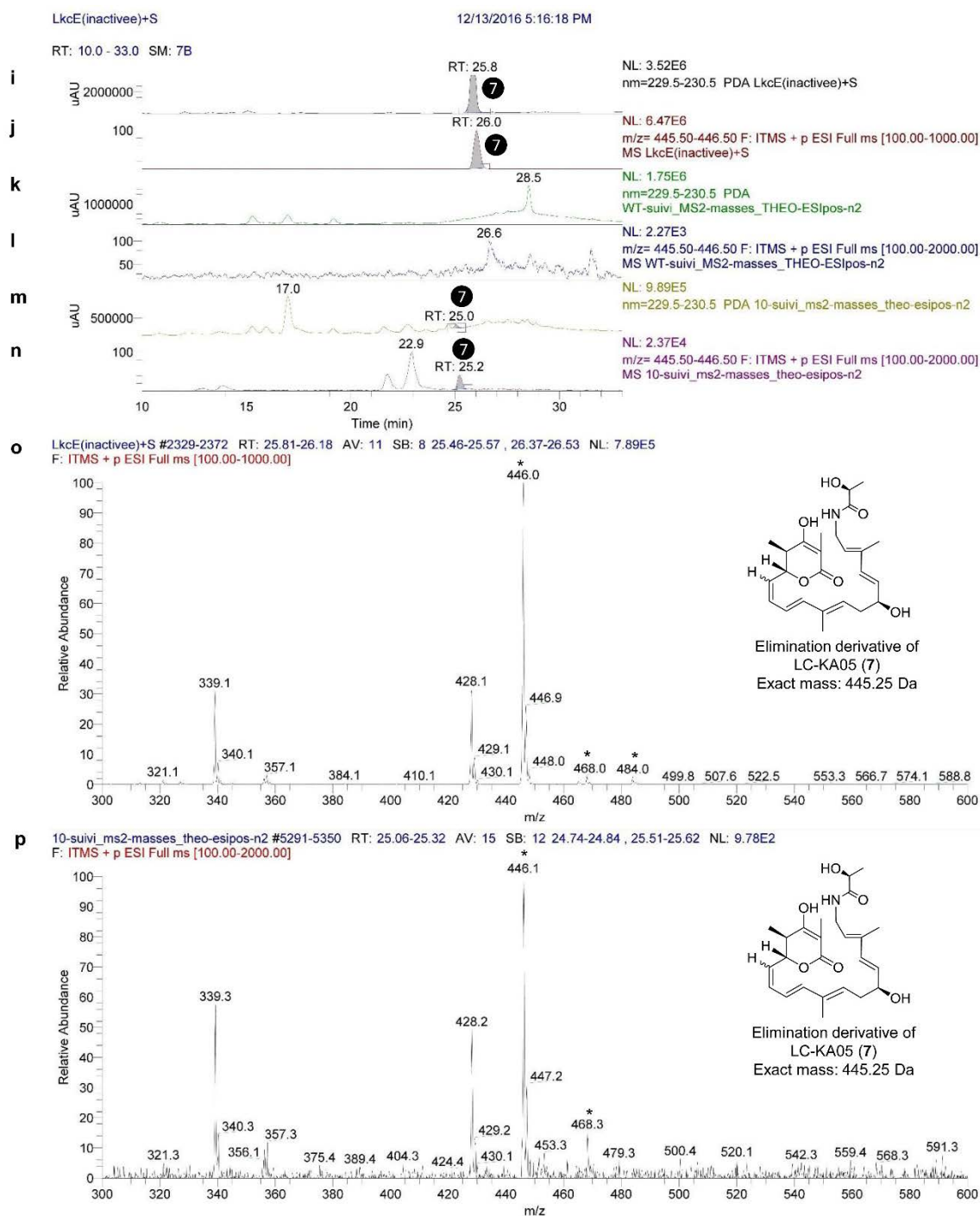
absent from the spectrum. (c) ^{13}C NMR spectrum of LC-KA05 (in CD_3OD), which is also consistent with the absence of a ketone group at C-3 and the presence of the C-2–C-3 double bond. (d) H-H COSY NMR spectrum of LC-KA05 (in CD_3OD). The structure indicates the observed correlations. (e) $[\text{}^1\text{H}, \text{}^{15}\text{N}]$ -HSQC NMR spectrum of LC-KA05 (in CD_3OD). (f) HMBC spectrum of LC-KA05 (in CD_3OD). The structure indicates the observed correlations. (g) ^1H NMR spectrum of the 7-OH form **6** (in CD_3OD). The spectrum is overall closely similar to that of LC-KA05 in (b), but notably shows the absence of a singlet at ca. 2 ppm corresponding to the three C-27 protons (H-27). (h) Comparative NMR analysis of **1** and **7**. In blue, the ^1H NMR spectrum of a 2:1 mixture of **1** and **7**. In red, the ^1H NMR spectrum of essentially pure **1** (a small amount of this material was obtained, but later decomposed during an attempt at complete purification). The difference spectrum (in green), shows that **7** lacks the signals for the protons H-6 and H-6' as expected, as well as the three C-27 protons (H-27). Additional new signals appear in the difference spectrum, but these could not be confidently assigned.



Supplementary Figure 10 | Analysis by HPLC-MS of extracts of wild type *S. rochei* and the $\Delta lkcE$ inactivation mutant for degradation products of LC-KA05, 6 and 7, and their cyclized equivalents 8 and 9. Scheme indicating how compounds 6–9 are formed.

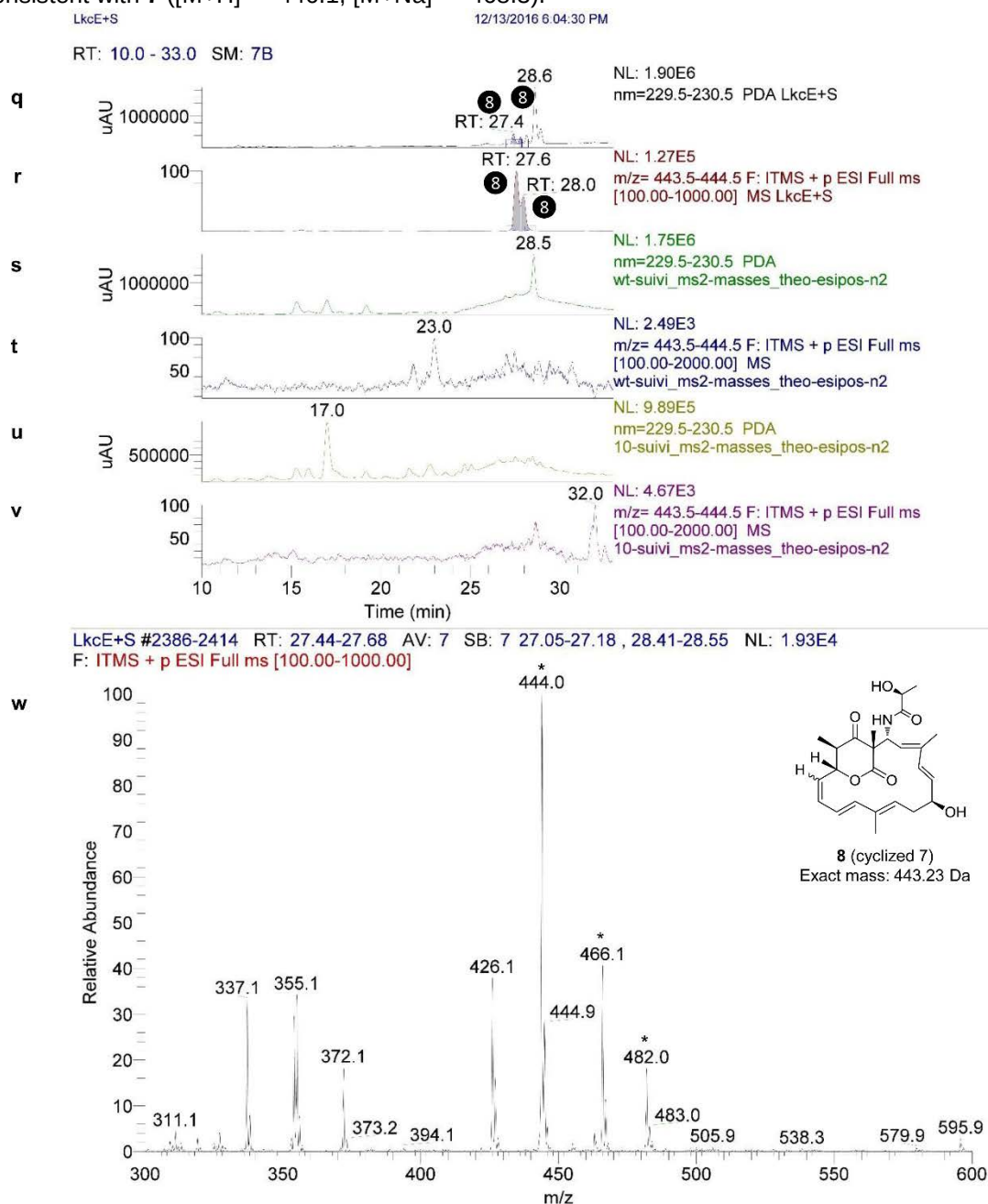


Supplementary Figure 10 | Analysis by HPLC-MS of extracts of wild type *S. rochei* and the Δ lkce inactivation mutant for degradation products of LC-KA05, 6 and 7, and their cyclized equivalents 8 and 9, cont. *m/z* values are reported to one significant figure. (a) Control for **6**, the deacetylated (7-OH) derivative of LC-KA05 (exact mass = 463.26 Da). Analysis of extracts of an in vitro assay containing inactivated Lkce and purified **6**. Trace of the absorbance at 230 nm, showing a single peak at RT = 15.0 min corresponding to **6** (greyed). (b) Analysis at *m/z* = 462 (ESI negative) of the same in vitro assay yields a single peak at RT = 15.2 min (greyed). (c) Trace of the absorbance at 230 nm of extracts of wild type *S. rochei*, showing the absence of a significant peak at RT = 15.0 min. (d) Analysis of wild type *S. rochei* at *m/z* = 462 (ESI negative) also did not reveal any peak corresponding to **6**. (e) Trace of the absorbance at 230 nm of extracts of the Δ lkce inactivation mutant. Two candidate peaks corresponding to **6** are observed (greyed). (f) Analysis of the Δ lkce inactivation mutant at *m/z* = 462 (ESI negative) reveals two candidate peaks corresponding to **6** with the same RT as in (e) (RT = 12.9 and 13.9). The identity of the peaks as **6** are supported by their mass spectra (see (h) for a representative spectrum). (g) Mass spectrum of the peak at 15.4 min in (b). The parent ion is starred. (h) Mass spectrum of the peak at RT = 13.9 min in (f) which is representative of that at RT = 12.9 min. The parent ion is starred.

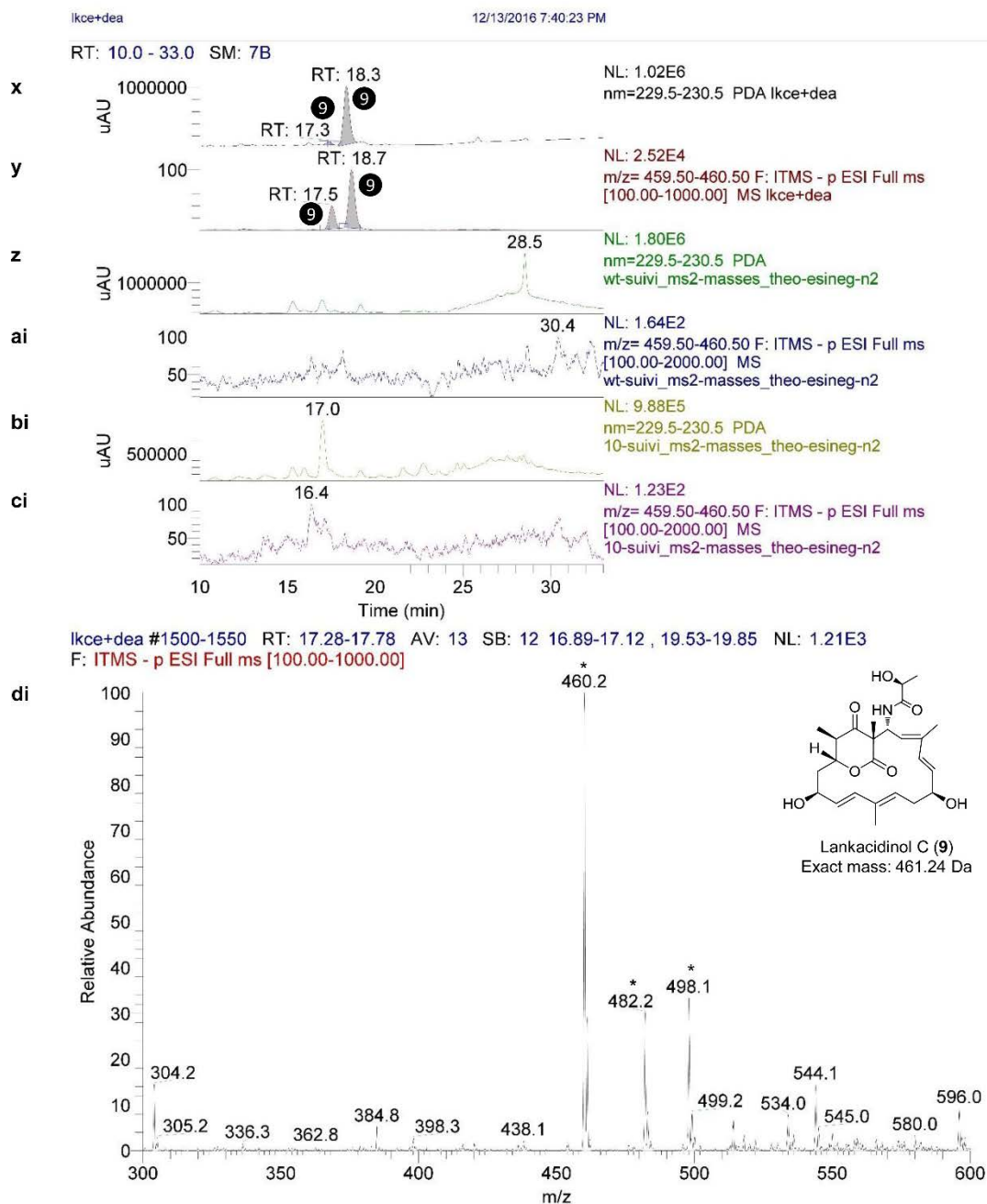


Supplementary Figure 10 | Analysis by HPLC-MS of extracts of wild type *S. rochei* and the $\Delta lkcE$ inactivation mutant for degradation products of LC-KA05, 6 and 7, and their cyclized equivalents 8 and 9, cont. (i) Control for 7, the acetate eliminated derivative of LC-KA05 (exact mass = 445.25 Da). Trace of the absorbance at 230 nm of an in vitro assay containing inactivated LkcE and the purified, 2:1 mixture of LC-KA05 1/eliminated derivative 7, showing a single peak at RT = 25.8 min corresponding to 7 (greyed). (j) Analysis of extracts of the same in vitro assay containing the 1/7 mixture. Analysis at $m/z = 446$ (ESI positive) yielded a single peak at RT = 26.0 min. (k) Trace of the absorbance at 230 nm of extracts of wild type *S. rochei*. No distinct peak corresponding to 7 is observed. (l) Analysis at $m/z = 446$ (ESI positive) of extracts of wild type *S. rochei* does not yield a clear peak corresponding to 7. (m) Trace of the absorbance at 230 nm of extracts of the $\Delta lkcE$ inactivation mutant, reveals a peak corresponding to 7 at RT = 25.0 min. (n) Analysis of the $\Delta lkcE$ inactivation mutant at $m/z = 446$ (ESI positive) yields the same peak (RT = 25.2 min). (o) Mass spectrum of the peak at RT = 26.0 min in (j) shows ions consistent with 7 ($[M+H]^+ = 446.0$, $[M+Na]^+ =$

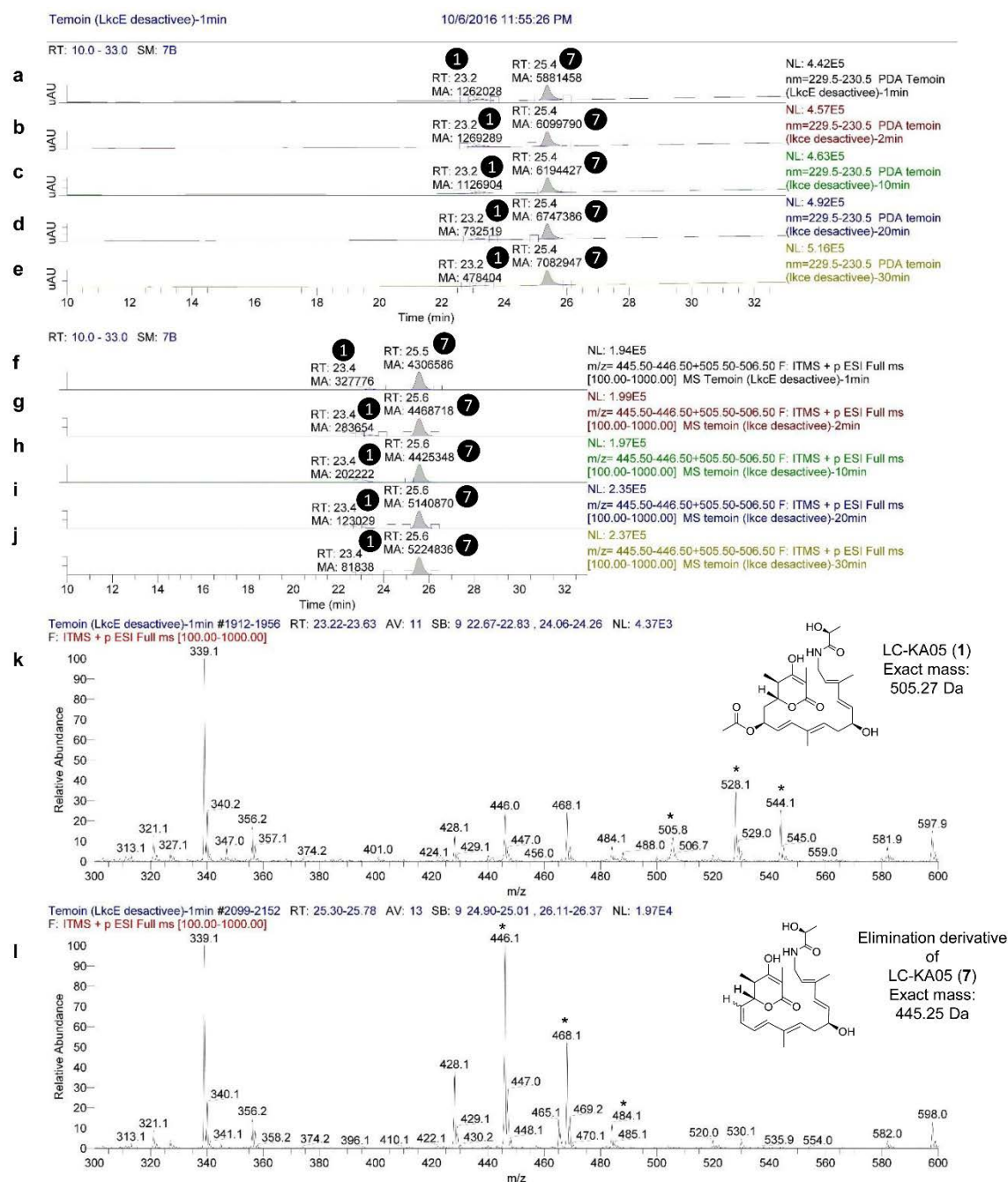
468.0, $[M+K]^+ = 484.0$, starred). (p) Mass spectrum of the peak at RT = 25.2 min in (n), shows ions consistent with **7** ($[M+H]^+ = 446.1$, $[M+Na]^+ = 468.3$).



Supplementary Figure 10 | Analysis by HPLC-MS of extracts of wild type *S. rochei* and the $\Delta lkcE$ inactivation mutant for degradation products of LC-KA05, **6 and **7**, and their cyclized equivalents **8** and **9**, cont.** (q) Control for **8** the cyclized form of the elimination derivative **7** (exact mass = 443.23 Da). Analysis of an in vitro assay containing active LkcE and a 2:1 mixture of LC-KA05 (**1**) and the eliminated derivative **7**. Trace of the absorbance at 230 nm yields two candidate peaks for the cyclized product **8** (RT = 27.4 and 28.6 min). (r) Analysis of the same in vitro assay at $m/z = 444$ (ESI positive) yields two peaks at essentially the same RT (RT = 27.6 and 28.0 min). (s) Trace of the absorbance at 230 nm of extracts of wild type *S. rochei*. No distinct peak corresponding to **8** is observed. (t) Analysis at $m/z = 444$ (ESI positive) of extracts of wild type *S. rochei* also does not yield a clear peak corresponding to **8**. (u) Trace of the absorbance at 230 nm of extracts of the *lkcE* inactivation mutant. No distinct peak corresponding to **8** is observed. (v) Analysis at $m/z = 444$ (ESI positive) of extracts of the *lkcE* inactivation mutant also does not yield a clear peak corresponding to **8**. (w) Mass spectrum of the peak at RT = 27.6 in (r), showing peaks consistent with **8** ($[M+H]^+ = 444.0$, $[M+Na]^+ = 466.1$, $[M+K]^+ = 482.0$, starred).

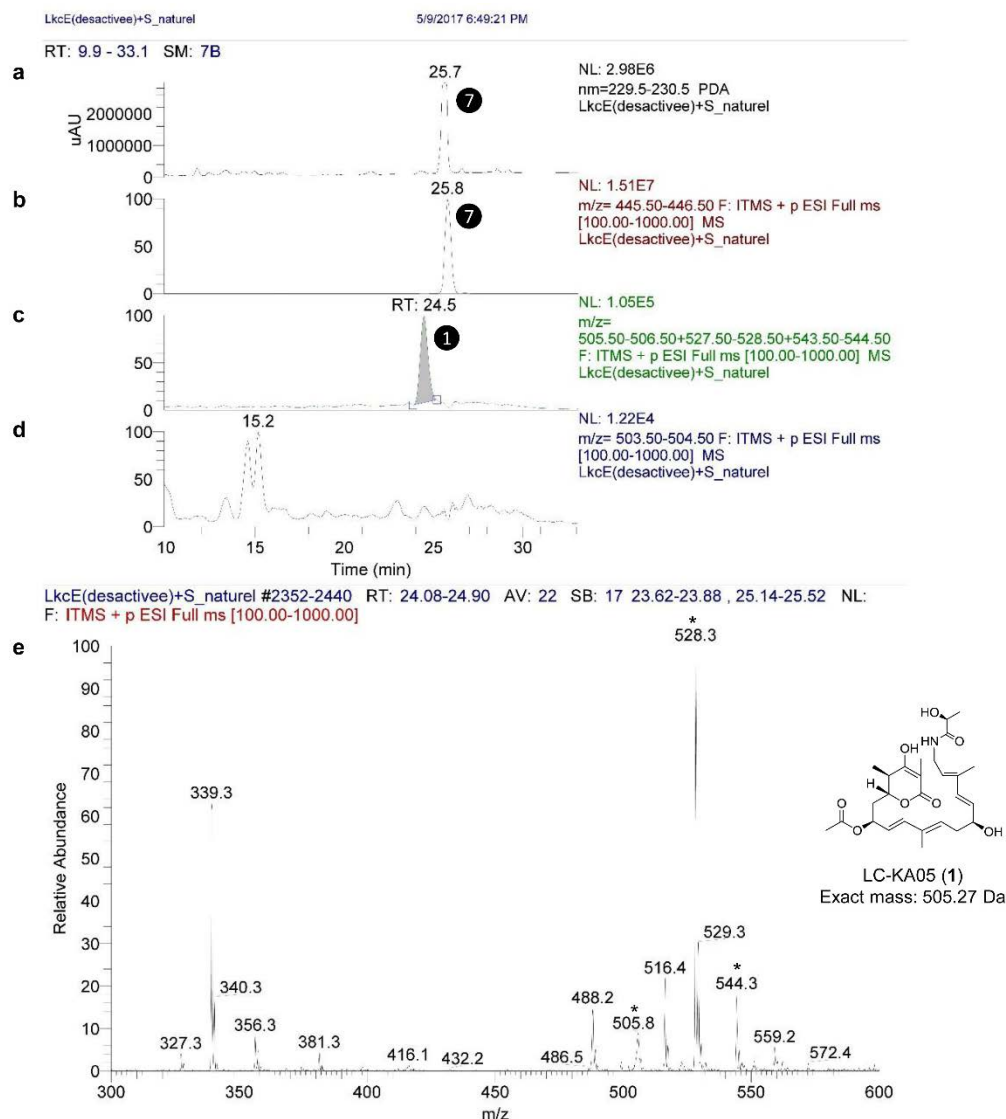
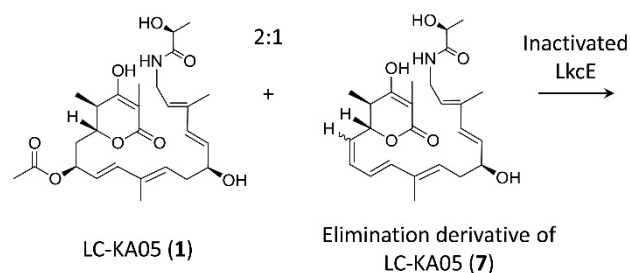


Supplementary Figure 10 | Analysis by HPLC-MS of extracts of wild type *S. rochei* and the Δ lkce inactivation mutant for degradation products of LC-KA05, 6 and 7, and their cyclized equivalents 8 and 9, cont. (x) Control for lankacidinol C (9), the cyclized form of 6 (exact mass = 461.24 Da). Analysis of extracts of an in vitro assay containing active LkcE and purified 6. Trace of the absorbance at 230 nm, showing two peaks at RT = 17.3 and 18.3 min corresponding to 9 (greyed). (y) Analysis at m/z = 460 (ESI negative) of the same in vitro assays also yields two peaks (RT = 17.5 and 18.7 min). (z) Trace of the absorbance at 230 nm of extracts of wild type *S. rochei*, showing the absence of peaks corresponding to 9. (ai) Analysis of extracts of wild type *S. rochei* at m/z = 460 (ESI negative) also did not reveal any peak corresponding to 9. (bi) Trace of the absorbance at 230 nm of extracts of Δ lkce inactivation mutant, showing the absence of peaks corresponding to 9, as expected. (ci) Analysis of extracts of Δ lkce inactivation mutant at m/z = 460 (ESI negative) also did not reveal any peak corresponding to 9. (di) Mass spectrum of the peak at RT = 17.5 in (y), showing peaks consistent with 9 ($[M-H]^-$ = 460.2, $[M-2H+Na]^-$ = 482.2, $[M-2H+K]^-$ = 498.1, starred).



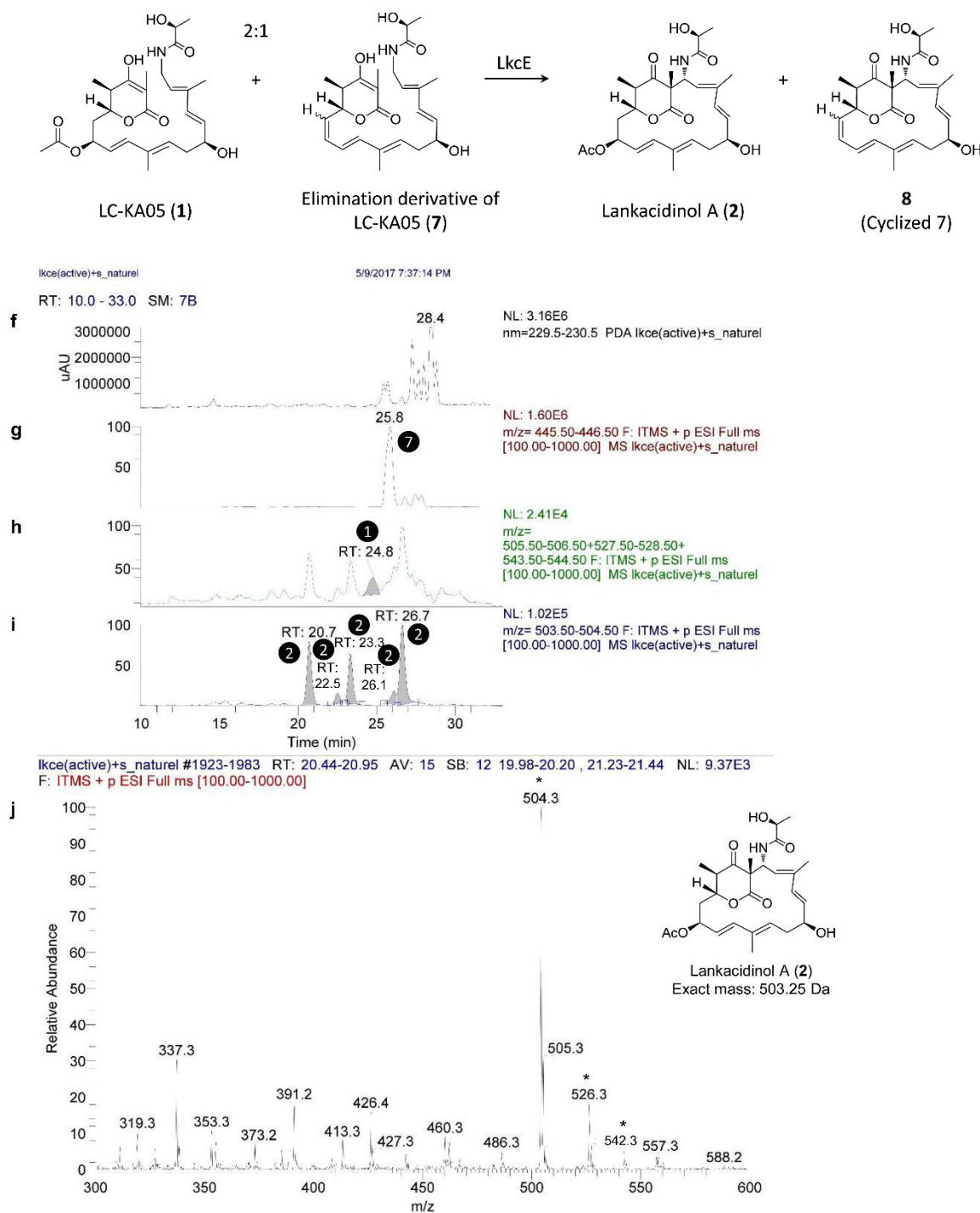
Supplementary Figure 11 | Analysis by HPLC-MS of the stability of LC-KA05 under standard LkcE in vitro assay conditions. Incubations were carried out in the presence of inactivated LkcE and the 2:1 mixture of LC-KA05 (1)/eliminated derivative 7, and the substrates extracted and analyzed at various time points. (a)–(e) Trace of the absorbance at 230 nm showing peaks at RT = 23.2 and 25.4 min, corresponding to 1 and 7, respectively. Integration of the peak areas shows that even in the absence of reaction, 1 decreased over time (1–30 min), while 7 increased correspondingly. (f)–(j) Analysis of the same time points at $m/z = 446$ (ESI positive) (for 7) and $m/z = 506$ (for 1). This analysis produced peaks with essentially the same retention times as in (a)–(e), and which showed the same behavior in terms of change in area with time. (k) Mass spectrum of the peak at RT = 23.4 min in (f) corresponding to 1. Although the retention time differs from that in **Supplementary Fig. 8 (d)**, the mass spectrum contains the same major peaks (including $[M+H]^+ = 505.8$, $[M+Na]^+ = 528.1$, $[M+K]^+ = 544.1$, starred). The difference in RT may be explained by the fact that 1 was analyzed within an extract in **Supplementary Fig. 8 (d)**, while the analysis here was carried out on purified compound following in vitro assay (and see **Table S4**). (l) Mass spectrum of the peak at RT = 25.5 min in (f) corresponding to 7 (for comparison with **Supplementary Fig. 10 (o)**).

Control reaction with 1/7 mixture



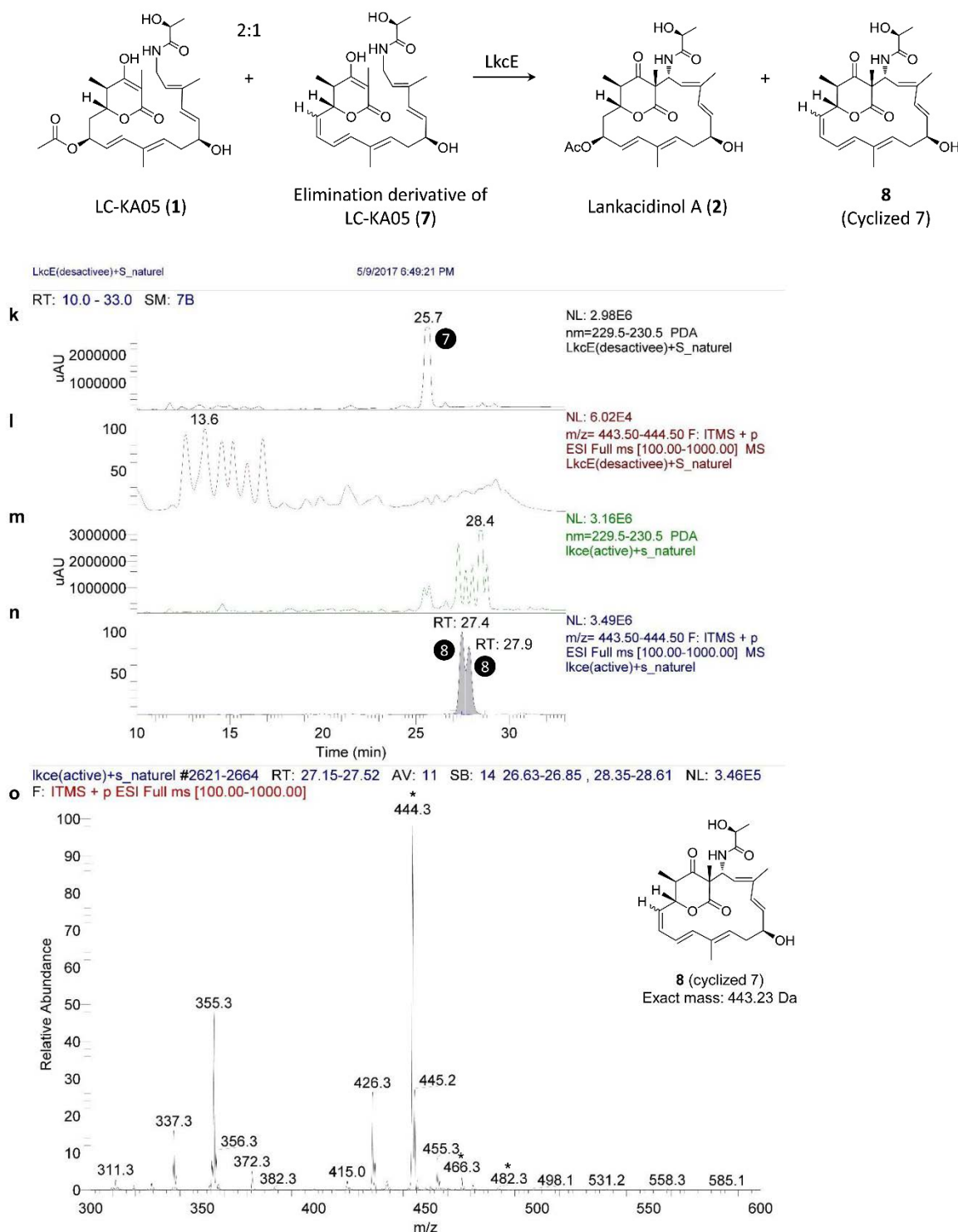
Supplementary Figure 12 | Demonstration by HPLC-MS that LkcE is active in vitro. A reaction scheme is provided for each experiment. **(a)** Trace of the absorbance at 230 nm of a control reaction (24 h incubation) with inactivated LkcE and the 2:1 native substrate LC-KA05 (**1**) (exact mass = 505.27 Da)/eliminated derivative **7** (exact mass = 445.25 Da) mixture, showing a peak (RT = 25.7 min) corresponding to **7**. **(b)** Analysis of the control reaction at $m/z = 446$ (ESI positive), gives a peak for **7** at essentially the same retention time as in **(a)** (RT = 25.8 min). **(c)** Analysis of the control reaction at $m/z = 506$ (ESI positive) for **1**, yields a peak at RT = 24.5 min. **(d)** Analysis of the control reaction for $m/z = 504$ corresponding to the expected mass of lankacidinol A (**2**) (exact mass = 503.25 Da), does not yield a peak with the appropriate retention time (20–27 min, **Supplementary Table 4**). **(e)** Mass spectrum of the peak at 24.5 min in **(c)**, showing peaks characteristic of **1** ($[M+H]^+ = 505.8$, $[M+Na]^+ = 528.3$, $[M+K]^+ = 544.3$, starred) (and see **Supplementary Table 4**).

Assay of active LkcE with 1/7 mixture



Supplementary Figure 12 | Demonstration by HPLC-MS that LkcE is active in vitro, cont. (f) Trace of the absorbance at 230 nm of a reaction with active LkcE and the 2:1 1/ 7 mixture, after 24 h incubation. **(g)** Analysis of the LkcE reaction at $m/z = 446$ (ESI positive) yields a peak corresponding to unreacted 7. **(h)** Analysis of the LkcE reaction at $m/z = 506$ (ESI positive) for 1, yields a peak at RT = 24.8 min corresponding to unreacted 1. **(i)** Analysis of the LkcE reaction at $m/z = 504$ for 2, yields 5 peaks (RT = 20.7, 22.5, 23.3, 26.1 and 26.7 min) which are notably not present in the control reaction **(d)**. **(j)** Mass spectrum of the peak in (i) at RT = 20.7 min, which is representative of the others. This shows peaks characteristic of 2 ($[M+H]^+ = 504.3$, $[M+Na]^+ = 526.3$, $[M+K]^+ = 542.3$, starred).

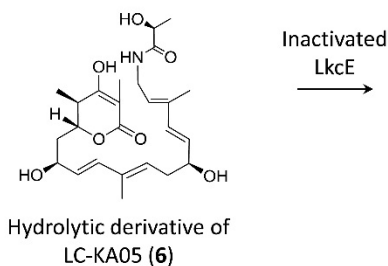
Assay of active LkcE with 1/7 mixture



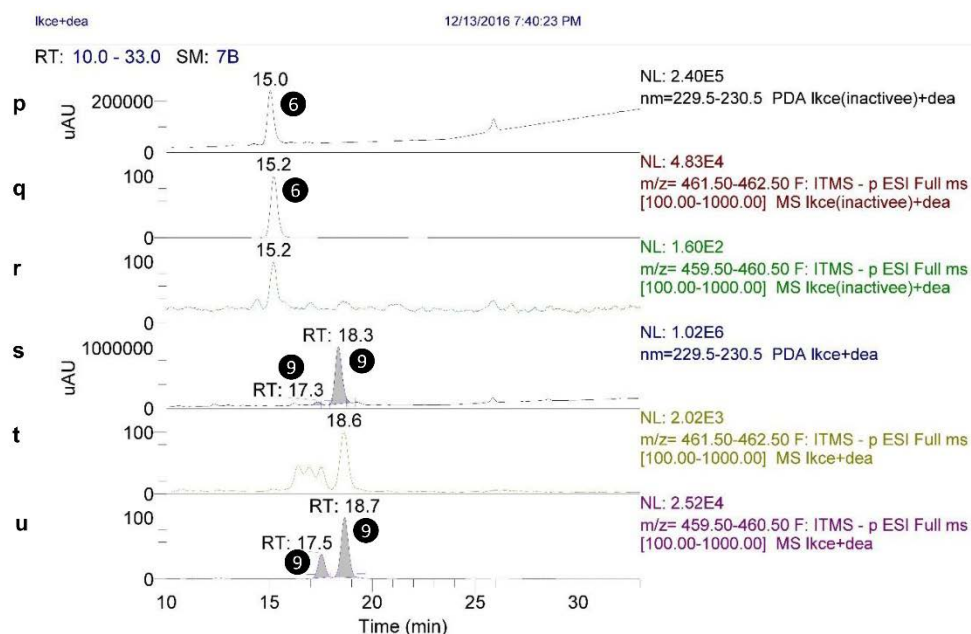
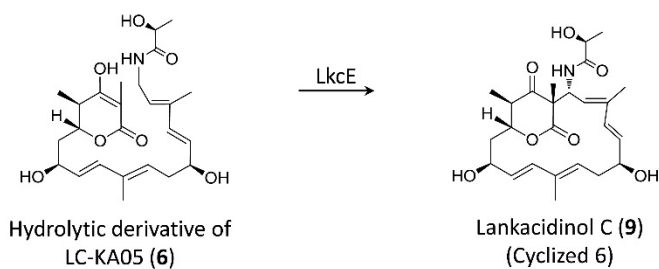
Supplementary Figure 12 | Demonstration by HPLC-MS that LkcE is active in vitro, cont. (k) Trace of the absorbance at 230 nm of a control reaction (24 h incubation) with inactivated LkcE and the 2:1 1/7 mixture, showing a peak (RT = 25.7 min) corresponding to 7. **(l)** Analysis of the control reaction at $m/z = 444$ (ESI positive) corresponding to 8 (cyclized 7), does not yield a candidate peak. **(m)** Trace of the absorbance at 230 nm of a reaction with active LkcE and the 2:1 1/7 mixture, after 24 h incubation, yields multiple candidate peaks for 8 in the 27–28 min range, which are not present in (l). **(n)** Analysis of the LkcE reaction at $m/z = 444$ for 8, yields 2 peaks (RT = 27.4 and 27.9 min) which are notably not present in the control reaction (d). **(o)** Mass spectrum of the peak at 27.4 min which is

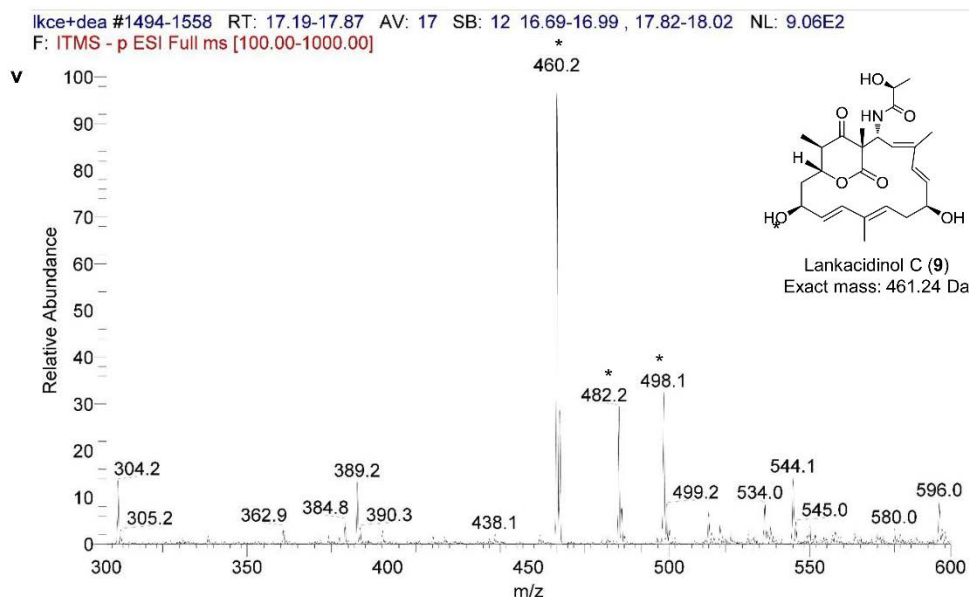
representative of the two, showing ions consistent with **8** ($[M+H]^+ = 444.3$, $[M+Na]^+ = 466.3$, $[M+K]^+ = 482.3$, starred).

Control reaction (panels p–r)



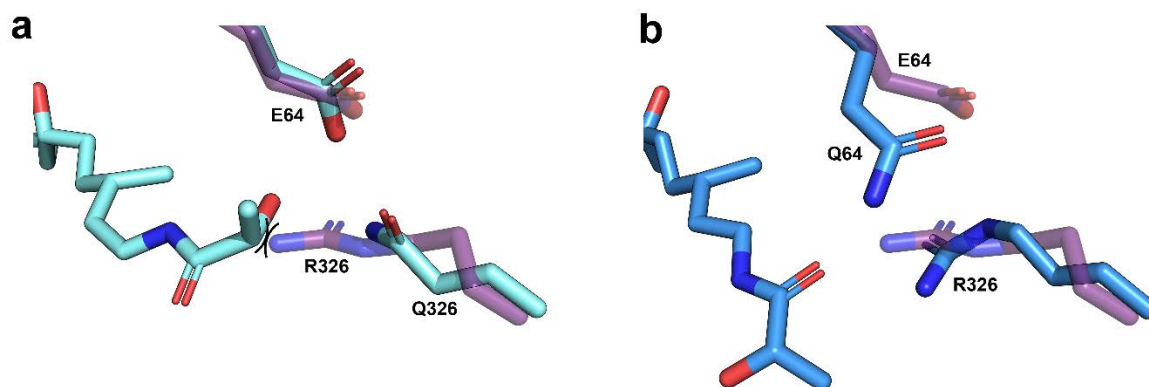
Assay of active LkcE with **6** (panels s–v)





Supplementary Figure 12 | Demonstration by HPLC-MS that LkcE is active in vitro, cont. (p)

Trace of the absorbance at 230 nm of a control reaction (24 h incubation) with inactivated LkcE and purified **6**, the deacetylated (7-OH) derivative of LC-KA05 (exact mass = 463.26 Da), showing a peak corresponding to unreacted **6** (RT = 15.0 min). (q) Analysis of the control reaction at $m/z = 462$ (ESI negative) for unreacted **6**, gives a peak at the same RT. (r) Analysis of the control reaction at $m/z = 460$ for lankacidinol C (**9**) (cyclized **6**) does not yield a candidate peak. (s) Trace of the absorbance at 230 nm of a reaction with active LkcE and purified **6**, yields two peaks not present in the control reaction (p) (RT = 17.3 and 18.4 min). (t) Analysis of the LkcE reaction at $m/z = 462$ (ESI negative) for unreacted **6** does not yield a peak at RT = 15 min, as observed in the control. (u) Analysis of the LkcE reaction at $m/z = 460$ for **9**, yields the two peaks not present in the control (RT = 17.5 and 18.7 min). (v) Mass spectrum (ESI negative) of the peak at RT = 17.5 in (u) yields peaks consistent with **9** ($[M-H]^- = 460.2$, $[M-2H+Na]^- = 482.2$, $[M-2H+K]^- = 498.1$, starred).



Supplementary Figure 13 | Analysis of LC-KA05 binding into the LkcE active site. (a) Superposition of the R326Q/LC-KA05 (**1**) structure (cyan) and that of the *holo* wild type (mauve) shows that if **1** bound into the wild type in the same orientation as in the R326Q/LC-KA05 complex, its C-24 hydroxyl would sterically clash with R326. (b) Superposition of the E64Q/LC-KA05 structure (blue) and that of the *holo* wild type (mauve). In the E64Q structure, the substrate adopts an orientation such that there is no longer a steric clash with R326.

Supplementary Tables

Supplementary Table 1 Oligonucleotide primers used in this study. Mutated sites are in bold.	
Name	Sequence, 5'-3' direction
<i>Primers used for lkcE deletion plasmid</i>	
LkcE deletion_F	GCAGGCCGAGCCCAGGAAGTGATGACCGTGGTAGAAGCAAT TCCGGGGATCCGTCGACC
LkcE deletion_R	GACAGGCGGGCGGGTGCTGGCTCACACGCCCATGAATTCTG TAGGCTGGAGCTGCTTC
<i>Primers used to verify lkcE deletion</i>	
PCR conf_F	TGCGGTGGTGGCAGACCGTG
PCR conf_R	GCGGCGAGGTCGGTGTATCG
PCR conf_I	ACTCGCCTGCAAGCCCGGTC
<i>Primers used for site-directed mutagenesis in the lkcE gene</i>	
E64A_F	GTCGACATGGGCGTG GC AACTTCAACGAGAAG
E64A_R	CTTCTCGTTGAAGTGT GC CACGCCCATGTCGAC
E64Q_F	GTCGACATGGGCGTG CA AACTTCAACGAGAAG
E64Q_R	CTTCTCGTTGAAGTGT TC CACGCCCATGTCGAC
Y182F_F	GACTACAACCTCATGT TC GTGCGGCTGTCGTT
Y182F_R	ACGACAGCGCGAC GA ACATGAGGTTGTAGTC
R326L_F	ATGGGCAGCCTGACC CT CAACTGCAACGTGCTC
R326L_R	GAGCACGTTGCAGTT GAG GGTCAGGCTGCCCCAT
R326Q_F	ATGGGCAGCCTGACC CA AACTGCAACGTGCTC
R326Q_R	GAGCACGTTGCAGTT CT GGGTCAGGCTGCCCCAT

Supplementary Table 2 List and description of vectors used in this study.			
Name	Description	Application	Reference
pBG102	Derivative of pET27	Used to generate N-terminal translational fusions with His ₆ and SUMO tags	Center for Structural Biology, Vanderbilt University
pBG102-LkcE	pBG102 containing the gene encoding LkcE	Production of recombinant LkcE	This study
pBG102-LkcE_E64A	Derived from pBG102-LkcE; LkcE has Glu64Ala mutation	Evaluation of the role of residue Glu64	This study
pBG102-LkcE_E64Q	Derived from pBG102-LkcE; LkcE has Glu64Gln mutation	Evaluation of the role of residue Glu64	This study
pBG102-LkcE_Y182F	Derived from pBG102-LkcE; LkcE has Tyr182Phe mutation	Evaluation of the role of residue Tyr182	This study
pBG102-LkcE_R326L	Derived from pBG102-LkcE; LkcE has Arg326Leu mutation	Evaluation of the role of residue Arg326	This study
pBG102-LkcE_R326Q	Derived from pBG102-LkcE; LkcE has Arg326Gln mutation	Evaluation of the role of residue Arg326	This study
pIJ773	Template plasmid containing the apramycin resistance gene <i>aac(3)IV</i> and the <i>oriT</i> of plasmid RP4	Used to generate the <i>lkcE</i> inactivation construct	1
Cosmid Lc2B12	Derivative of SuperCos1 containing a portion of the lankacidin gene cluster	Used to inactivate <i>lkcE</i> in <i>S. rochei</i>	2
pIJ790	Temperature sensitive λ Red recombination helper plasmid	λ RED recombination plasmid	1

Supplementary Table 3 | Data collection and refinement statistics on LkcE crystal structures.

	SeLkcE	LkcE _{WT}	R326Q	E64Q
Data collection				
Space group	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	125.30, 125.30, 154.53	125.50, 125.50; 156.69	125.55, 125.55; 156.66	124.83, 124.83, 157.39
α , β , γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	48.66-3.15* (3.2-3.15)	48.98-2.80* (3.00-2.80)**	48.99-2.50* (2.60-2.50)**	48.90-3.03* (3.11-3.03)**
<i>R</i> _{merge}	0.131 (0.802)	0.05 (0.56)	0.05 (0.64)	0.11 (0.76)
<i>I</i> / σ <i>I</i>	13.62 (2.58)	22.3 (3.8)	20.24 (2.42)	30.69 (5.67)
CC(1/2)	1.00 (0.78)	0.99 (0.89)	0.99 (0.75)	1.00 (0.95)
Completeness (%)	99.9 (99.3)	99.9 (100.0)	99.6 (99.4)	99.8 (97.2)
Redundancy	7.6 (7.0)	7.4 (7.5)	4.4 (4.4)	29.7 (27.9)
Refinement				
Resolution (Å)	48.66-3.15	48.98-2.80	48.99-2.50	48.90-3.03
No. reflections	21897	31467	41613	23588
<i>R</i> _{work} / <i>R</i> _{free}	0.198 / 0.240	0.237 / 0.256	0.206 / 0.244	0.191 / 0.247
No. atoms				
Protein	6807	6835	6888	6907
Ligand/ion	112	195	129	79
Water	216	181	230	136
<i>B</i> -factors				
Protein [§]	70.52	87.81	61.75	70.26
Ligand/ion	67.96	77.40	58.16	77.01
Water	56.01	107.13	73.00	52.88
R.m.s. deviations				
Bond lengths (Å)	0.010	0.008	0.015	0.012
Bond angles (°)	1.13	1.10	1.79	1.61

*A single crystal was used for each data set; **Values in parentheses are for highest-resolution shell.

[§]The corresponding *B*-factors for the protein from the Wilson plots are: 75.3, 82.2, 60.8 and 72.9, respectively. There is good agreement between these two sets of *B*-factors, indicative of significant dynamic disorder at the atomic level in the structures.

Supplementary Table 4 | Analytical data obtained by HPLC-MS on compounds investigated in this study.

Compound	Calc. <i>m/z</i> (Da)	Number of peaks	Retention time(s) (min) ^a	Characteristic MS peaks (ionization mode)
1 LC-KA05	505.27	2	Culture extracts: 21.8, 22.9 (Fig. S7) In vitro: 23.4 (Fig. S10) and 24.5 (Fig. S11)	544 ([M+K] ⁺), 528 ([M+Na] ⁺), 506 ([M+H] ⁺), 446, 428, 339 (ESI positive)
2 Lankacidinol A	503.25	3–5	Culture extracts: 21.8, 22.6, 23.0 (Fig. S7) In vitro: 20.7, 22.5, 23.3, 26.1, 26.7 (Fig. S11)	542 ([M+K] ⁺), 526 ([M+Na] ⁺), 504 ([M+H] ⁺), 486, 426, 354, 337 (ESI positive)
Lankacidin A	501.24	1	Culture extracts: 28.7 (Fig. S7)	524 ([M+Na] ⁺), 502 ([M+H] ⁺), 484, 424, 354, 337 (ESI positive)
3 Lankacidin C	459.22	1	Culture extracts: 25.6 (Fig. S7)	460 ([M+H] ⁺), 400 (ESI positive)
6 Hydrolytic (deacetylated) derivative of LC-KA05	463.26	1 or 2	Culture extracts: 12.9, 13.9 In vitro: 15.2 (Both in Fig. S9)	462 ([M–H] [–]) (ESI negative) 502 ([M+K] ⁺), 486 ([M+Na] ⁺), 464 (M+H) ⁺ , 446, 428, 339 (ESI positive)
7 Eliminated derivative of LC-KA05	445.25	1	Culture extracts: 25.2 In vitro: 26.0 (Both in Fig. S9)	484 ([M+K] ⁺), 468 ([M+Na] ⁺), 446 ([M+H] ⁺), 428, 339 (ESI positive)
8 Cyclized 7	443.23	2	In vitro: 27.6, 28.0 (Fig. S9)	482 ([M+K] ⁺), 466 ([M+Na] ⁺), 444 ([M+H] ⁺), 426, 355, 337 (ESI positive)
9 Lankacidinol C (cyclized 6)	461.24	2	In vitro: 17.5, 18.7 (Fig. S9)	498 ([M–2H+K] [–]), 482 ([M–2H+Na] [–]), 460 ([M–H] [–]), 384 (ESI negative)

^aAnalysis of several compounds (**1**, **2**, **6**, **8** and **9**) gave multiple peaks, which we attribute to the presence of isomerization in the two conjugated systems (C8–C11 and C14–C17). In addition, for compounds **1**, **2**, **6** and **7**, we noted differences both in retention times and the number of peaks when analysis was carried out on culture extracts vs. substrates/products of in vitro assays. We can cite several potential reasons for these variations. Although all of the analyses were performed under the same conditions (column, elution profile, etc.), the two-year period over which they were carried out may have seen some degradation of the column material. In addition, differential exposure to light between the extracts and the in vitro assays may have led to differing extents of isomerization in the conjugated systems. We could nonetheless make confident peak assignments based on the mass spectra and comparison to the control analyses (extracts of the *lkcE* mutant vs. the wild type, analysis of active *LkcE* in vitro vs. inactivated protein).

Supplementary References

1. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1541–1546 (2003).
2. Dickschat, J. S. *et al. ChemBioChem* **12**, 2408–2412 (2011).
3. Arakawa, K., Sugino, F., Kodama, K., Ishii, T. & Kinashi, H. *Chem. Biol.* **12**, 249–256 (2005).
4. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. Mol. Biol.* **215**, 403–410 (1990).
5. Ng, B.-G., Han, J.-W., Lee, D. W., Choi, G. J., Kim, B. S. *J. Antibiot.* (2018). doi.org/10.1038/s41429-017-0023-x.
6. Kachalova, G., Decker, K., Holt, A., & Bartunik, H. D. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4800–4805 (2011).
7. Binda, C., Newton-Vinson, P., Hubálek, F., Edmondson, D. E. & Mattevi, A. *Nat. Struct. Biol.* **9**, 22–26 (2002).

DISCUSSION ET PERSPECTIVES

En raison de leur caractère modulaire et de la grande diversité structurale et thérapeutique des molécules qu'elles produisent, les PKS de type I modulaires ont été une cible préférentielle d'ingénierie ces trente dernières années (Weissman, 2016). Malheureusement, plusieurs stratégies prometteuses de modification de ces PKS, telles que l'inactivation, l'échange ou l'insertion de domaines, se sont avérées peu efficaces (Weissman, 2016). Les études structurales les plus récentes suggèrent que les modules subissent d'importants changements conformationnels au cours de chaque cycle d'élongation/modification. Ils impliquent des mouvements significatifs et synergiques des domaines guidés par la formation d'interfaces compatibles entre elles (Whicher et al., 2014). Ce phénomène expliquerait en grande partie l'échec des expériences d'ingénierie, et surtout celles impliquant des changements de domaine. Plusieurs solutions encourageantes permettant de contourner ces problèmes existent cependant. Une première consiste à échanger entre différents systèmes PKS des modules voire des sous-unités entières de manière à pallier ces problèmes structuraux, car les unités fonctionnelles des PKS restent intactes.

Domaines de *docking* : études fondamentales et exploitation en biologie synthétique

Dans le but de mettre en place cette stratégie de manière efficace et sensée, des connaissances supplémentaires sont nécessaires, notamment vis-à-vis des domaines de *docking* (DD). Ces petits domaines structuraux situés aux extrémités C_{ter} et N_{ter} des sous-unités de PKS assurent les interactions spécifiques entre ces dernières. Les DD sont des garants du bon acheminement de l'intermédiaire polycétidique d'une sous-unité à la suivante et assurent ainsi en partie l'obtention de la molécule finale active. Ces domaines sont donc prometteurs pour l'ingénierie, car ils offrent la possibilité de générer des combinaisons de modules/sous-unités par création de nouvelles interfaces productives. Plusieurs classes structurales de DD ont déjà été caractérisées chez les PKS *cis*-AT (Broadhurst et al., 2003 ; Buchholz et al., 2009 ; Whicher et al., 2013). Elles partagent plusieurs points communs : les DD sont structurés en hélices α et ils interagissent ensemble grâce à des interactions hydrophobes mais aussi électrostatiques. Ces dernières établissent un « code de spécificité », même s'il n'est pas toujours parfaitement compris, pour les DD appartenant à la même classe. Les ^NDD sont tous homodimériques, *via* la formation d'un *coiled-coil*. Ils promeuvent ainsi l'homodimérisation de leur propre sous-unité, ce qui participerait à sa stabilisation (Aparicio et al., 1996 ; Gokhale, 1999). Les interactions médiées par ces domaines sont toutes d'une affinité moyenne, comprise entre 2 et 24 μ M.

Au cours de mon travail de thèse, j'ai eu l'occasion de travailler sur plusieurs interfaces de *docking*. La première est située entre les deux premières sous-unités (VirA et VirFG) d'une PKS *trans*-AT, la virginiamycine (Vir) synthase. Jusqu'en 2015, l'existence de DD chez ces PKS n'était qu'hypothétique, car aucune information n'était disponible. Dans ce contexte, nous étions la première équipe à avoir mis en évidence un couple de DD chez les PKS *trans*-AT. La structure RMN d'une protéine de fusion entre les deux partenaires a révélé que la stratégie d'interaction est très similaire à celle trouvée chez les *cis*-AT : des interactions hydrophobes accompagnées d'interactions électrostatiques, mais à la différence des DD caractérisés chez les *cis*-AT, aucun de ces deux DD de la PKS *trans*-AT n'est homodimérique. La structure respective des deux partenaires est pourtant assez similaire : deux hélices α reliées par une boucle et disposées de manière parallèle pour former un fagot de 4 hélices α . VirA ^CDD et VirFG ^NDD constituent ainsi une nouvelle classe de DD qui serait spécifique aux PKS *trans*-AT.

De plus, nous avons pu mettre en évidence, par le biais de plusieurs techniques (RMN, CD et SAXS), que l'un des deux partenaires, le VirFG ^NDD, est une région intrinsèquement désordonnée de la PKS. Lorsque ce domaine est seul en solution, il n'aborde que très peu de structuration secondaire, c'est-à-dire qu'il ne se structure en hélices α de manière stable que lors de l'interaction avec son partenaire. C'est la première fois qu'une région intrinsèquement désordonnée (IDR) est mise en évidence chez les PKS, mais aussi dans une voie de synthèse appartenant à du métabolisme secondaire chez les bactéries. Globalement les IDR et les protéines intrinsèquement désordonnées (IDP) sont rares chez les procaryotes (Williamson et Potts, 2012), mais elles présentent la caractéristique intéressante d'offrir une interaction à la fois très spécifique et d'affinité moyenne voire faible (Uversky et Dunker, 2013). En effet, un K_d de 5,8 μ M a été déterminé entre le VirA ^CDD et le VirFG ^NDD par ITC.

Par une approche couplée de mutagenèse dirigée et d'ITC, nous avons confirmé la présence de trois couples clés de résidus impliqués dans des interactions électrostatiques entre les deux partenaires. Au cours de ces expériences nous avons pu noter que des mutations du VirFG ^NDD induisaient une structuration de ce dernier. Il devenait donc impossible de savoir si la perte d'affinité observée était due réellement à la mutation introduite ou à une perte de la capacité de ce dernier à subir une structuration induite native. Cette caractéristique est typique de ces protéines, puisque leur caractère désordonné est intrinsèque à leur structure primaire. Une mutation peut alors changer ce caractère par l'introduction d'un acide aminé dit « structurant » (Oldfield & Dunker, 2014).

Cette nouvelle classe de DD serait présente à plusieurs autres jonctions de sous-unités de PKS *trans*-AT : par exemple, au niveau d'interfaces classiques ACP/ KS, mais aussi ACP/KR ou ER au niveau de modules « splités », et entre ACP et domaine de condensation ou d'hétérocyclisation appartenant à des modules NRPS (Dorival et al., 2016). Malheureusement, aucun code électrostatique clair n'a pu être mis en évidence au sein de ces nouvelles jonctions identifiées. Au vu de ces incertitudes et de la nature intrinsèquement désordonnée de certains des partenaires de cette classe, il est impossible d'envisager des mutations rationnelles de ces domaines dans le cadre d'expériences d'ingénierie visant à la communication non native de sous-unités de PKS *trans*-AT. L'échange direct d'un couple de DD appartenant serait donc plus approprié que des mutations dirigées.

Suite à cette publication, quatre couples de DD appartenant à la même classe que VirA^{CDD} et VirFG^{NDD} ont été mis en évidence chez la PKS synthétisant la macrolactine (Zeng, et al., 2016). Chez cette PKS, quatre modules « splités » sont présents. Trois d'entre eux consistent en une interface DH/KR, la quatrième en une interface KS/KR ; dans tous les cas, des DD de la famille de VirA^{CDD}/VirFG^{NDD} seraient présents à ces interfaces. La structure de l'un des domaines KR ainsi que son DD en N_{ter}, situé à l'interface KS/KR, a été résolue par cristallographie à 1,72 Å de résolution (Figure 54 page 104).

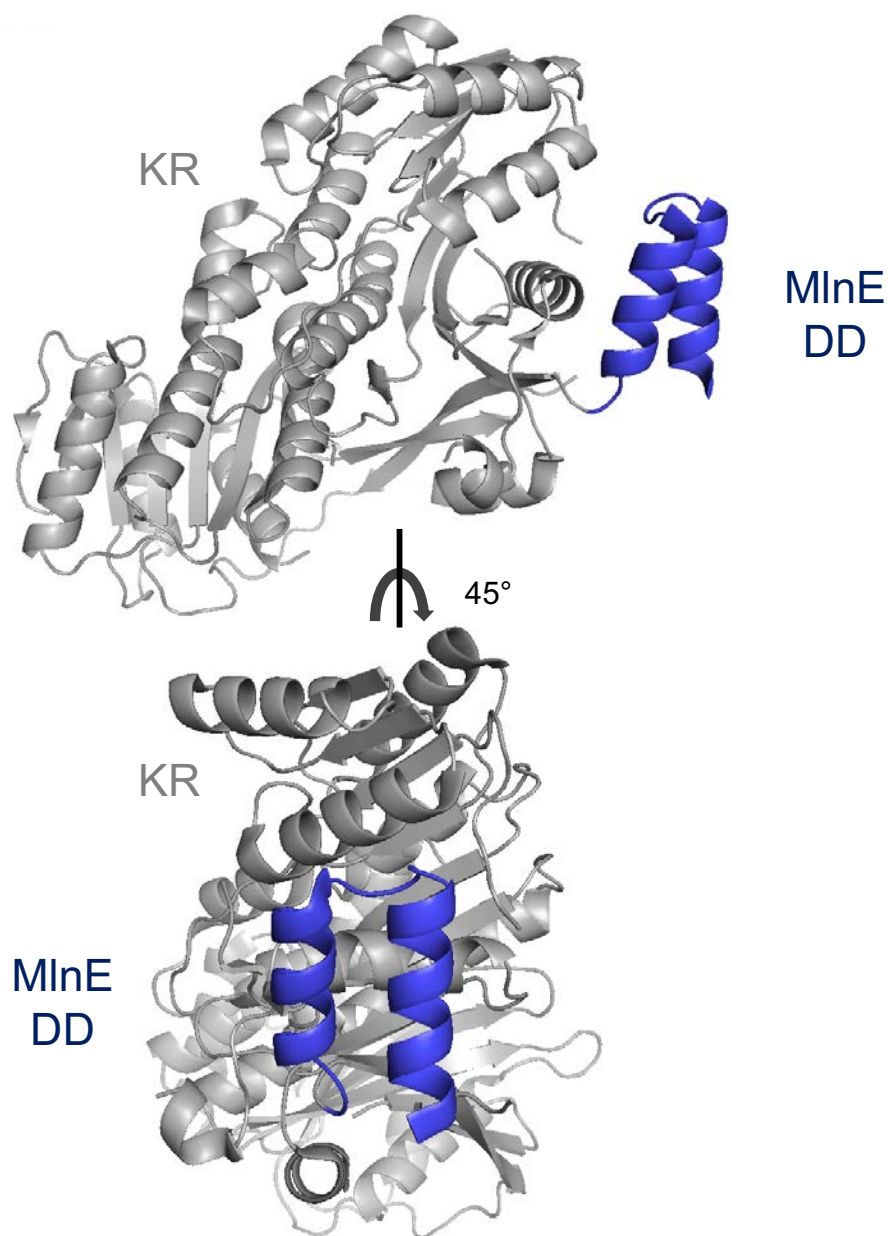


Figure 54 : Structure de la KR N_{ter} (en gris) de la sous-unité MlnE présentant un DD (en bleu) de la classe de VirA/VirFG (PDB : 5D2E).

La KR présente en N_{ter} un DD, MlnE $^{\text{NDD}}$, permettant l'interaction avec son partenaire, MlnD $^{\text{CDD}}$, situé en C_{ter} du domaine KS de la sous unité précédente. La structure du MlnE $^{\text{NDD}}$ est identique à celle adoptée par le VirFG $^{\text{NDD}}$ dans le contexte de la protéine de fusion, bien qu'ici le partenaire du MlnE $^{\text{NDD}}$ soit absent (le *rmsd* entre les $C\alpha$ des deux DD est de 1,16 Å). Des interactions impliquent la deuxième hélice α du DD avec des résidus du *linker* le reliant à la KR mais aussi avec des résidus du domaine. Ces interactions n'expliquent cependant pas à elles seules la structuration du $^{\text{NDD}}$. En effet, en faisant apparaître les protéines symétriques du système cristallin, nous voyons qu'une autre copie du

MlnE^{NDD} entre en interaction avec celui-ci (Dorival, 2016). Ce symétrique mime la présence du DD partenaire et induit vraisemblablement son repliement (Figure 55 page 105).

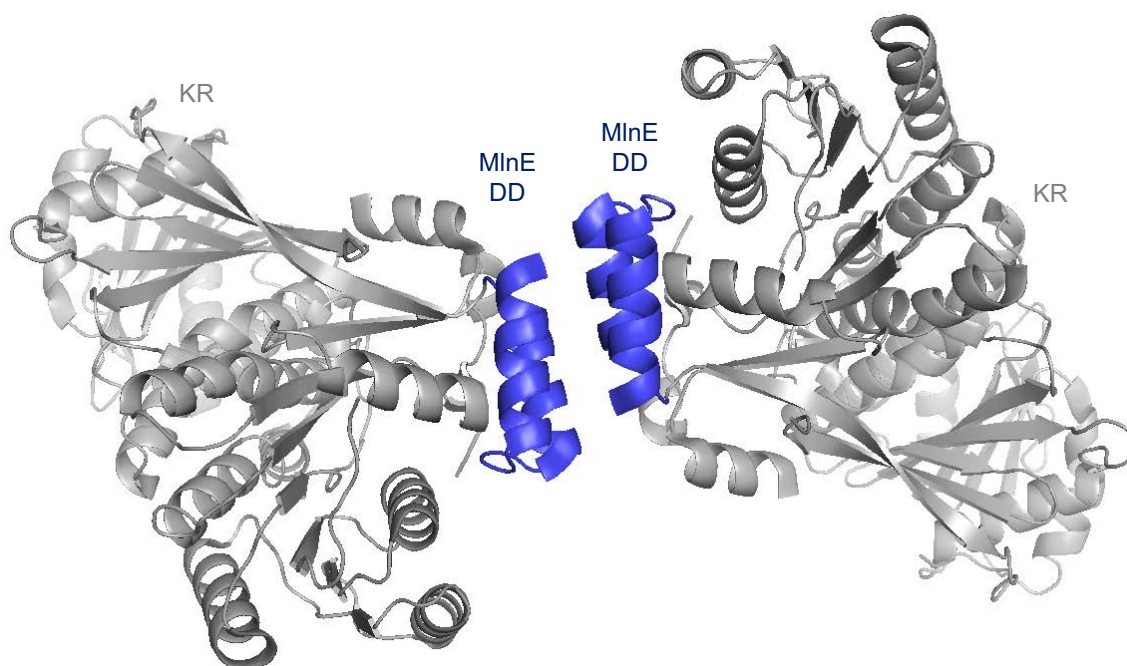


Figure 55 : Visualisation de la symétrie cristalline de la KR (en gris), s'homodimérisant par l'intermédiaire du DD (en bleu) (PDB : 5D2E).

L'homodimérisation du DD peut expliquer sa structuration. Il faut noter que les deux hélices α de chaque DD se disposent de manière antiparallèle, contrairement aux DD de VirA et VirFG dont les hélices α se positionnent à 127°.

Zeng et collaborateurs ont aussi démontré la portabilité des DD de cette famille. Les ^CDD de ce système ont été échangés et accolés à d'autres domaines situés à l'extrémité C_{ter} de leur sous-unité et ont permis la communication avec leur partenaire ^NDD natif (Zeng et al., 2016). Ces résultats n'ont été obtenus que par chromatographie d'exclusion de taille, et demanderaient donc à être confirmés par d'autres techniques plus précises, afin de déterminer si l'affinité d'interaction entre les partenaires est impactée ou non. Ces résultats sont tout de même prometteurs quant à la possibilité d'utiliser les DD de cette famille dans les expériences d'échange de modules et de sous-unités chez les PKS *trans*-AT. Du fait que ces derniers soient aussi bien trouvés au niveau d'interfaces classiques (KS/ACP) qu'au niveau de modules « splités », la panoplie d'échanges possibles et donc de nouveaux dérivés polycétidiques obtenus est très large.

Au cours de notre étude du couple VirA ^CDD/VirFG ^NDD, nous nous sommes aussi intéressés à replacer l'interaction de ces derniers dans leur contexte modulaire d'interaction. Pour ce faire, nous avons superposé deux enveloppes SAXS distinctes qui ont pour point

commun de comporter l'ACP_{5b}. Nous avons pu ainsi proposer la première structure d'une interface entre deux sous-unités consécutives de PKS *trans*-AT. La première structure (KS₅-ACP_{5a}-ACP_{5b}-^cDD) (Davison et al, 2014) montrait une forme de pince ou d'arche, dont le sommet est fourni par la KS homodimérique, à laquelle est accolé l'ACP_{5a}. Les deux bras projetés à partir de ce sommet comprenaient un *linker* suivi de l'ACP_{5b} et du ^cDD, qui semblait en excellente position, projeté vers l'extérieur, pour aller contacter son partenaire. En effet, en superposant la deuxième structure (ACP_{5b}-^cDD-ⁿDD-KS₆), les bras de l'arche accommodent parfaitement la KS₆ suivie des DD en complexe. Ainsi même si les deux DD sont monomériques, l'interface d'interaction entre ces derniers est présente deux fois grâce à l'homodimérisation des KS₅ et KS₆. La surface totale d'interaction entre les DD est alors de 1460 Å², ce qui est du même ordre de grandeur que celles déterminées chez les DD de PKS *cis*-AT, qui sont comprises entre 1100 et 1800 Å². En revanche, la disposition latérale du VirFG ⁿDD par rapport à la KS₆ adjacente est inhabituelle, puisque chez les PKS *cis*-AT le ⁿDD formant un *coiled-coil* est projeté de manière axiale, facilitant l'ancrage du ^cDD et le transfert du polycétide de l'ACP vers la KS.

L'interaction entre les DD assure la formation d'une structure close, en forme d'arche, similaire à celle adoptée par le module 5 de la pikromycine déterminée par cryo-EM (Dutta et al., 2014) (Figure 56 page 106).

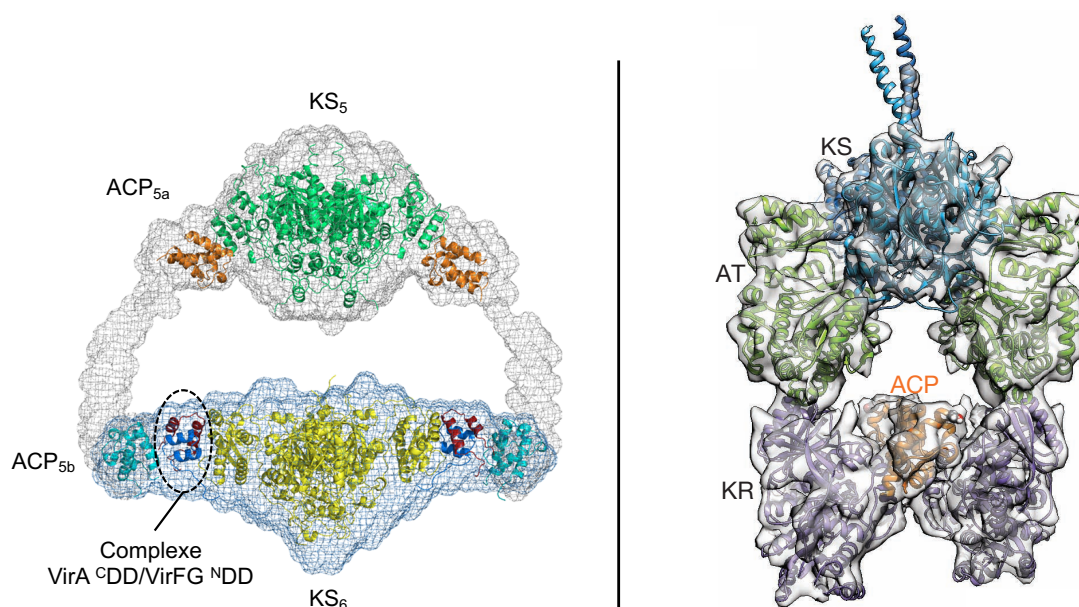


Figure 56 : Comparaison de la disposition en forme d'arche de l'interface VirA/VirFG et du module 5 de la pikromycine synthase (adapté de Dorival et al., 2016 et de Dutta et al., 2014).

A gauche la reconstruction par SAXS de l'interface entre les modules 5 et 6 de la virginiamycine synthase, situés aux extrémités des sous-unités VirA et VirFG, respectivement. En gris, l'enveloppe SAXS moyenne calculée, dans laquelle est replacé un modèle par homologie des KS. A droite la structure cryo-EM du module 5 de la pikromycine synthase, en transparent avec les structures à haute résolution positionnées au sein de la carte EM.

Chez la virginiamycine synthase, cette disposition particulière pourrait faciliter la formation d'un complexe multi-enzymatique avec les protéines de la cassette de β -méthylation intervenant au niveau des ACP du module 5. En outre, la structure *apo* proposée ici n'est pas compatible avec le transfert de la chaîne polycétidique, puisque la sérine de l'ACP_{5b}, portant normalement le groupement Ppant et l'intermédiaire, est séparée de 80 Å de la cystéine catalytique de la KS₆. Ces résultats sont pourtant en accord avec les structures cryo-EM citées précédemment (Whicher et al., 2014) : des interactions stables entre l'ACP et ses partenaires ne sont observées que lorsque ce premier est modifié avec un substrat. Des travaux complémentaires sont actuellement en cours au laboratoire dans le but de : i) déterminer les changements conformationnels subis par le module 5 de la virginiamycine synthase lorsque les ACP sont modifiés par un analogue de substrat, et ii) mieux comprendre les interactions entre ces derniers et les protéines de la cassette de β -méthylation.

Afin d'étoffer nos connaissances sur les DD et sur les mécanismes moléculaires responsables de la spécificité d'interaction entre sous-unités d'un même système, nous nous sommes également intéressés aux six interfaces de l'enacyloxine synthase. Cette PKS est attrayante pour plusieurs raisons : i) le nombre relativement élevé d'interfaces (Helfrich et Piel, 2016 ; Medema et al., 2015), avec pour conséquence que ce système doit présenter plusieurs barrières moléculaires au mésappariement, ii) l'enacyloxine est une molécule efficace contre *Acinetobacter baumannii*, une bactérie pathogène responsable d'infections nosocomiales sévères. Malheureusement, la molécule n'est pas assez stable pour être utilisée en clinique (Mahenthiralingam et al., 2011). Les résultats obtenus permettraient d'envisager des expériences d'ingénierie visant à optimiser cette molécule. Et iii) cette PKS est un hybride *cis*-/*trans*-AT, le deuxième seulement mis en évidence à ce jour (Mahenthiralingam et al., 2011), les cinq premières sous-unités étant de type *cis*-AT, et seule la dernière de type *trans*-AT. Ainsi l'étude de ce système permettrait de mieux comprendre comment ces deux types de PKS, aux origines évolutives distinctes, communiquent ensemble et des hybrides pourraient être reproduits de manière efficace.

Le premier travail bio-informatique n'a pas permis de déterminer de manière fiable les bornes des ^cDD potentiels : les régions en C_{ter} du dernier domaine conservé comprenaient en moyenne une cinquantaine de résidus au sein desquels une à deux hélices α pouvaient se former. Différentes versions ont dû être testées avec leur partenaire respectif afin d'éliminer les versions de DD les moins probables.

Par la suite, plusieurs éléments caractéristiques à toutes les interfaces ont émergé. En utilisant une approche combinant analyse bio-informatique, CD, ITC, SLS et SAXS, nous avons pu mettre en évidence que tous les ^CDD sont monomériques, alors que tous les ^NDD sont homodimériques. Tous les ^CDD et tous les ^NDD ont aussi la capacité de se replier en hélices α . Ces deux caractéristiques sont présentes chez les deux classes de DD de PKS *cis*-AT connues. De plus, tous les ^CDD présentent des pI acides alors que la majorité des ^NDD présentent des pI basiques.

D'autre part, par CD, SAXS et RMN, nous avons pu mettre en évidence que tous les ^CDD du système sont des IDR. Il faut noter qu'ils sont tous présents au niveau de sous-unités PKS *cis*-AT, mais jusqu'alors, les IDR n'ont été confirmées que chez les PKS *trans*-AT et les NRPS (Richter et al., 2008 ; Dorival et al., 2016 ; Jenner et al., 2018). L'enacyloxine étant le troisième système pour lequel la présence d'IDR est confirmée, cette stratégie d'interaction semble donc répandue de manière importante au sein des PKS et non pas seulement chez les *trans*-AT.

Chez les mammifères, les IDP et les IDR sont retrouvées au niveau de protéines impliquées dans des voies de signalisation complexes. Cette caractéristique particulière permet à ces protéines de pouvoir interagir de manière transitoire avec de nombreux partenaires (Fuxreiter et al., 2008). La région C_{ter} de la protéine p53 adopte ainsi différentes conformations selon le partenaire avec lequel elle interagit (Oldfield et al., 2008). Il semble que le fait que les IDR puissent interagir avec plusieurs partenaires réduit intrinsèquement leur capacité à interagir de manière stable. Au sein des PKS, les IDR doivent garantir l'interaction, certes de manière transitoire, mais avec un seul partenaire spécifique.

De manière à comprendre comment la spécificité d'interaction est garantie au sein de la PKS, nous nous sommes intéressés de plus près à la classe à laquelle les DD appartiennent. Par RMN, nous avons réussi à déterminer la structure à haute résolution de protéines de fusion entre les paires de DD de la première interface ainsi que de la dernière correspondant à l'hybride *cis*-/*trans*. Dans le premier cas, la topologie d'interaction adoptée par les partenaires est très proche de celle des DD de classe 2 (Whicher et al., 2013), alors que la dernière interface appartiendrait à la classe 1b (Buchholz et al., 2009). D'autre part, par alignements de séquences, nous avons pu mettre en évidence que ces deux classes, ainsi que la classe 1a, seraient présentes au niveau des autres interfaces. Le code électrostatique permettrait la discrimination entre les partenaires appartenant à la même classe, car bien que les pI des ^NDD soient globalement basiques, des variations sont tout de même à noter. Par exemple l'interface

5923/5922 appartiendrait elle aussi aux DD de classe 2 (Whicher et al., 2013) ou classe 2-like, tout comme la première interface. Cependant, dans le cas de la première interface du système, le 5924 ^NDD présente un pI de 9,7 alors que le 5922 ^NDD présente un pI de 5,4. Des interactions électrostatiques spécifiques entre partenaires doivent donc avoir lieu pour permettre de discriminer les appariements natifs des appariements non naturels entre DD appartenant à la même classe.

La structure RMN des DD de la première interface a d'ailleurs permis de proposer trois couples de résidus intervenant dans des interactions électrostatiques (Ser2622/Arg17 ; Glu2623/Lys32 ; Glu2620/Gln36). Nous avons donc cherché à connaître l'impact de ces résidus sur l'affinité d'interaction. Les trois résidus d'intérêt du ^NDD ont été mutés en un résidu de charge inverse (Glu), et en un résidu neutre (Ala). Étonnamment, aucun mutant n'a eu d'impact important sur le K_d . Il semblerait que l'interaction ne peut pas être disséquée en paires individuelles de résidus interagissant ensemble ; la spécificité d'appariement émergerait d'un ensemble de contacts à l'interface, ce qui la rendrait moins sensible aux mutations ponctuelles.

Une autre barrière au mésappariement entre sous-unités de la PKS résiderait également dans une interface d'interaction étendue aux domaines adjacents aux DD. Ce serait le cas notamment au niveau du module « splité » pour lequel aucune interaction n'a pu être observée avec les DD seuls, mais aussi au niveau de l'interface 5922/5921 chez laquelle le ^NDD seul est monomérique contrairement aux autres ^NDD du système. Des études précédentes ont mis en évidence le rôle important des domaines ACP et KS dans l'interaction entre sous-unités (Wu, Cane et Khosla, 2002 ; Weissman 2006a). Elles ont porté uniquement sur la dernière interface de l'érythromycine synthase. Il est donc important d'élargir ces premiers résultats à d'autres systèmes PKS de manière à préciser : les interactions mises en place entre les DD et leur domaine adjacent, l'impact des domaines sur le repliement des DD ainsi que les interfaces et résidus mis à contribution dans l'interaction entre ces domaines adjacents.

L'ensemble de ces résultats démontrent que la spécificité d'interaction entre DD n'émerge pas seulement d'un nombre limité de contacts polaires qui peuvent être ciblés par de la mutagenèse dirigée. Nous confirmons grâce à cette seconde étude que cette stratégie ne serait pas non plus la plus adaptée pour les DD de PKS *cis*-AT, et ce pour deux raisons principales : i) les IDR semblent être également répandues au sein de ces systèmes, et ii) il est compliqué de déterminer un code électrostatique clair en se basant sur des alignements de

séquences uniquement. La stratégie la plus opportune serait donc bien d'échanger directement des paires natives de DD, et nous avons au cours de cette étude pu en caractériser cinq de plus.

Quant à l'interface hybride *cis*-AT/*trans*-AT, la structure RMN du complexe formé par les DD à cette jonction révèle que ces derniers appartiendraient à la classe 1b. Grâce à une étude bio-informatique, nous pouvons émettre l'hypothèse que l'interface hybride caractérisée chez la kiromycine synthase appartiendrait à la classe 1a. Ainsi la classe 1 pourrait être utilisée de manière extensive afin de mettre en communication des sous-unités *cis*-AT avec des sous-unités *trans*-AT.

Caractérisation d'une enzyme post-PKS : LkcE

Une autre stratégie attrayante pour modifier les polycétides, consiste à utiliser des enzymes post-PKS. Afin d'élargir cette trousse à outils, nous nous sommes penchés sur LkcE. Cette enzyme est particulièrement intéressante puisqu'elle catalyse une réaction de macrocyclisation (réaction de Mannich intramoléculaire) couplée à une activité amide oxydase. Les groupements fonctionnels du substrat, amide et méthyl, reconnus par l'enzyme et impliqués dans la réaction, sont très répandus au sein des métabolites hybrides polycétide/NRP. Cette enzyme pourrait donc être exploitée de manière à circulariser de nouvelles molécules.

La détermination de la structure de LkcE par cristallographie des rayons X révèle un homodimère. Les monomères présentent un repliement caractéristique d'amine oxydase, mais l'interface d'homodimérisation est totalement différente de ses plus proches homologues structuraux, les monoamines oxydases humaines A et B. Grâce à la mise au point d'un système cinétique couplé, les paramètres cinétiques de l'enzyme ont pu être déterminés, de même que les résidus importants intervenant dans la catalyse. Nous avons ainsi pu mettre en évidence que chez LkcE, le résidu Arg326 joue un rôle catalytique, vraisemblablement en stabilisant le substrat dans sa conformation pré-cyclique. Chez d'autres amine oxydases, le résidu en même position est compris dans l'interface d'homodimérisation. Le mode alternatif d'oligomérisation de LkcE aurait donc permis l'émergence de son activité macrocyclase.

Nous avons pu obtenir une structure cristallographique d'un mutant catalytique de l'enzyme en présence du substrat naturel en trempant les cristaux. Les C2 et C18 du substrat, impliqués dans la réaction de cyclisation, y sont séparés de 13,2 Å. Cette structure obtenue

présenterait donc l'enzyme dans une conformation de pré-cyclisation. Cette dernière devrait subir un changement conformationnel important afin de rapprocher les deux fonctions. Au contraire des amine oxydases, le domaine de fixation du cofacteur FAD de LkcE n'est pas impliqué dans cette interface, ce qui laisserait de la liberté aux monomères de subir ce changement conformationnel. Malheureusement nous n'avons jamais réussi à co-cristalliser le mutant catalytique en présence du substrat en raison de l'instabilité des cristaux. Des expériences de dynamiques moléculaires restent à être menées afin d'élucider cet aspect important de la catalyse.

D'autre part nous avons pu mettre en évidence que LkcE est capable d'accommoder plusieurs substrats présentant des variations au niveau des carbones C6 et C7. Nous avons démontré que l'enzyme est aussi capable de fixer les analogues structuraux du substrat naturel qui sont d'un poids moléculaire bien moindre. Ces résultats sont très encourageants mais des études complémentaires visant à cartographier plus précisément la spécificité de substrat de cette enzyme seront nécessaires pour pouvoir l'utiliser dans des voies de synthèses de nouveaux composés.

Ainsi nos données supportent que LkcE et son activité particulière d'amide oxydase/macrocyclase auraient émergé au cours de l'évolution *via* la mise en place d'un mode d'homodimérisation alternatif à partir d'une enzyme ancestrale de type monoamine oxydase. Ces résultats seront utiles dans l'avenir pour favoriser des efforts de modification des sites actifs enzymatiques existants afin de catalyser des nouvelles réactions.

BIBLIOGRAPHIE

Agarwal V., Moore B. S. « Fungal polyketide engineering comes of age ». *Proc. Natl. Acad. Sci.* [En ligne]. 2014. Vol. 111, n°34, p. 12278-12279. Disponible sur : < <https://doi.org/10.1073/pnas.1412946111> >

Alberts A. W. « Discovery, biochemistry and biology of lovastatin ». *Am. J. Cardiol.* [En ligne]. 1988. Vol. 62, n°15, p. J10-J15. Disponible sur : < [https://doi.org/10.1016/0002-9149\(88\)90002-1](https://doi.org/10.1016/0002-9149(88)90002-1) >

Alekseyev V. Y., Liu C. W., Cane D. E., Puglisi J. D., Khosla C. « Solution structure and proposed domain-domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase ». *Protein Sci.* [En ligne]. 2007. Vol. 16, n°10, p. 2093-2107. Disponible sur : < <https://doi.org/10.1110/ps.073011407> >

Ames B. D., Haynes S. W., Gao X., Evans B. S., Kelleher N. L., Tang Y., Walsh C. T. « Complexity Generation in Fungal Peptidyl Alkaloid Biosynthesis: Oxidation of Fumiquinazoline A to the Heptacyclic Hemiaminal Fumiquinazoline C by the Flavoenzyme Af12070 from *Aspergillus fumigatus* ». *Biochemistry* [En ligne]. 2011. Vol. 50, n°40, p. 8756-8769. Disponible sur : < <https://doi.org/10.1021/bi201302w> >

Annaval T., Paris C., Leadlay P. F., Jacob C., Weissman K. J. « Evaluating Ketoreductase Exchanges as a Means of Rationally Altering Polyketide Stereochemistry ». *ChemBioChem* [En ligne]. 2015. Vol. 16, n°9, p. 1357-1364. Disponible sur : < <https://doi.org/10.1002/cbic.201500113> >

Aparicio J. F., Molnár I., Schwecke T., König A., Haydock S. F., Khaw L. E., Staunton J., Leadlay P. F. « Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase ». *Gene*. 1996. Vol. 169, n°1, p. 9-16. Disponible sur : < [10.1016/0378-1119\(95\)00800-4](https://doi.org/10.1016/0378-1119(95)00800-4) >

Arakawa K., Mochizuki S., Yamada K., Noma T., Kinashi H. « Gamma-Butyrolactone autoregulator-receptor system involved in lankacidin and lankamycin production and morphological differentiation in *Streptomyces rochei* ». *Microbiology* [En ligne]. 2007. Vol. 153, n°6, p. 1817-1827. Disponible sur : < <https://doi.org/10.1099/mic.0.2006/002170-0> >

Arakawa K., Sugino F., Kodama K., Ishii T., Kinashi H. « Cyclization Mechanism for the Synthesis of Macrocyclic Antibiotic Lankacidin in *Streptomyces rochei* ». *Chem. Biol.* [En ligne]. 2005. Vol. 12, n°2, p. 249-256. Disponible sur : < <https://doi.org/10.1016/j.chembiol.2005.01.009> >

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.* [En ligne]. 2010. Vol. 107, n°5, p. 1983-1988. Disponible sur : < <https://doi.org/10.1073/pnas.0914100107> >

Baerga-Ortiz A., Popovic B., Siskos A. P., O'Hare H. M., Spiteller D., Williams M. G., Campillo N., Spencer J. B., Leadlay P. F. « Directed Mutagenesis Alters the Stereochemistry of Catalysis by Isolated Ketoreductase Domains from the Erythromycin Polyketide Synthase ». *Chem. Biol.* [En ligne]. 2006. Vol. 13, n°3, p. 277-285. Disponible sur : < <https://doi.org/10.1016/j.chembiol.2006.01.004> >

Baltz R. H. « Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? ». *J. Ind. Microbiol. Biotechnol.* [En ligne]. 2006. Vol. 33, n°7, p. 507-513. Disponible sur : < <https://doi.org/10.1007/s10295-005-0077-9> >

Behnken S., Hertweck C. « Cryptic Polyketide Synthase Genes in Non-Pathogenic *Clostridium SPP* ». *PLoS ONE* [En ligne]. 2012. Vol. 7, n°1, p. e29609. Disponible sur : < <https://doi.org/10.1371/journal.pone.0029609> >

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.* [En ligne]. 2011. Vol. 108, n°7, p. 2717-2722. Disponible sur : < <https://doi.org/10.1073/pnas.1019406108> >

Bevitt D. J., Cortes J., Haydock S. F., Leadlay P. F. « 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme ». *Eur. J. Biochem.* [En ligne]. 1992. Vol. 204, n°1, p. 39-49. Disponible sur : < <https://doi.org/10.1111/j.1432-1033.1992.tb16603.x> >

Bevitt D. J., Staunton J., Leadlay P. F. « Mutagenesis of the dehydratase active site in the erythromycin-producing polyketide synthase ». *Biochem. Soc. Trans.* [En ligne]. 1993. Vol. 21, p. 30S. Disponible sur : < <https://doi.org/10.1042/bst021030s> >

Birch A. J., Massy-Westropp R. A., Moye C. J. « Studies in relation to biosynthesis. VII. 2-Hydroxy-6-methylbenzoic acid in *Penicillium griseofulvum* Dierckx ». Aust. J. Chem. 1955. Vol. 8, n°4, p. 539-544. Disponible sur : < <https://doi.org/10.1071/CH9550539> >

Breitling R., Cenicerio A., Jankevics A., Takano E. « Metabolomics for Secondary Metabolite Research ». Metabolites [En ligne]. 2013. Vol. 3, n°4, p. 1076-1083. Disponible sur : < <https://doi.org/10.3390/metabo3041076> >

Brignole E. J., Smith S., Asturias F. J. « Conformational flexibility of metazoan fatty acid synthase enables catalysis ». Nat. Struct. Mol. Biol. [En ligne]. 2009. Vol. 16, n°2, p. 190-197. Disponible sur : < <https://doi.org/10.1038/nsmb.1532> >

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. 2003. Vol. 10, n°8, p. 723–731. Disponible sur : < [https://doi.org/10.1016/S1074-5521\(03\)00156-X](https://doi.org/10.1016/S1074-5521(03)00156-X) >

Buchan D. W. A., Jones D. T. « The PSIPRED Protein Analysis Workbench: 20 years on ». Nucleic Acids Res. [En ligne]. 2019. Vol. 47, n°W1, p. W402-W407. Disponible sur : < <https://doi.org/10.1093/nar/gkz297> >

Buchholz T. J., Geders T. W., Bartley 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. [En ligne]. 2009. Vol. 4, n°1, p. 41-52. Disponible sur : < <https://doi.org/10.1021/cb8002607> >

Caffrey P. « Conserved Amino Acid Residues Correlating With Ketoreductase Stereospecificity in Modular Polyketide Synthases ». ChemBioChem [En ligne]. 2003. Vol. 4, n°7, p. 654-657. Disponible sur : < <https://doi.org/10.1002/cbic.200300581> >

Calderone C. T., Bumpus S. B., Kelleher N. L., Walsh C. T., Magarvey N. A. « A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both α - and β -ketone reduction during chain growth ». Proc. Natl. Acad. Sci. [En ligne]. 2008. Vol. 105, n°35, p. 12809-12814. Disponible sur : < <https://doi.org/10.1073/pnas.0806305105> >

Cheng Y.-Q., Tang G.-L., Shen B. « Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis ». Proc. Natl. Acad. Sci. 2003. 2003. Vol. 100, n°6, p. 3149–3154. Disponible sur : < <https://doi.org/10.1073/pnas.0537286100> >

Cocito C. « Antibiotics of the Virginiamycin Family, Inhibitors Which Contain Synergistic Components ». 1979. Vol. 43, n°2, p. 145-198. PMCID : PMC281470

Cortes J., Haydock S. F., Roberts G. A., Bevitt D. J., Leadlay P. F. « An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea* ». Nature [En ligne]. 1990. Vol. 348, n°6297, p. 176-178. Disponible sur : < <https://doi.org/10.1038/348176a0> >

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. [En ligne]. 2009. Vol. 42, n°5, p. 892-900. Disponible sur : < <https://doi.org/10.1107/S0021889809029288> >

Davies J., Davies D. « Origins and evolution of antibiotic resistance ». Microbiol. Mol. Biol. Rev. MMBR [En ligne]. 2010. Vol. 74, n°3, p. 417-433. Disponible sur : < <https://doi.org/10.1128/MMBR.00016-10> >

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 [En ligne]. 2014. Vol. 5, n°8, p. 3081-3095. Disponible sur : < <https://doi.org/10.1039/C3SC53511H> >

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. [En ligne]. 2003. Vol. 30, n°8, p. 489-494. Disponible sur : < <https://doi.org/10.1007/s10295-003-0062-0> >

Demain A. L. « From natural products discovery to commercialization: a success story ». J. Ind. Microbiol. Biotechnol. [En ligne]. 2006. Vol. 33, n°7, p. 486-495. Disponible sur : < <https://doi.org/10.1007/s10295-005-0076-x> >

Demain A. L. « Importance of microbial natural products and the need to revitalize their discovery ». J. Ind. Microbiol. Biotechnol. [En ligne]. 2014. Vol. 41, n°2, p. 185-201. Disponible sur : < <https://doi.org/10.1007/s10295-013-1325-z> >

Demain A. L., Fang A. « The natural functions of secondary metabolites ». Adv. Biochem. Eng. Biotechnol. 2001. Vol. 69, p. 1-39. Disponible sur : < https://doi.org/10.1007/3-540-44964-7_1 >

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. [En ligne]. 2004. Vol. 20, n°6, p. 1660-1665. Disponible sur : < <https://doi.org/10.1021/bp0497435> >

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 [En ligne]. 2011. Vol. 12, n°16, p. 2408-2412. Disponible sur : < <https://doi.org/10.1002/cbic.201100474> >

Donadio S., Katz L. « Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea* ». Gene [En ligne]. 1992. Vol. 111, n°1, p. 51-60. Disponible sur : < [https://doi.org/10.1016/0378-1119\(92\)90602-L](https://doi.org/10.1016/0378-1119(92)90602-L) >

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. 1993. Vol. 90, n°15, p. 7119–7123. Disponible sur : < <https://doi.org/10.1073/pnas.90.15.7119> >

Donadio S., Staver M. J., McAlpine J. B., Swanson S. J., Katz L. « Modular organization of genes required for complex polyketide biosynthesis ». Science. 1991. Vol. 252, n°5006, p. 675–679. Disponible sur : < <http://doi.org/10.1126/science.2024119> >

Dorival J. La biologie structurale des polykétide synthases de type trans-AT. [s.l.] : Université de Lorraine, 2016.

Dorival J., Annaïval T., Risser F., Collin S., Roblin P., Jacob C., Gruez A., Chagot B., Weissman K. J. « Characterization of Intersubunit Communication in the Virginiamycin trans-Acyl Transferase Polyketide Synthase ». J. Am. Chem. Soc. [En ligne]. 2016. Vol. 138, n°12, p. 4155-4167. Disponible sur : < <https://doi.org/10.1021/jacs.5b13372> >

Dunn B. J., Khosla C. « Engineering the acyltransferase substrate specificity of assembly line polyketide synthases ». J. R. Soc. Interface [En ligne]. 2013. Vol. 10, n°85, p. 20130297-20130297. Disponible sur : < <https://doi.org/10.1098/rsif.2013.0297> >

Dunn B. J., Watts K. R., Robbins T., Cane D. E., Khosla C. « Comparative Analysis of the Substrate Specificity of trans - versus cis- Acyltransferases of Assembly Line Polyketide Synthases ». Biochemistry [En ligne]. 2014. Vol. 53, n°23, p. 3796-3806. Disponible sur : < <https://doi.org/10.1021/bi5004316> >

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., Skinotis G. « Structure of a modular polyketide synthase ». *Nature* [En ligne]. 2014. Vol. 510, n°7506, p. 512-517. Disponible sur : < <https://doi.org/10.1038/nature13423> >

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.* [En ligne]. 2014. Vol. 426, n°11, p. 2229-2245. Disponible sur : < <https://doi.org/10.1016/j.jmb.2014.03.015> >

Fagan P. F., Palfey B. A. « Comprehensive Natural Products II ». In : *Compr. Nat. Prod.* II. 2010. p. 37-114.

Fischbach M. A., Walsh C. T. « Antibiotics for emerging pathogens ». *Science*. 2009. Vol. 325, n°5944, p. 1089–1093. Disponible sur : < <https://doi.org/10.1126/science.1176667> >

Fischbach M. A., Walsh C. T. « Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms ». *Chem. Rev.* [En ligne]. 2006. Vol. 106, n°8, p. 3468-3496. Disponible sur : < <https://doi.org/10.1021/cr0503097> >

Fitzpatrick P. F. « Oxidation of amines by flavoproteins ». *Arch. Biochem. Biophys.* [En ligne]. 2010. Vol. 493, n°1, p. 13-25. Disponible sur : < <https://doi.org/10.1016/j.abb.2009.07.019> >

Foulston L. « Genome mining and prospects for antibiotic discovery ». *Curr. Opin. Microbiol.* [En ligne]. 2019. Vol. 51, p. 1-8. Disponible sur : < <https://doi.org/10.1016/j.mib.2019.01.001> >

Fuxreiter M., Tompa P., Simon I., Uversky V. N., Hansen J. C., Asturias F. J. « Malleable machines take shape in eukaryotic transcriptional regulation ». *Nat. Chem. Biol.* [En ligne]. 2008. Vol. 4, n°12, p. 728-737. Disponible sur : < <https://doi.org/10.1038/nchembio.127> >

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*: Hybrid macrolide biosynthesis in *S. erythraea* ». *Mol. Microbiol.* [En ligne]. 2002. Vol. 44, n°3, p. 771-781. Disponible sur : < <https://doi.org/10.1046/j.1365-2958.2002.02910.x> >

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.* [En ligne]. 2013. Vol. 135, n°44, p. 16324-16327. Disponible sur : < <https://doi.org/10.1021/ja408944s> >

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.* [En ligne]. 2014. Vol. 136, n°29, p. 10190-10193. Disponible sur : < <https://doi.org/10.1021/ja5056998> >

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.* [En ligne]. 2016. Vol. 193, n°3, p. 196-205. Disponible sur : < <https://doi.org/10.1016/j.jsb.2015.12.011> >

Ghasemzadeh A., Jaafar H. Z. E. « Effect of CO(2) enrichment on synthesis of some primary and secondary metabolites in ginger ». *Int. J. Mol. Sci.* [En ligne]. 2011. Vol. 12, n°2, p. 1101-1114. Disponible sur : < <https://doi.org/10.3390/ijms12021101> >

Go M. K., Wongsantichon J., Cheung V. W. N., Chow J. Y., Robinson R. C., Yew W. S. « Synthetic Polyketide Enzymology: Platform for Biosynthesis of Antimicrobial Polyketides ». *ACS Catal.* [En ligne]. 2015. Vol. 5, n°7, p. 4033-4042. Disponible sur : < <https://doi.org/10.1021/acscatal.5b00477> >

Gokhale R. S. « Dissecting and Exploiting Intermodular Communication in Polyketide Synthases ». *Science* [En ligne]. 1999. Vol. 284, n°5413, p. 482-485. Disponible sur : < <https://doi.org/10.1126/science.284.5413.482> >

Gregory M. A., Petkovic H., Lill R. E., Moss S. J., Wilkinson B., Gaisser S., Leadlay P. F., Sheridan R. M. « Mutasynthesis of Rapamycin Analogues through the Manipulation of a Gene Governing Starter Unit Biosynthesis ». *Angew. Chem. Int. Ed.* [En ligne]. 2005. Vol. 44, n°30, p. 4757-4760. Disponible sur : < <https://doi.org/10.1002/anie.200462784> >

Hacker C., Cai X., Kegler C., Zhao L., Weickmann A. K., Wurm J. P., Bode H. B., Wöhnert J. « Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS ». *Nat. Commun.* [En ligne]. 2018. Vol. 9, n°1. Disponible sur : < <https://doi.org/10.1038/s41467-018-06712-1> >

Hans M., Hornung A., Dziarnowski A., Cane D. E., Khosla C. « Mechanistic Analysis of Acyl Transferase Domain Exchange in Polyketide Synthase Modules ». J. Am. Chem. Soc. [En ligne]. 2003. Vol. 125, n°18, p. 5366-5374. Disponible sur : < <https://doi.org/10.1021/ja029539i> >

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. [En ligne]. 2003. Vol. 330, n°5, p. 1061-1075. Disponible sur : < [https://doi.org/10.1016/S0022-2836\(03\)00668-5](https://doi.org/10.1016/S0022-2836(03)00668-5) >

Hayashi T., Suenaga I., Narukawa N., Yamazaki T. « In vitro and in vivo activities of sedecamycin against *Treponema hyodysenteriae*. » Antimicrob. Agents Chemother. [En ligne]. 1988. Vol. 32, n°4, p. 458-461. Disponible sur : < <https://doi.org/10.1128/AAC.32.4.458> >

He J., Müller M., Hertweck C. « Formation of the Aureothin Tetrahydrofuran Ring by a Bifunctional Cytochrome P450 Monooxygenase ». J. Am. Chem. Soc. [En ligne]. 2004. Vol. 126, n°51, p. 16742-16743. Disponible sur : < <https://doi.org/10.1021/ja046104h> >

Helfrich E. J. N., Piel J. « Biosynthesis of polyketides by trans-AT polyketide synthases ». Nat. Prod. Rep. [En ligne]. 2016. Vol. 33, n°2, p. 231-316. Disponible sur : < <https://doi.org/10.1039/C5NP00125K> >

Hertweck C., Luzhetskyy A., Rebets Y., Bechthold A. « Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork ». Nat Prod Rep [En ligne]. 2007. Vol. 24, n°1, p. 162-190. Disponible sur : < <https://doi.org/10.1039/B507395M> >

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. [En ligne]. 1998. Vol. 5, n°1, p. 35-47. Disponible sur : < [https://doi.org/10.1016/S1074-5521\(98\)90085-0](https://doi.org/10.1016/S1074-5521(98)90085-0) >

Hong H., Leadlay P. F., Staunton J. « The changing patterns of covalent active site occupancy during catalysis on a modular polyketide synthase multienzyme revealed by ion-trap mass spectrometry: Changing patterns of covalent active-site occupancy ». FEBS J. [En ligne]. 2009. Vol. 276, n°23, p. 7057-7069. Disponible sur : < <https://doi.org/10.1111/j.1742-4658.2009.07418.x> >

Howard A., O'Donoghue M., Feeney A., Sleator R. D. « *Acinetobacter baumannii*: An emerging opportunistic pathogen ». Virulence [En ligne]. 2012. Vol. 3, n°3, p. 243-250. Disponible sur : < <https://doi.org/10.4161/viru.19700> >

Jain M., Cox J. S. « Interaction between Polyketide Synthase and Transporter Suggests Coupled Synthesis and Export of Virulence Lipid in *M. tuberculosis* ». PLoS Pathog. [En ligne]. 2005. Vol. 1, n°1, p. e2. Disponible sur : < <https://doi.org/10.1371/journal.ppat.0010002> >

Jenke-Kodama H., Dittmann E. « Bioinformatic perspectives on NRPS/PKS megasynthases: Advances and challenges ». Nat. Prod. Rep. [En ligne]. 2009. Vol. 26, n°7, p. 874. Disponible sur : < <https://doi.org/10.1039/b810283j> >

Jenner M., Kosol S., Griffiths D., Prasongpholchai P., Manzi L., Barrow A. S., Moses J. E., Oldham N. J., Lewandowski J. R., Challis G. L. « Mechanism of intersubunit ketosynthase-dehydratase interaction in polyketide synthases ». Nat. Chem. Biol. [En ligne]. 8 2018. Disponible sur : < <https://doi.org/10.1038/nchembio.2549> >

Jez J. M., Bowman M. E., Noel J. P. « Structure-Guided Programming of Polyketide Chain-Length Determination in Chalcone Synthase ». Biochemistry [En ligne]. 2001. Vol. 40, n°49, p. 14829-14838. Disponible sur : < <https://doi.org/10.1021/bi015621z> >

Kabear J. N., Semana E., Mussa A. R., He X. « Plant Secondary Metabolites: Biosynthesis, Classification, Function and Pharmacological Properties ». J. Pharm. Pharmacol. 2014. Vol. 2, n°7, p. 377-392.

Kalkreuter E., CroweTipton J. M., Lowell A. N., Sherman D. H., Williams G. J. « Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender Units ». J. Am. Chem. Soc. [En ligne]. 2019. Vol. 141, n°5, p. 1961-1969. Disponible sur : < <https://doi.org/10.1021/jacs.8b10521> >

Kao C. M., McPherson M., McDaniel R. N., Fu H., Cane D. E., Khosla C. « Gain of Function Mutagenesis of the Erythromycin Polyketide Synthase. 2. Engineered Biosynthesis of an Eight-Membered Ring Tetraketide Lactone ». J. Am. Chem. Soc. [En ligne]. 1997. Vol. 119, n°46, p. 11339-11340. Disponible sur : < <https://doi.org/10.1021/ja972609e> >

Katz L., Baltz R. H. « Natural product discovery: past, present, and future ». J. Ind. Microbiol. Biotechnol. [En ligne]. 2016. Vol. 43, n°2-3, p. 155-176. Disponible sur : < <https://doi.org/10.1007/s10295-015-1723-5> >

Keatinge-Clay A. « Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase ». J. Mol. Biol. [En ligne]. 2008. Vol. 384, n°4, p. 941-953. Disponible sur : < <https://doi.org/10.1016/j.jmb.2008.09.084> >

Keatinge-Clay A. T. « A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides ». *Chem. Biol.* [En ligne]. 2007. Vol. 14, n°8, p. 898-908. Disponible sur : < <https://doi.org/10.1016/j.chembiol.2007.07.009> >

Keatinge-Clay A. T., Maltby D. A., Medzihradszky K. F., Khosla C., Stroud R. M. « An antibiotic factory caught in action ». *Nat. Struct. Mol. Biol.* [En ligne]. 2004. Vol. 11, n°9, p. 888-893. Disponible sur : < <https://doi.org/10.1038/nsmb808> >

Keatinge-Clay A. T., Stroud R. M. « The Structure of a Ketoreductase Determines the Organization of the β -Carbon Processing Enzymes of Modular Polyketide Synthases ». *Structure* [En ligne]. 2006. Vol. 14, n°4, p. 737-748. Disponible sur : < <https://doi.org/10.1016/j.str.2006.01.009> >

Kellenberger L., Galloway I. S., Sauter G., Böhm G., Hanefeld U., Cortés J., Staunton J., Leadlay P. F. « A Polylinker Approach to Reductive Loop Swaps in Modular Polyketide Synthases ». *ChemBioChem* [En ligne]. 2008. Vol. 9, n°16, p. 2740-2749. Disponible sur : < <https://doi.org/10.1002/cbic.200800332> >

Kim M.-S., Cho W.-J., Song M. C., Park S.-W., Kim K., Kim E., Lee N., Nam S.-J., Oh K.-H., Yoon Y. J. « Engineered biosynthesis of milbemycins in the avermectin high-producing strain *Streptomyces avermitilis* ». *Microb. Cell Factories* [En ligne]. 2017. Vol. 16, n°1, Disponible sur : < <https://doi.org/10.1186/s12934-017-0626-8> > (consulté le 15 mars 2019)

Kinashi H., Mori E., Hatani A., Nimi O. « Isolation and characterization of linear plasmids from lankacidin-producing *Streptomyces* species ». *J. Antibiot. (Tokyo)*. 1994. Vol. 47, n°12, p. 1447-1455. Disponible sur : < <https://doi.org/10.7164/antibiotics.47.1447> >

Kolb H. C., Finn M. G., Sharpless K. B. « Click Chemistry: Diverse Chemical Function from a Few Good Reactions ». *Angew. Chem. Int. Ed.* [En ligne]. 2001. Vol. 40, n°11, p. 2004-2021. Disponible sur : < [https://doi.org/10.1002/1521-3773\(20010601\)40:11<2004::AID-ANIE2004>3.0.CO;2-5](https://doi.org/10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5) >

Koryakina I., Kasey C., McArthur J. B., Lowell A. N., Chemler J. A., Li S., Hansen D. A., Sherman D. H., Williams G. J. « Inversion of Extender Unit Selectivity in the Erythromycin Polyketide Synthase by Acyltransferase Domain Engineering ». *ACS Chem. Biol.* [En ligne]. 2017. Vol. 12, n°1, p. 114-123. Disponible sur : < <https://doi.org/10.1021/acschembio.6b00732> >

Kosol S., Jenner M., Lewandowski J. R., Challis G. L. « Protein–protein interactions in trans-AT polyketide synthases ». *Nat. Prod. Rep.* [En ligne]. 2018. Vol. 35, n°10, p. 1097-1109. Disponible sur : < <https://doi.org/10.1039/C8NP00066B> >

Kosol S., Gallo A., Griffiths D., Valentic T. R., Masschelein J., Jenner M., De los Santos E. L. C., Manzi L., Sydor P. K., Rea D., Zhou S., Fülöp V., Oldham N. J., Tsai S.-C., Challis G. L., Lewandowski J. R. « Structural basis for chain release from the enacyloxin polyketide synthase ». *Nat. Chem.* [En ligne]. 2019. Vol. 11, n°10, p. 913-923. Disponible sur : < <https://doi.org/10.1038/s41557-019-0335-5> >

Kushnir S., Sundermann U., Yahiaoui S., Brockmeyer A., Janning P., Schulz F. « Minimally Invasive Mutagenesis Gives Rise to a Biosynthetic Polyketide Library ». *Angew. Chem. Int. Ed.* [En ligne]. 2012. Vol. 51, n°42, p. 10664-10669. Disponible sur : < <https://doi.org/10.1002/anie.201202438> >

Laureti L., Song L., Huang S., Corre C., Leblond P., Challis G. L., Aigle B. « Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens* ». *Proc. Natl. Acad. Sci.* [En ligne]. 2011. Vol. 108, n°15, p. 6258-6263. Disponible sur : < <https://doi.org/10.1073/pnas.1019077108> >

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 ». *Bioorganic Chem.* [En ligne]. 2004. Vol. 32, n°6, p. 549-559. Disponible sur : < <https://doi.org/10.1016/j.bioorg.2004.06.002> >

Lešnik U., Lukežič T., Podgoršek A., Horvat J., Polak T., Šala M., Jenko B., Harmrolfs K., Ocampo-Sosa A., Martínez-Martínez L., Herron P. R., Fujs Š., Kosec G., Hunter I. S., Müller R., Petković H. « Construction of a New Class of Tetracycline Lead Structures with Potent Antibacterial Activity through Biosynthetic Engineering ». *Angew. Chem.* [En ligne]. 2015. Vol. 127, n°13, p. 4009-4012. Disponible sur : < <https://doi.org/10.1002/ange.201411028> >

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.* [En ligne]. 1993. Vol. 90, n°15, p. 7010-7014. Disponible sur : < <https://doi.org/10.1073/pnas.90.15.7010> >

Linpinski C., Lombardo F., Dominy B. W., Feeney P. J. « Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings ». *Adv. Drug Deliv. Rev.* [En ligne]. 1997. Vol. 23, n°1-3, p. 1-25. Disponible sur : < [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1) >

Liu R., Li X., Lam K. S. « Combinatorial chemistry in drug discovery ». *Curr. Opin. Chem. Biol.* [En ligne]. 2017. Vol. 38, p. 117-126. Disponible sur : < <https://doi.org/10.1016/j.cbpa.2017.03.017> >

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.* [En ligne]. 2011. Vol. 18, n°5, p. 665-677. Disponible sur : < <https://doi.org/10.1016/j.chembiol.2011.01.020> >

Maier T., Leibundgut M., Ban N. « The Crystal Structure of a Mammalian Fatty Acid Synthase ». *Science* [En ligne]. 2008. Vol. 321, n°5894, p. 1315-1322. Disponible sur : < <https://doi.org/10.1126/science.1161269> >

Marsden A. F. A., Wilkinson B., Cortés J., Dunster N. J., Staunton J., Leadlay P. F. « Engineering Broader Specificity into an Antibiotic-Producing Polyketide Synthase ». *Science* [En ligne]. 1998. Vol. 279, n°5348, p. 199-202. Disponible sur : < <https://doi.org/10.1126/science.279.5348.199> >

Martens E., Demain A. L. « The antibiotic resistance crisis, with a focus on the United States ». *J. Antibiot. (Tokyo)* [En ligne]. 2017. Vol. 70, n°5, p. 520-526. Disponible sur : < <https://doi.org/10.1038/ja.2017.30> >

Masschelein J., Sydor P. K., Hobson C., Howe R., Jones C., Roberts D. M., Ling Yap Z., Parkhill J., Mahenthiralingam E., Challis G. L. « A dual transacylation mechanism for polyketide synthase chain release in enacyloxin antibiotic biosynthesis ». *Nat. Chem.* [En ligne]. 2019. Vol. 11, n°10, p. 906-912. Disponible sur : < <https://doi.org/10.1038/s41557-019-0309-7> >

Mast Y., Weber T., Götz M., Ort-Winklbauer R., Gondran A., Wohlleben W., Schinko E. « Characterization of the ‘pristinamycin supercluster’ of *Streptomyces pristinaespiralis*: Pristinamycin gene region of *S. pristinaespiralis* ». *Microb. Biotechnol.* [En ligne]. 2011. Vol. 4, n°2, p. 192-206. Disponible sur : < <https://doi.org/10.1111/j.1751-7915.2010.00213.x> >

McDaniel R., Thamchaipenet A., Gustafsson C., Fu H., Betlach Melanie, Betlach Mary, Ashley G. « Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel “unnatural” natural products ». *Proc. Natl. Acad. Sci.* 1999. Vol. 96, n°5, p. 1846–1851.

Medema M. H., Kottmann R., Yilmaz P., Cummings M., Biggins J. B., Blin K., De Bruijn I., Chooi Y. H., Claesen J., Coates R. C., Cruz-Morales P., Duddela S., Düsterhus S., Edwards D. J., Fewer D. P., Garg N., Geiger C., Gomez-Escribano J. P., Greule A., Hadjithomas M., Haines A. S., Helfrich E. J. N., Hillwig M. L., Ishida K., Jones A. C., Jones C. S., Jungmann K., Kegler C., Kim H. U., Kötter P., Krug D., Masschelein J., Melnik A. V., Mantovani S. M., Monroe E. A., Moore M., Moss N., Nützmann H.-W., Pan G., Pati A., Petras D., Reen F. J., Rosconi F., Rui Z., Tian Z., Tobias N. J., Tsunematsu Y., Wiemann P., Wyckoff E., Yan X., Yim G., Yu F., Xie Y., Aigle B., Apel A. K., Balibar C. J., Balskus E. P., Barona-Gómez F., Bechthold A., Bode H. B., Borriss R., Brady S. F., Brakhage A. A., Caffrey P., Cheng Y.-Q., Clardy J., Cox R. J., De Mot R., Donadio S., Donia M. S., Van der Donk W. A., Dorrestein P. C., Doyle S., Driessen A. J. M., Ehling-Schulz M., Entian K.-D., Fischbach M. A., Gerwick L., Gerwick W. H., Gross H., Gust B., Hertweck C., Höfte M., Jensen S. E., Ju J., Katz L., Kaysser L., Klassen J. L., Keller N. P., Kormanec J., Kuipers O. P., Kuzuyama T., Kyrpides N. C., Kwon H.-J., Lautru S., Lavigne R., Lee C. Y., Linquan B., Liu X., Liu W., Luzhetskyy A., Mahmud T., Mast Y., Méndez C., Metsä-Ketelä M., Micklefield J., Mitchell D. A., Moore B. S., Moreira L. M., Müller R., Neilan B. A., Nett M., Nielsen J., O’Gara F., Oikawa H., Osbourn A., Osburne M. S., Ostash B., Payne S. M., Pernodet J.-L., Petricek M., Piel J., Ploux O., Raaijmakers J. M., Salas J. A., Schmitt E. K., Scott B., Seipke R. F., Shen B., Sherman D. H., Sivonen K., Smanski M. J., Sosio M., Stegmann E., Süssmuth R. D., Tahlan K., Thomas C. M., Tang Y., Truman A. W., Viaud M., Walton J. D., Walsh C. T., Weber T., Van Wezel G. P., Wilkinson B., Willey J. M., Wohlleben W., Wright G. D., Ziemert N., Zhang C., Zotchev S. B., Breitling R., Takano E., Glöckner F. O. « Minimum Information about a Biosynthetic Gene cluster ». *Nat. Chem. Biol.* [En ligne]. 2015. Vol. 11, n°9, p. 625-631. Disponible sur : < <https://doi.org/10.1038/nchembio.1890> >

Mészáros B., Erdős G., Dosztányi Z. « IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding ». *Nucleic Acids Res.* [En ligne]. 2018. Vol. 46, n°W1, p. W329-W337. Disponible sur : < <https://doi.org/10.1093/nar/gky384> >

Menzella H. G., Carney J. R., Santi D. V. « Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases ». *Chem. Biol.* [En ligne]. 2007. Vol. 14, n°2, p. 143-151. Disponible sur : < <https://doi.org/10.1016/j.chembiol.2006.12.002> >

Metz J. G. « Production of Polyunsaturated Fatty Acids by Polyketide Synthases in Both Prokaryotes and Eukaryotes ». *Science* [En ligne]. 2001. Vol. 293, n°5528, p. 290-293. Disponible sur : < <https://doi.org/10.1126/science.1059593> >

Minowa Y., Araki M., Kanehisa M. « Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes ». *J. Mol. Biol.* [En ligne]. 2007. Vol. 368, n°5, p. 1500-1517. Disponible sur : < <https://doi.org/10.1016/j.jmb.2007.02.099> >

Musiol E. M., Greule A., Härtner T., Kulik A., Wohlleben W., Weber T. « The AT 2 Domain of KirCI Loads Malonyl Extender Units to the ACPs of the Kirromycin PKS ». *ChemBioChem* [En ligne]. 2013. Vol. 14, n°11, p. 1343-1352. Disponible sur : < <https://doi.org/10.1002/cbic.201300211> >

Musiol E. M., Härtner T., Kulik A., Moldenhauer J., Piel J., Wohlleben W., Weber T. « Supramolecular Templating in Kirromycin Biosynthesis: The Acyltransferase KirCII Loads Ethylmalonyl-CoA Extender onto a Specific ACP of the trans-AT PKS ». *Chem. Biol.* [En ligne]. 2011. Vol. 18, n°4, p. 438-444. Disponible sur : < <https://doi.org/10.1016/j.chembiol.2011.02.007> >

Musiol-Kroll E. M., Zubeil F., Schafhauser T., Härtner T., Kulik A., McArthur J., Koryakina I., Wohlleben W., Grond S., Williams G. J., Lee S. Y., Weber T. « Polyketide Bioderivatization Using the Promiscuous Acyltransferase KirCII ». *ACS Synth. Biol.* [En ligne]. 2017. Vol. 6, n°3, p. 421-427. Disponible sur : < <https://doi.org/10.1021/acssynbio.6b00341> >

Nafissi N., Slavcev R. « Bacteriophage recombination systems and biotechnical applications ». *Appl. Microbiol. Biotechnol.* [En ligne]. 2014. Vol. 98, n°7, p. 2841-2851. Disponible sur : < <https://doi.org/10.1007/s00253-014-5512-2> >

Namwat W., Kamioka Y., Kinoshita H., Yamada Y., Nihira T. « Characterization of virginiamycin S biosynthetic genes from *Streptomyces virginiae* ». *Gene* [En ligne]. 2002. Vol. 286, n°2, p. 283-290. Disponible sur : < [https://doi.org/10.1016/S0378-1119\(02\)00424-9](https://doi.org/10.1016/S0378-1119(02)00424-9) >

Newman D. J. « Natural Products as Leads to Potential Drugs: An Old Process or the New Hope for Drug Discovery? ». *J. Med. Chem.* [En ligne]. 2008. Vol. 51, n°9, p. 2589-2599. Disponible sur : < <https://doi.org/10.1021/jm0704090> >

Newman D. J., Cragg G. M. « Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010 ». *J. Nat. Prod.* [En ligne]. 2012. Vol. 75, n°3, p. 311-335. Disponible sur : < <https://doi.org/10.1021/np200906s> >

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.* [En ligne]. 2008. Vol. 26, n°2, p. 225-233. Disponible sur : < <https://doi.org/10.1038/nbt1379> >

Niccolai D., Tarsi L., Thomas R. J. « The renewed challenge of antibacterial chemotherapy ». *Chem. Commun.* [En ligne]. 1997. n°24, p. 2333-2342. Disponible sur : < <https://doi.org/10.1039/a704497f> >

Nilius A. « Ketolides: the future of the macrolides? ». *Curr. Opin. Pharmacol.* [En ligne]. 2002. Vol. 2, n°5, p. 493-500. Disponible sur : < [https://doi.org/10.1016/S1471-4892\(02\)00198-4](https://doi.org/10.1016/S1471-4892(02)00198-4) >

O'Connell K. M. G., Galloway W. R. J. D., Spring D. R. « The Basics of Diversity-Oriented Synthesis ». In : *Divers. Oriented Synth.* [En ligne]. Hoboken, NJ, USA : John Wiley & Sons, Inc., 2013. p. 1-26. Disponible sur : < <https://doi.org/10.1002/9781118618110.ch1> > (consulté le 8 novembre 2018) ISBN : 978-1-118-61811-0.

O'Hare H. M., Baerga-Ortiz A., Popovic B., Spencer J. B., Leadlay P. F. « High-Throughput Mutagenesis to Evaluate Models of Stereochemical Control in Ketoreductase Domains from the Erythromycin Polyketide Synthase ». *Chem. Biol.* [En ligne]. 2006. Vol. 13, n°3, p. 287-296. Disponible sur : < <https://doi.org/10.1016/j.chembiol.2006.01.003> >

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.* [En ligne]. 2010. Vol. 27, n°4, p. 571. Disponible sur : < <https://doi.org/10.1039/b911956f> >

Oldfield C. J., Dunker A. K. « Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions ». *Annu. Rev. Biochem.* [En ligne]. 2014. Vol. 83, n°1, p. 553-584. Disponible sur : < <https://doi.org/10.1146/annurev-biochem-072711-164947> >

Oldfield C. J., Meng J., Yang J. Y., Yang M. Q., Uversky V. N., Dunker A. K. « Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners ». BMC Genomics [En ligne]. 2008. Vol. 9, n°Suppl 1, p. S1. Disponible sur : < <https://doi.org/10.1186/1471-2164-9-S1-S1> >

Oliynyk M., Brown M. J. B., Cortés J., Staunton J., Leadlay P. F. « A hybrid modular polyketide synthase obtained by domain swapping ». Chem. Biol. [En ligne]. 1996. Vol. 3, n°10, p. 833-839. Disponible sur : < [https://doi.org/10.1016/S1074-5521\(96\)90069-1](https://doi.org/10.1016/S1074-5521(96)90069-1) >

Oliynyk M., Samborsky 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. [En ligne]. 2007. Vol. 25, n°4, p. 447-453. Disponible sur : < <https://doi.org/10.1038/nbt1297> >

Palmer C. M., Alper H. S. « Expanding the Chemical Palette of Industrial Microbes: Metabolic Engineering for Type III PKS-Derived Polyketides ». Biotechnol. J. [En ligne]. 2019. Vol. 14, n°1, p. 1700463. Disponible sur : < <https://doi.org/10.1002/biot.201700463> >

Parfait R., Cocito C. « Lasting damage to bacterial ribosomes by reversibly bound virginiamycin M. ». Proc. Natl. Acad. Sci. [En ligne]. 1980. Vol. 77, n°9, p. 5492-5496. Disponible sur : < <https://doi.org/10.1073/pnas.77.9.5492> >

Parke J. L., Gurian-Sherman D. « Diversity of the *Burkholderia Cepacia* Complex and Implications for Risk Assessment of Biological Control Strains ». Annu. Rev. Phytopathol. [En ligne]. 2001. Vol. 39, n°1, p. 225-258. Disponible sur : < <https://doi.org/10.1146/annurev.phyto.39.1.225> >

Parmeggiani A., Nissen P. « Elongation factor Tu-targeted antibiotics: Four different structures, two mechanisms of action ». FEBS Lett. [En ligne]. 2006. Vol. 580, n°19, p. 4576-4581. Disponible sur : < <https://doi.org/10.1016/j.febslet.2006.07.039> >

Piel J. « Biosynthesis of polyketides by trans-AT polyketide synthases ». Nat. Prod. Rep. [En ligne]. 2010. Vol. 27, n°7, p. 996-1047. Disponible sur : < <https://doi.org/10.1039/b816430b> >

Piel J. « A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles ». Proc. Natl. Acad. Sci. 2002. Vol. 99, n°22, p. 14002–14007. Disponible sur : < <https://doi.org/10.1073/pnas.222481399> >

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 [En ligne]. 2007. Vol. 393, n°1-2, p. 31-42. Disponible sur : < <https://doi.org/10.1016/j.gene.2006.12.035> >

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. [En ligne]. 2005. Vol. 280, n°8, p. 6399-6408. Disponible sur : < <https://doi.org/10.1074/jbc.M409814200> >

Ranganathan A., Timoney M., Bycroft M., Cortés J., Thomas I. P., Wilkinson B., Kellenberger L., Hanefeld U., Galloway I. S., Staunton J., Leadlay P. F. « Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: a route to simple statin analogues ». Chem. Biol. 1999. Vol. 6, n°10, p. 731-741. Disponible sur : < [https://doi.org/10.1016/S1074-5521\(00\)80020-4](https://doi.org/10.1016/S1074-5521(00)80020-4) >

Reid R., Piagentini M., Rodriguez E., Ashley G., Viswanathan N., Carney J., Santi D. V., Hutchinson C. R., McDaniel R. « A Model of Structure and Catalysis for Ketoreductase Domains in Modular Polyketide Synthases ». Biochemistry [En ligne]. 2003. Vol. 42, n°1, p. 72-79. Disponible sur : < <https://doi.org/10.1021/bi0268706> >

Richter C. D., Nietlispach D., Broadhurst R. W., Weissman K. J. « Multienzyme docking in hybrid megasynthetases ». Nat. Chem. Biol. [En ligne]. 2008. Vol. 4, n°1, p. 75-81. Disponible sur : < <https://doi.org/10.1038/nchembio.2007.61> >

Rix U., Fischer C., Remsing L. L., Rohr J. « Modification of post-PKS tailoring steps through combinatorial biosynthesis ». Nat. Prod. Rep. [En ligne]. 2002. Vol. 19, n°5, p. 542-580. Disponible sur : < <https://doi.org/10.1039/b103920m> >

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. [En ligne]. 2001. Vol. 8, n°5, p. 475-485. Disponible sur : < [https://doi.org/10.1016/S1074-5521\(01\)00024-2](https://doi.org/10.1016/S1074-5521(01)00024-2) >

Rudd B. A. M., Hopwood D. A. « Genetics of Actinorhodin Biosynthesis by *Streptomyces coelicolor* A3(2) ». J. Gen. Microbiol. [En ligne]. 1979. Vol. 114, n°1, p. 35-43. Disponible sur : < <https://doi.org/10.1099/00221287-114-1-35> >

Scheffler R. J., Colmer S., Tynan H., Demain A. L., Gullo V. P. « Antimicrobials, drug discovery, and genome mining ». *Appl. Microbiol. Biotechnol.* [En ligne]. 2013. Vol. 97, n°3, p. 969-978. Disponible sur : < <https://doi.org/10.1007/s00253-012-4609-8> >

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.* [En ligne]. 2008. Vol. 6, n°18, p. 3315. Disponible sur : < <https://doi.org/10.1039/b807914e> >

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* [En ligne]. 2001. Vol. 413, n°6858, p. 814-821. Disponible sur : < <https://doi.org/10.1038/35101544> >

Scotti C., Piatti M., Cuzzonia A., Perania P., Tognonib A., Grandib G., Galizzi A., Albertini A. M. « A *Bacillus subtilis* large ORF coding for a polypeptide highly similar to polyketide synthases ». *Gene*. 1993. Vol. 130, n°1, p. 65-71. Disponible sur : < [https://doi.org/10.1016/0378-1119\(93\)90347-6](https://doi.org/10.1016/0378-1119(93)90347-6) >

Seyedsayamdost M. R., Chandler J. R., Blodgett J. A. V., Lima P. S., Duerkop B. A., Oinuma K.-I., Greenberg E. P., Clardy J. « Quorum-Sensing-Regulated Bactobolin Production by *Burkholderia thailandensis* E264 ». *Org. Lett.* [En ligne]. 2010. Vol. 12, n°4, p. 716-719. Disponible sur : < <https://doi.org/10.1021/ol902751x> >

Shen B. « Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms ». *Curr. Opin. Chem. Biol.* [En ligne]. 2003. Vol. 7, n°2, p. 285-295. Disponible sur : < [https://doi.org/10.1016/S1367-5931\(03\)00020-6](https://doi.org/10.1016/S1367-5931(03)00020-6) >

Sillitoe I., Lewis T. E., Cuff A., Das S., Ashford P., Dawson N. L., Furnham N., Laskowski R. A., Lee D., Lees J. G., Lehtinen S., Studer R. A., Thornton J., Orengo C. A. « CATH: comprehensive structural and functional annotations for genome sequences ». *Nucleic Acids Res.* [En ligne]. 2015. Vol. 43, n°D1, p. D376-D381. Disponible sur : < <https://doi.org/10.1093/nar/gku947> >

Simunovic V., Gherardini F. C., Shimkets L. J. « Membrane Localization of Motility, Signaling, and Polyketide Synthetase Proteins in *Myxococcus xanthus* ». *J. Bacteriol.* [En ligne]. 2003. Vol. 185, n°17, p. 5066-5075. Disponible sur : < <https://doi.org/10.1128/JB.185.17.5066-5075.2003> >

- Squire C. M., Goss R. J. M., Hong H., Leadlay P. F., Staunton J. « Catalytically Active Tetramodular 6-Deoxyerythronolide B Synthase Fusion Proteins ». *ChemBioChem* [En ligne]. 2003. Vol. 4, n°11, p. 1225-1228. Disponible sur : < <https://doi.org/10.1002/cbic.200300671> >
- Staunton J., Caffrey P., Aparicio J. F., Roberts G. A., Bethell S. S., Leadlay P. F. « Evidence for a double-helical structure for modular polyketide synthases ». *Nat. Struct. Biol.* 1996. Vol. 3, n°2, p. 188-192. Disponible sur : < <https://doi.org/10.1038/nsb0296-188> >
- Staunton J., Weissman K. J. « Polyketide biosynthesis: a millennium review ». *Nat. Prod. Rep.* [En ligne]. 2001. Vol. 18, n°4, p. 380-416. Disponible sur : < <https://doi.org/10.1039/a909079g> >
- Stinear T. P., Mve-Obiang A., Small P. L., Frigui W., Pryor M. J., Brosch R., Jenkin G. A., Johnson P. D., Davies J. K., Lee R. E. « Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans* ». *Proc. Natl. Acad. Sci. U. S. A.* 2004. Vol. 101, n°5, p. 1345–1349. Disponible sur : < <https://doi.org/10.1073/pnas.0305877101> >
- 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.* [En ligne]. 2007. Vol. 104, n°1, p. 305-310. Disponible sur : < <https://doi.org/10.1073/pnas.0609073103> >
- 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* [En ligne]. 2000. Vol. 246, n°1-2, p. 123-131. Disponible sur : < [https://doi.org/10.1016/S0378-1119\(00\)00060-3](https://doi.org/10.1016/S0378-1119(00)00060-3) >
- Tacconelli E., Magrini N. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. [s.l.] : World Health Organization, 2017.
- Tan D. S. « Diversity-oriented synthesis: exploring the intersections between chemistry and biology ». *Nat. Chem. Biol.* 2005. Vol. 1, n°2, p. 11. Disponible sur : < <https://doi.org/10.1038/nchembio0705-74> >
- Tang L., Fu H., McDaniel R. « Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases ». *Chem. Biol.* 2000. Vol. 7, n°2, p. 77–84. Disponible sur : < [https://doi.org/10.1016/S1074-5521\(00\)00073-9](https://doi.org/10.1016/S1074-5521(00)00073-9) >

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.* [En ligne]. 2007. Vol. 14, n°8, p. 931-943. Disponible sur : < <https://doi.org/10.1016/j.chembiol.2007.07.012> >

Tang Y., Kim C.-Y., Mathews I. I., Cane D. E., Khosla C. « The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase ». *Proc. Natl. Acad. Sci.* 2006. Vol. 103, n°30, p. 11124–11129. Disponible sur : < <https://doi.org/10.1073/pnas.0601924103> >

Tang Y., Tsai S.-C., Khosla C. « Polyketide Chain Length Control by Chain Length Factor ». *J. Am. Chem. Soc.* [En ligne]. 2003. Vol. 125, n°42, p. 12708-12709. Disponible sur : < <https://doi.org/10.1021/ja0378759> >

Tatsuno S., Arakawa K., Kinashi H. « Analysis of Modular-iterative Mixed Biosynthesis of Lankacidin by Heterologous Expression and Gene Fusion ». *J. Antibiot.* (Tokyo) [En ligne]. 2007. Vol. 60, n°11, p. 700-708. Disponible sur : < <https://doi.org/10.1038/ja.2007.90> >

Tatsuno S., Arakawa K., Kinashi H. « Extensive Mutational Analysis of Modular-Iterative Mixed Polyketide Biosynthesis of Lankacidin in *Streptomyces rochei* ». *Biosci. Biotechnol. Biochem.* [En ligne]. 2009. Vol. 73, n°12, p. 2712-2719. Disponible sur : < <https://doi.org/10.1271/bbb.90591> >

Thattai M., Burak Y., Shraiman B. I. « The origins of specificity in polyketide synthase protein interactions ». *PLoS Comput. Biol.* 2007. Vol. 3, n°9, p. e186. Disponible sur : < <https://doi.org/10.1371/journal.pcbi.0030186> >

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.* [En ligne]. 2002. Vol. 9, n°7, p. 781-787. Disponible sur : < [https://doi.org/10.1016/S1074-5521\(02\)00164-3](https://doi.org/10.1016/S1074-5521(02)00164-3) >

Till M., Race P. R. « Progress challenges and opportunities for the re-engineering of trans-AT polyketide synthases ». *Biotechnol. Lett.* [En ligne]. 2014. Vol. 36, n°5, p. 877-888. Disponible sur : < <https://doi.org/10.1007/s10529-013-1449-2> >

Traitcheva N., Jenke-Kodama H., He J., Dittmann E., Hertweck C. « Non-Colinear Polyketide Biosynthesis in the Aureothin and Neoareothin Pathways: An Evolutionary Perspective ». *ChemBioChem* [En ligne]. 2007. Vol. 8, n°15, p. 1841-1849. Disponible sur : < <https://doi.org/10.1002/cbic.200700309> >

Tsai S.-C., Miercke L. J. W., Krucinski J., Gokhale R., Chen J. C.-H., 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.* [En ligne]. 2001. Vol. 98, n°26, p. 14808-14813. Disponible sur : < <https://doi.org/10.1073/pnas.011399198> >

Tsuji S. Y., Cane D. E., Khosla C. « Selective Protein-Protein Interactions Direct Channeling of Intermediates between Polyketide Synthase Modules ». *Biochemistry* [En ligne]. 2001. Vol. 40, n°8, p. 2326-2331. Disponible sur : < <https://doi.org/10.1021/bi002463n> >

Uversky V. N., Dunker A. K. « The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure ». *F1000 Biol. Rep.* 2013. Vol. 5. Disponible sur : < <https://doi.org/10.3410/B5-1> >

Verdine G. L. « The combinatorial chemistry of nature ». *Nature* [En ligne]. 1996. Vol. 384, n°6604 Suppl, p. 11-13. Disponible sur : < <https://doi.org/10.1038/384011a0> >

Vial L., Lepine F., Milot S., Groleau M.-C., Dekimpe V., Woods D. E., Deziel E. « *Burkholderia pseudomallei*, *B. thailandensis*, and *B. ambifaria* Produce 4-Hydroxy-2-Alkylquinoline Analogues with a Methyl Group at the 3 Position That Is Required for Quorum-Sensing Regulation ». *J. Bacteriol.* [En ligne]. 2008. Vol. 190, n°15, p. 5339-5352. Disponible sur : < <https://doi.org/10.1128/JB.00400-08> >

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.* [En ligne]. 2008. Vol. 26, n°1, p. 131-146. Disponible sur : < <https://doi.org/10.1080/07391102.2008.10507230> >

Walsh C. T., Chen H., Keating T. A., Hubbard B. K., Losey H. C., Luo L., Marshall C. G., Miller D. A., Patel H. M. « Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines ». *Curr. Opin. Chem. Biol.* [En ligne]. 2001. Vol. 5, n°5, p. 525-534. Disponible sur : < [https://doi.org/10.1016/S1367-5931\(00\)00235-0](https://doi.org/10.1016/S1367-5931(00)00235-0) >

Walsh C. T., Wencewicz T. A. « Prospects for new antibiotics: a molecule-centered perspective ». *J. Antibiot. (Tokyo)* [En ligne]. 2014. Vol. 67, n°1, p. 7-22. Disponible sur : < <https://doi.org/10.1038/ja.2013.49> >

Watve M., Tickoo R., Jog M., Bhole B. « How many antibiotics are produced by the genus *Streptomyces* ? ». Arch. Microbiol. [En ligne]. 2001. Vol. 176, n°5, p. 386-390. Disponible sur : < <https://doi.org/10.1007/s002030100345> >

Weber T., Laiple K. J., Pross E. K., Textor A., Grond S., Welzel K., Pelzer S., Vente A., Wohlleben W. « Molecular Analysis of the Kirromycin Biosynthetic Gene Cluster Revealed β -Alanine as Precursor of the Pyridone Moiety ». Chem. Biol. [En ligne]. 2008. Vol. 15, n°2, p. 175-188. Disponible sur : < <https://doi.org/10.1016/j.chembiol.2007.12.009> >

Weissman K. J. « Chapter 1 Introduction to Polyketide Biosynthesis ». In : Methods Enzymol. [En ligne]. [s.l.] : Elsevier, 2009. p. 3-16. Disponible sur : < [https://doi.org/10.1016/S0076-6879\(09\)04601-1](https://doi.org/10.1016/S0076-6879(09)04601-1) >

Weissman K. J. « The structural biology of biosynthetic megaenzymes ». Nat. Chem. Biol. [En ligne]. 2015. Vol. 11, n°9, p. 660-670. Disponible sur : < <https://doi.org/10.1038/nchembio.1883> >

Weissman K. J. « Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology ». Nat. Prod. Rep. [En ligne]. 2016. Vol. 33, n°2, p. 203-230. Disponible sur : < <https://doi.org/10.1039/C5NP00109A> >

Weissman K. J. « Mutasynthesis – uniting chemistry and genetics for drug discovery ». Trends Biotechnol. [En ligne]. 2007. Vol. 25, n°4, p. 139-142. Disponible sur : < <https://doi.org/10.1016/j.tibtech.2007.02.004> >

Weissman K. J. « Uncovering the structures of modular polyketide synthases ». Nat. Prod. Rep. [En ligne]. 2017. Vol. 34, n°8, p. 1035. Disponible sur : < <https://doi.org/10.1039/c7np90027a> >

Weissman K. J. « Single Amino Acid Substitutions Alter the Efficiency of Docking in Modular Polyketide Biosynthesis ». ChemBioChem [En ligne]. 2006a. Vol. 7, n°9, p. 1334-1342. Disponible sur : < <https://doi.org/10.1002/cbic.200600185> >

Weissman K. J. « The Structural Basis for Docking in Modular Polyketide Biosynthesis ». ChemBioChem [En ligne]. 2006b. Vol. 7, n°3, p. 485-494. Disponible sur : < <https://doi.org/10.1002/cbic.200500435> >

Weissman K. J., Leadlay P. F. « Combinatorial biosynthesis of reduced polyketides ». Nat. Rev. Microbiol. [En ligne]. 2005. Vol. 3, n°12, p. 925-936. Disponible sur : < <https://doi.org/10.1038/nrmicro1287> >

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* [En ligne]. 2014. Vol. 510, n°7506, p. 560-564. Disponible sur : < <https://doi.org/10.1038/nature13409> >

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.* [En ligne]. 2013. Vol. 20, n°11, p. 1340-1351. Disponible sur : < <https://doi.org/10.1016/j.chembiol.2013.09.015> >

Williamson M. P., Potts J. R. « Intrinsically disordered proteins: administration not executive ». *Biochem. Soc. Trans.* [En ligne]. 2012. Vol. 40, n°5, p. 945-949. Disponible sur : < <https://doi.org/10.1042/BST20120188> >

Winn M., Fyans J. K., Zhuo Y., Micklefield J. « Recent advances in engineering nonribosomal peptide assembly lines ». *Nat. Prod. Rep.* [En ligne]. 2016. Vol. 33, n°2, p. 317-347. Disponible sur : < <https://doi.org/10.1039/C5NP00099H> >

Wong F. T., Chen A. Y., Cane D. E., Khosla C. « Protein–Protein Recognition between Acyltransferases and Acyl Carrier Proteins in Multimodular Polyketide Synthases ». *Biochemistry* [En ligne]. 2010. Vol. 49, n°1, p. 95-102. Disponible sur : < <https://doi.org/10.1021/bi901826g> >

Wu L.-L., Xiang Y., Yang D.-C., Guan Z., He Y.-H. « Biocatalytic asymmetric Mannich reaction of ketimines using wheat germ lipase ». *Catal. Sci. Technol.* [En ligne]. 2016. Vol. 6, n°11, p. 3963-3970. Disponible sur : < <https://doi.org/10.1039/C5CY01923K> >

Wong F. T., Jin X., Mathews I. I., Cane D. E., Khosla C. « Structure and Mechanism of the trans-Acting Acyltransferase from the Disorazole Synthase ». *Biochemistry* [En ligne]. 2011. Vol. 50, n°30, p. 6539-6548. Disponible sur : < <https://doi.org/10.1021/bi200632j> >

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* [En ligne]. 2002. Vol. 41, n°15, p. 5056-5066. Disponible sur : < <https://doi.org/10.1021/bi012086u> >

Wu N., Tsuji S. Y., Cane D. E., Khosla C. « Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules ». *J. Am. Chem. Soc.* 2001. Vol. 123, n°27, p. 6465-6474. Disponible sur : < <https://doi.org/10.1021/ja010219t> >

Xu J. « Identification of tailoring genes involved in the modification of the polyketide backbone of rifamycin B by *Amycolatopsis mediterranei* S699 ». Microbiology [En ligne]. 2005. Vol. 151, n°8, p. 2515-2528. Disponible sur : < <https://doi.org/10.1099/mic.0.28138-0> >

Xu Y., Zhou T., Zhang S., Espinosa-Artiles P., Wang L., Zhang W., Lin M., Gunatilaka A. A. L., Zhan J., Molnar I. « Diversity-oriented combinatorial biosynthesis of benzenediol lactone scaffolds by subunit shuffling of fungal polyketide synthases ». Proc. Natl. Acad. Sci. [En ligne]. 2014. Vol. 111, n°34, p. 12354-12359. Disponible sur : < <https://doi.org/10.1073/pnas.1406999111> >

Yamamoto S., He Y., Arakawa K., Kinashi H. « Gamma-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. [En ligne]. 2008. Vol. 190, n°4, p. 1308-1316. Disponible sur : < <https://doi.org/10.1128/JB.01383-07> >

Yu H.-N., Liu X.-Y., Gao S., Sun B., Zheng H.-B., Ji M., Cheng A.-X., Lou H.-X. « Structural and biochemical characterization of the plant type III polyketide synthases of the liverwort *Marchantia paleacea* ». Plant Physiol. Biochem. [En ligne]. 2018. Vol. 125, p. 95-105. Disponible sur : < <https://doi.org/10.1016/j.plaphy.2018.01.030> >

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. [En ligne]. 2016. Vol. 11, n°9, p. 2466-2474. Disponible sur : < <https://doi.org/10.1021/acscchembio.6b00345> >

Zhang Q., Pang B., Ding W., Liu W. « Aromatic Polyketides Produced by Bacterial Iterative Type I Polyketide Synthases ». ACS Catal. [En ligne]. 2013. Vol. 3, n°7, p. 1439-1447. Disponible sur : < <https://doi.org/10.1021/cs400211x> >

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. [En ligne]. 2012. Vol. 8, n°7, p. 615-621. Disponible sur : < <https://doi.org/10.1038/nchembio.964> >

Les polycétides sont des métabolites secondaires produits par divers organismes et présentant un large spectre d'activité thérapeutique. L'organisation modulaire des enzymes responsables de leur synthèse, les polycétide synthases (PKS), en font des cibles attrayantes pour la biologie synthétique visant l'obtention de nouvelles structures de polycétides. L'une des stratégies les plus prometteuses à ce jour consiste à échanger des sous-unités entières entre différents systèmes PKS. Cependant, le succès de cette stratégie dépend essentiellement de la compréhension et de l'utilisation des domaines de *docking*. Ces petits éléments structuraux, situés aux extrémités C- et N-terminales des sous-unités de PKS, sont responsables de leur bon appariement et donc du transfert correct de l'intermédiaire polycétidique. Pour approfondir nos connaissances sur les DD, nous avons étudié plusieurs interfaces entre sous-unités de PKS *trans*-AT et *cis*-AT. Ces travaux ont notamment conduit à l'identification de la première classe de DD chez les PKS *trans*-AT, et nous avons pu caractériser une interface complète, formée entre deux sous-unités consécutives au sein de la PKS virginiamycine. De plus, nous avons montré qu'au moins un des DD de paires appariées est souvent une région intrinsèquement désordonnée (IDR), ce mode de reconnaissance permet d'assurer une interaction spécifique couplée à une affinité moyenne. En effet, chez l'enacyloxine synthase, une PKS hybride *cis*-AT/*trans*-AT, les six interfaces de *docking* sont médiées par une IDR C-terminal. En outre, nous avons démontré que ce système présentait plusieurs classes structurales de DD, mais que des variations du code électrostatique au sein d'une classe individuelle pouvaient également être utilisées pour garantir la spécificité. Ensemble, ces résultats fournissent des indications importantes pour les futures tentatives d'utilisation des DD en ingénierie des sous-unités de PKS.

Une autre stratégie innovante permettant d'obtenir de nouveaux dérivés polycétidiques, consiste à utiliser des enzymes post-PKS. Ces dernières modifient de manière extensive l'intermédiaire et lui confèrent bien souvent son activité biologique. Dans ce contexte, nous avons étudié LkcE, une enzyme bi-fonctionnelle qui catalyse une réaction rare d'oxydation d'amide, suivie d'une réaction de Mannich intramoléculaire. Nous avons résolu quatre structures cristallographiques de l'enzyme et ces données couplées aux études cinétiques, nous ont permis de proposer un mécanisme catalytique détaillé pour LkcE, impliquant un changement conformationnel à grande échelle de l'enzyme pour amener le substrat à une conformation de pré-cyclisation. De plus, nous avons démontré que LkcE possède une certaine tolérance vis-à-vis de la structure de son substrat, suggérant son utilité comme catalyseur général de cyclisation/ligation en biologie synthétique et synthèse chimique.

Mots clés : polycétide, polycétide synthases, domaines de *docking*, enzymes post-PKS, RMN, SAXS.

Complex polyketides are secondary metabolites which are produced by a range of different organisms, and which present a broad spectrum of therapeutic activity. The modular organization of the enzymes responsible for their synthesis, the polyketide synthases (PKS), makes them attractive targets for synthetic biology aimed at obtaining new polyketide structures. One of the most promising strategies to date consists in swapping of whole sub-units between different PKS systems. However, the success of this strategy critically depends on understanding and exploiting 'docking domains' – the protein sequences at the C- and N-terminal extremities of the subunits which are responsible for correctly ordering the polypeptides, and therefore for faithful chain transfer. To increase our knowledge of DDs, we investigated several interfaces in both *trans*-AT and *cis*-AT PKSs. This work led notably to the identification of the first family of DDs from *trans*-AT PKSs, and we were further able to characterize a complete interface formed between two consecutive subunits within the virginiamycin PKS. In addition, we showed that at least one DD of matched pairs is often an intrinsically disordered region (IDR), as this type of interaction motif allows for specific but medium affinity contacts. Indeed, in the enacyloxin hybrid *cis*-AT/*trans*-AT PKS which we also investigated extensively, docking at every interface is mediated by a C-terminal IDR. In addition, we demonstrated that multiple structural classes of DD are present within the system, but that variations of the electrostatic 'code' within an individual structural class can also be used to ensure specificity. Taken together, these results provide important guidelines for future attempts to deploy DDs in subunit engineering.

Another attractive target for synthetic biology are the so-called 'post-PKS' enzymes, which chemically decorate the initially-formed structure, and are often essential for their bioactivity. In this context, we studied LkcE, a bi-functional enzyme that catalyzes a rare amide oxidation followed by an intramolecular Mannich reaction to yield the lankacidin macrocycle – both to understand its unusual mechanism and to evaluate its suitability as a general polyketide modifying enzyme. We solved four crystal structures of the enzyme, and characterized it kinetically. Together, our data allowed us to propose a detailed catalytic mechanism for LkcE, involving a large-scale conformational change of the enzyme to bring the substrate into a cyclisation-ready state. Moreover, we showed that LkcE displays a certain tolerance toward its substrate structures, suggesting its usefulness as a general catalyst for cyclisation/ligation reaction in synthetic biology and chemical synthesis.

Key words: polyketide, polyketide synthases, docking domains, post-PKS enzymes, NMR, SAXS.