THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*

Discipline ou spécialité :
BIOLOGIE STRUCTURALE ET FONCTIONNELLE

Présentée et soutenue par
Alexandre FAILLE

Le 27 Novembre *2013*

Titre :

Etude structurale et fonctionnelle de la polykétide synthase PpsC et de son activateur PptT chez Mycobacterium tuberculosis

---

JURY

Timm MAIER, Professor, Biozentrum, Bâle	*Rapporteur*
Pedro ALZARI, Directeur de Recherche, Institut Pasteur, Paris	*Rapporteur*
Thierry VERNET, Directeur de Recherche, CEA, Grenoble	*Examinateur*
Eric CLOTTES, Professeur, Université Toulouse III-Paul Sabatier	*Examinateur*

Ecole doctorale : *Biologie-Santé, Biotechnologies*

Unité de recherche : *Institut de Pharmacologie et de Biologie Structurale, CNRS UMR5089*

Directeur(s) de Thèse : Jean-Denis PEDELACQ & Lionel MOUREY

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*

Discipline ou spécialité :
BIOLOGIE STRUCTURALE ET FONCTIONNELLE

Présentée et soutenue par
Alexandre FAILLE

Le 27 Novembre *2013*

Titre :

Etude structurale et fonctionnelle de la polykétide synthase PpsC et de son activateur PptT chez Mycobacterium tuberculosis

---

JURY

Timm MAIER, Professor, Biozentrum, Bâle	*Rapporteur*
Pedro ALZARI, Directeur de Recherche, Institut Pasteur, Paris	*Rapporteur*
Thierry VERNET, Directeur de Recherche, CEA, Grenoble	*Examinateur*
Eric CLOTTES, Professeur, Université Toulouse III-Paul Sabatier	*Examinateur*

Ecole doctorale : *Biologie-Santé, Biotechnologies*

Unité de recherche : *Institut de Pharmacologie et de Biologie Structurale, CNRS UMR5089*

Directeur(s) de Thèse : Jean-Denis PEDELACQ & Lionel MOUREY

Je tiens tout d'abord à adresser mes remerciements les plus sincères aux membres du jury :

Dr Timm Maier,

For accepting to come in Toulouse to be the reviewer of my thesis was a very kind gesture. Thank you for your interest in my work and for the quality of the discussion we had during your stay here. Danke Schön.

Dr Pedro Alzari,

Merci d'avoir accepté d'être rapporteur de cette thèse, pour la curiosité et la considération apportée à mon travail et le souci d'améliorer la valorisation de celui-ci.

Dr Thierry Vernet,

Merci d'avoir participé à mon jury de thèse en qualité d'examinateur. Merci pour votre curiosité ainsi que pour votre scientifique et technique sur le sujet.

Enfin, merci cher Président, d'avoir accepté de mener à bien ce fabuleux cérémonial. Merci **Eric** pour ton implication, tes conseils et tes corrections de fautes d'orthographe (notamment en couverture…). Sois assuré de ma profonde reconnaissance.

Après ces 5 années passés à vos côtés, voire 6 au moment où j'écris ces quelques lignes, ou plutôt soyons précis 2190 jours, enfin 2190 après-midi… Bref, après ces merveilleuses 6 années dans le "Groupe de Biophysique Structurale" (qui a changé de nom depuis mais je ne m'en rappelle jamais), mes premiers mots (après le jury bien sûr) se devaient d'être pour vous. Grâce à vous chers cristallographes et autres biochimistes, anciens et nouveaux collègues du GBS, ces années furent des années de bonheur. D'apprentissage aussi, "tu séquenceras toujours tes plasmides avant de cristalliser la mauvaise protéine", ils auraient dû commencer par ça. J'ai appris à cristalliser la même protéine dans une douzaine de conditions différentes, j'ai appris à déboucher des colonnes de gel filtration sans que Valérie s'en rende compte (j'ai fait ça bien Val c'est promis), j'ai appris que filtrer du PEG 8000 40% à la seringue pouvait me permettre d'aller fumer une clope avec Caro, et de lui refiler ma

blouse à laver tant que j'y étais. J'ai même appris à déshydrater une déshydratase (sans trop de succès on fait de la recherche hein) c'est pour vous dire !

Mais je n'ai pas appris que de la science, après tout si vous en voulez, de la science, je ne me suis pas cassé la tête (et celle de mes directeurs de thèse) à écrire les quelques centaines de pages qui suivent pour rien. J'ai aussi appris la simplicité, dans le bon sens du terme, une qualité rare chez les surdiplômés que nous sommes, désormais. J'ai aussi appris l'esprit d'équipe, la collaboration, le nettoyage de paillasse, la quasi ponctualité, et que la patience est une vertu qui se mange froide, surtout en réunion d'équipe, comme les sandwichs d'ailleurs. J'ai aussi appris à superviser des étudiants, les faire rire, mais aussi et surtout les faire pleurer. Je leur souhaite d'ailleurs un meilleur avenir. Enfin, si toutefois j'en doutais encore, j'ai appris grâce à vous que travailler dans la recherche, à défaut de rapporter de l'argent, est une excellente manière de s'épanouir professionnellement et bien au-delà...

Je ne vous oublierai jamais et je ne peux que me souhaiter de retrouver un jour, si cela est possible, une équipe dans laquelle il fait si bon vivre et où le temps devient une notion plus que relative, 2190 après-midi rendez-vous compte... moi je n'y arrive toujours pas.

J'aimerais aussi exprimer ma gratitude aux membres un peu particuliers de cette belle équipe, aux profs de biochimie,

A mes profs devrais-je dire, puisque j'ai eu la chance de tous vous cotoyer du début à la fin de mes études. Vous êtes ceux qui m'avez donné envie de continuer en licence de biochimie, en partie je l'avoue, pour ne pas louper les hilarantes histoires de bouilleurs de cru russes d'Eric. Puis de continuer en master où j'ai été happé par le miracle de la cristallo, à moins que cela ne soit la passion contagieuse de Sam et Laurent, probablement les deux. Vous êtes aussi ceux qui m'avez convaincu de m'exercer au même métier que vous par la suite, sans regret aucun, ce fut une expérience formidable. Je ne sais pas ce que j'aurais bien pu finir par faire sans votre intervention dans ma scolarité, en tout cas je n'aurais certainement pas eu la chance de faire cette thèse dans cette équipe alors juste pour ça, Merci. Continuez de transmettre votre passion, elle créé et j'en suis la preuve, des vocations.

A mon chef et directeur de thèse, à toi Lionel,

Moi qui suis un peu fâché, par nature certainement, avec la hiérarchie, je dois dire que cette fois j'ai du mal à critiquer. Tu es parvenu à me comprendre, et tout en me donnant la liberté dont j'avais besoin, tu n'as jamais manqué de m'épauler. Tu m'as appris que l'on pouvait courir 24h d'affilée, enfin il parait, j'étais pas là, et comme dirait St Thomas “j'ai des doutes !”. De toute façon mes beaux-cousins ils te mettent à l'amande, ou plutôt à la pomme, je sais que t'aimes ça les pommes ! Bref, avoir à la fois un chef positif et encourageant, et un homme profondément gentil et dévoué comme tu es, est une vraie chance pour cette équipe. Je te souhaite sincèrement le meilleur pour la suite. Quant à moi ne t'en fais pas, car oui je sais que tu t'en fais toujours un peu, tu m'as donné tout ce qu'il fallait pour réussir. A très bientôt ou plutôt Bowie Ken et Alain Di !

Il fallait bien te trouver une place dans ces remerciements, ce n'est pas par souci de protocole, toi plus que n'importe qui sais ce que j'en pense du protocole... mais c'est qu'en ayant commencé par l'équipe, puis les amis et la famille, je ne sais toujours pas où parler de toi, donc ce sera ici.

A toi JD, mon collègue, mon directeur de thèse, mon correspondant américain, et surtout mon ami,

Je crois qu'on pourrait écrire un bouquin, ou faire un film plutôt, tant ce sont des images qui vont me rester éternellement. Un road movie serait bien adapté, ça commencerait par des tours de vélo à l'ESRF, des tours de voiture à l'ESRF, des courses d'avion, à l'ESRF aussi (si si rappelle toi). Heureusement il n'y a pas eu que l'ESRF, il y a eu la coupe du monde de foot dans le bureau, les courses de tank remplis d'azote aussi, “damn!”, les skype outre atlantique interminables, et le saquage de texte, ah qu'est-ce que t'aimais ça saquait ma prose.. Non ceci n'est pas une allusion mal placée... même si nous étions, sommes, et serons j'espère toujours très proches. Ces fameuses 6 années auraient certainement été beaucoup moins joyeuses sans toi, et en plus nous n'aurions pas pu inventer la nitrovivose ! J'ai certainement dû oublier de nos nombreuses anecdotes, je suis sûr que nous aurons largement le loisir de nous les remémorer, et d'en inventer d'autres par le futur.

A vous mes chers pairs, qui m’avez souvent aidé par vos précieux conseils scientifiques et autres,

Christophe, Alain, KT, Isabelle, Yannick, Christian, et plus généralement les membres des équipes Daffé et Guilhot.

A vous mes copains du labo, qui m'avez aussi aidé, scientifiquement bien sûr mais aussi à faire en sorte que le temps passé au labo soit encore plus agréable. La liste est un peu longue tant l'environnement à l'IPBS est amical, c'est un peu comme au jeu des 7 familles, j'espère n'oublier personne: chez les Parini je voudrais Romain, Claire, Anna et Vero; chez les Khamlichi je voudrais Claire, chez les Teissier je voudrais Caro, Greg, Mathilde et Vincent; chez les Milon je voudrais Pascal; chez les Daffé je voudrais Clément, Fabienne, Nathalie, Nawel, Annaik et Hedia; chez les Guilhot je voudrais Wlad, Ainhoa, Sophie, Pierre-Yves et Henar; chez les Neyrolles je voudrais Firmin, GC, Ania et Anthony; chez les Montsarrat je voudrais Bertrand; chez les Puzo je voudrais Emilyne; chez les Salomé je voudrais Cheikh.

Mention spéciale à mes copains de clope avec qui l'on refait le monde comme on le fait au bar mais sobrement cette fois: Sam, Pascal, Stéphane, Fabienne, Nathalie, Caro, et Sabine et son verre d'eau.

Et mention encore plus spéciale à mon ami Wlad ainsi qu'à son adorable petite famille.

A vous mes amis, qui m'avez supporté, pendant que je parlais de science insatiablement, ou que je parlais beaucoup tout simplement,

Mes amis catalans avec qui j'ai appris (confirmé ?) que la fac c'était aussi (plus ?) sympa en dehors que sur ses bancs, que j'ai un peu perdu de vue, mais que je n'oublierai jamais: Fingo, Neymo, Adrien, Guilhem, Morgan et Boris. Mes autres amis catalans, Juju le magicien de l'acromio-epaulo-coudo contracture et Nat ma mangeuse de pain préférée. Mes amis désormais parisiens Sev la dresseuse de serpent, Hugues, et leur smala. Mes amis lotois Bilou et Tif chercheurs, de truffes eux, joyaux festoyeurs de village et joueurs de pétanque. Mes amis tarnais la Miche et Marion importateurs d'alcool aux parfums exotiques. Mes amis tarnais Bardou et Bouissou importateurs de "rhum" et "travailleurs". Mes amis néo-saint-léonains Marine et Tomtom, tous deux festoyeurs et lui pagayeur résistant au froid. Mon amie clermontoise et mon ami marocain Agathe et Youssef, joueurs de poker jusqu'au petit jour, maîtres du houmous traditionnel, et "travailleurs" eux aussi. Je vous passe

gracieusement mes amis du labo, dont certains déjà cités plus haut auraient très bien pu figurer dans ce paragraphe.

A ma belle-famille, que je cite puisque je suis un peu “Faillo” et que je ne veux pas me faire engueuler. Bien sûr, ce n’est pas la raison première pour laquelle je vous remercie, mais c’est bien pour votre soutien, nos discussions scientifiques enrichissantes sur les “microbes” et autres “saloperies”, et surtout pour les dons généreux et réguliers de produits du Roussillon, du Capcir et de l’Ariège.

A vous ma famille, mes parents, mon adorable neveu, ma sœur et mes tantes,

Merci pour votre soutien à toute épreuve. Pour être venu me ramasser après mes gamelles à moto. Pour m'avoir épaulé, financé, et surtout encouragé dans tous mes choix. Grâce à vous j'ai eu la liberté de faire de mon métier une passion, ou l'inverse, je ne sais plus trop. J’en suis conscient, c’est un luxe, en même temps vous m'avez toujours gâté depuis tout petit, j'en attendais donc pas moins. Je ne vous apprends rien si je vous dis que sans vous je ne serais pas là, quand bien même, sans vous je ne serais pas là. Je vous aime.

A mon ami, qui m'a appris, entre autres choses, que la Gersie n'était pas peuplé que de sauvages et de canards, à toi Rami. Tu m'as aussi appris qu'il existait un peu plus de 3 espèces de poissons, de serpents, d'arbres ou encore de planctons. Je ne cite certainement pas les plus importantes, mais à ton contact je n'ai jamais cessé d'apprendre de nouvelles choses. Bon, malgré les nombreuses saisons achevées, pour les roulettes à fifa on repassera… Merci aussi d'avoir équilibré mon canapé pendant notre année de master, afin que celui-ci ne s'affaisse pas que d'un côté.

A mon double, mon binôme de TP, mon ami, à toi Etienne. Moi non plus je ne remercierai jamais assez Boris de nous avoir fait galérer ce jour-là de 1ere année dans le bois de la fac. Que dire de tout ce que nous avons partagé depuis. Peut-être retiendra-t-on que je suis un peu rancunier quand même. Tu m'as appris le soutien et la loyauté infaillibles dont pouvaient faire preuve des amis. Tu m'as aussi appris qu'il existait une autre biologie que la biochimie, même si j'avoue, je ne suis pas encore complétement convaincu. Avec toi j'ai redéfini ma notion d'amitié à une hauteur que je n'avais jamais imaginée jusqu'alors. Enfin, on s'en est offert des cadeaux, des débilités sans nom du style ouvre-bouteilles phalliques, mais tu m'en as fait un que je

ne saurai jamais égaler car grâce à toi j'ai rencontré Charlotte, le plus beau des cadeaux.

A toi Charlotte, ma moitié, ma chérie, mon amour,

Et dire que je ne jurais que par les beuveries entre copains… Je n'oublierai jamais cette nuit d'hiver enneigée. Ce fut une révélation, ma vie a changé comme je ne pouvais m'y attendre, depuis je n'arrête pas de me demander comment je vivais avant. Tu m'as appris l'amour, à moi, un cartésien qui en a le prénom, certainement une prouesse dont toi seule était capable. Avec toi j'ai traversé la France, les continents et les océans… et les différents services de la clinique St Jean. Pour autant, même à St Orens, chaque jour qui passe, mon amour pour toi s'étend. Je t'aime plus que tout.

ABSTRACT

Polyketides are a remarkable class of compounds. In addition to exhibiting a great structural and functional diversity, they often possess drug activity that has been used for decades in many pharmaceutical fields. They are biosynthesized by highly complex machineries called polyketide synthases. Therefore, these enzymes represent a very promising tool for the production of compounds with either new or enhanced drug activity. *Mycobacterium tuberculosis* uses polyketides as virulence factors, and thus polyketide synthases also represent potential drug targets. Tuberculosis is the second cause of death due to a single infectious agent and is still considered as a global health concern. Nowadays, the need for new drugs targeting new pathways is critical.

PpsC is a type-I mono-modular polyketide synthase essential for the synthesis of the well-characterized phthiocerol dimycocerosates virulence factors in *Mycobacterium tuberculosis*. In an attempt to both help the search for new antituberculosis drugs, and decipher the complex mechanism used by polyketide synthases, we have undertaken the functional and structural study of PpsC using X-ray crystallography. Type-I polyketide synthases are made of several domains, each possessing a different catalytic activity. Due to their size and intrinsic flexibility these enzymes are difficult to characterize in their full-length form. We have then used the recently developed domain trapping method to identify soluble fragments representing each domain of PpsC. The study of these fragments has led to the structural determination of the acyltransferase domain, the dehydratase domain, and the nucleotide-binding subdomain of the enoylreductase domain of PpsC. In addition, we have deciphered the catalytic mechanism of the dehydratase domain, with the help of the structure of a substrate-enzyme complex and *in vitro* enzymatic tests.

Polyketide synthases need to be activated by the transfer of a prosthetic group onto their acyl carrier protein domain, catalyzed by 4'-phosphopantetheinyl transferases. In *Mycobacterium tuberculosis*, PptT is responsible for the activation of polyketide synthases and is thus essential to its virulence. Therefore, we were also interested in the functional and structural characterization of this promising drug

target. Using the split-GFP technology, we have designed a protocol for the production and purification of PptT, where its co-factors play an essential role for its stability both *in vivo* and *in vitro*. Activity of the enzyme was assessed by *in vitro* transfer of the prosthetic group onto the acyl carrier protein domain of PpsC. Finally, we have solved the three dimensional structure of PptT at 1.4 A resolution using X-ray crystallography, which will allow the structure-based design of inhibitors.

LIST OF ABBREVIATIONS

ACP	acyl carrier protein
AT	acyltransferase
CoA	coenzyme-A
Da	Dalton
DAT	diacyl trehaloses
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
DIM	phthiocerol dimycocerosates
DSF	differential scanning fluorimetry
ER	enoylreductase
FAS	fatty acid synthase
IPBS	institut de pharmacologie et de biologie structurale
IPTG	isopropyl-β-D-thiogalactopyranoside
KR	ketoreductase
KRc	catalytic ketoreductase
KRs	structural ketoreductase
KS	ketosynthase
Mtb	*Mycobacterium tuberculosis*
NRPS	non ribosomal peptide synthase
PAT	polyacyl trehaloses
PEG	polyethylene glycol
PKS	polyketide synthase
P-Pant	4'-phosphopantetheinyl
PPTase	4'-phosphopantetheinyl transferase
RMSD	root mean square deviation
SL	sulfolipids
TB	tuberculosis
WHO	world health organization

TABLE OF CONTENTS

I. INTRODUCTION

Polyketides make good drugs

Of the ~10,000 known polyketides, 1% possess drug activity

From standard pharmaceutical screens, the typical hit rate is <0.001%.

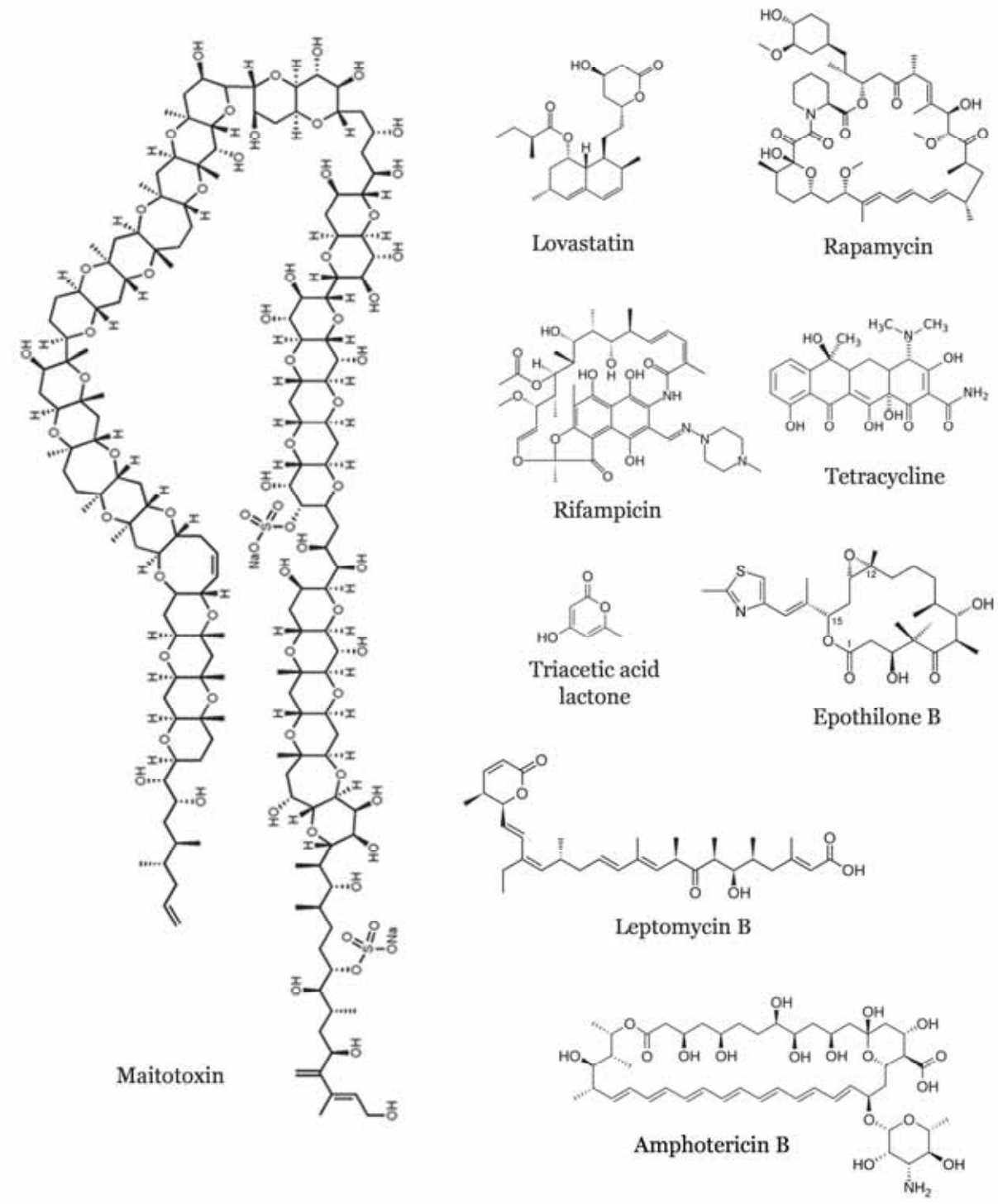


Figure 1. Structures of some polyketides illustrating the chemical diversity among this class of compounds.

A. POLYKETIDES

1. PHARMACEUTICAL INTEREST

Polyketides cover a large variety of structure and shape. They can adopt aliphatic, aromatic or macrolidic structures with multiple modes of folding and ring formation (**Figure 1**). They can be as small as the 6 carbons triacetic acid lactone or up to 164 carbons like the 3422 Da maitotoxin, they possess variable molecular properties and occupy a large chemical space. Polyketides have therefore found applications in numerous pharmaceutical fields as antibacterial, anticholesterolemic, immunosuppressant, antifungal, anti-cancer, antiparasitic, insecticide and probably other activities that remain to be characterized. The lovastatins are the most successful cardiovascular drugs of all time. Erythromycin has been in clinical usage for 53 years while its derivative azithromycin is one of the best-selling antibiotics in the world. Rifampicin is one of the most commonly used antibiotics to treat the second most common cause of death from an infectious disease, *i.e.* tuberculosis. The enediyines, especially calicheamicins, are among the most powerful anti-cancer agents yet discovered. By targeting the bacterial ribosome, tetracycline and its derivatives have been widely used as broad-spectrum antibiotics. There are numerous examples of pharmaceutical success stories concerning polyketides (**Table 1**). In fact, more than one third of natural products and derivative compounds approved as drugs between 2005 and 2007 were polyketides. As testimony to their importance, polyketide-derived pharmaceuticals comprise 20% of the top-selling drugs and annual sales routinely reach US$20 billion.

Polyketide	Example of commercial name	Drug Activity	Producing organism	Discovery
Tetracycline	Sumycin	Broad-spectrum antibiotic	*Streptomyces genus*	1945
Erythromycin	Ilosone	Broad-spectrum antibiotic	*Saccharopolyspora erythrae*	1949
Nystatin	Nystan	Antifungal	*Streptomyces noursei*	1950
Rifampicin	Rifadin	Broad-spectrum antibiotic	*Amycolatopsis rifamycininca*	1959
Monensin	Monensin	Antiobiotic / Antimalarial	*Streptomyces cinnamonensis*	1967
Rapamycin	Sirolimus	Antifungal / Immunosuppressant	*Streptomyces hygroscopicus*	1975
Lovastatin	Mevacor	Statin: lowers LDL cholesterol	*Aspergillus terreus*	1978
Avermectin	Ivomec	Broad-spectrum antiparasitic	*Streptomyces avermitilis*	1978
Azithromycin	Zithromax	Broad-spectrum antibiotic	*Saccharopolyspora erythrae*	1980
Spinosyns	Spinosad	Insecticide	*Saccharopolyspora spinosa*	1982
Tacrolimus	FK-506	Immunosuppressant	*Streptomyces tsukubaensis*	1984
Calicheamicin	Mylotarg	Anti-cancer	*Micromonospora echinospora*	1989

Table 1. Examples of famous polyketides.

2. A BIT OF HISTORY

In 1893 at London University, James Collie serendipitously discovered a remarkable set of synthetic reactions. While working on the elucidation of the structure of dehydroacetic acid by degradation chemistry [1], he discovered that boiling dehydroacetic acid with barium hydroxide yielded an aromatic compound, orcinol, as one of the reaction products. Collie showed that orcinol and related compounds did not fit into any of the previously described chemical classes, and therefore proposed a new chemical class named polyketides which contain the structure: CH_3-CO-$(CH_2$-CO$)_n$-X. Later, Collie proposed a rough mechanism for the synthesis of orcinol and related compounds, based solely on their structure, implying repetitive condensation or polymerization reactions [2].

However, Collie's hypotheses were met with indifference until the 1950s, when Arthur Birch came up with the theory that polyketides could be generated from acetate units by repeated condensation reactions. He applied his theory to the synthesis of the polyketide 6-methyl-salicylic acid (6-MSA) produced by the fungus *Penicillium patulum*. The linkage of four acetate units would be necessary to generate a first intermediate, followed by a sequence of aldol condensation, dehydration and enolization that would finally lead to the aromatic natural product. To test this hypothesis, he fed a polyketide-producing organism with a ^{14}C radiolabelled version of his proposed precursor. Then he isolated the generated products and subjected them to radioactivity measurement. The results were consistent with his predictive pattern of incorporation [3]. This new methodology inaugurated a great activity in the natural product community. Subsequent experiments established not only that many polyphenolic aromatic molecules were biosynthesized from acetate units but also that many non-aromatic compounds were also formed by further transformations of such products.

The results on natural product biosynthesis coincided with the development of nuclear magnetic resonance and mass spectrometry as powerful new techniques for the structural elucidation of complex molecules. A structural study that might take several years of intensive work using the decomposition methodology, might be solved in only a matter of weeks or even days by the new spectroscopic technologies.

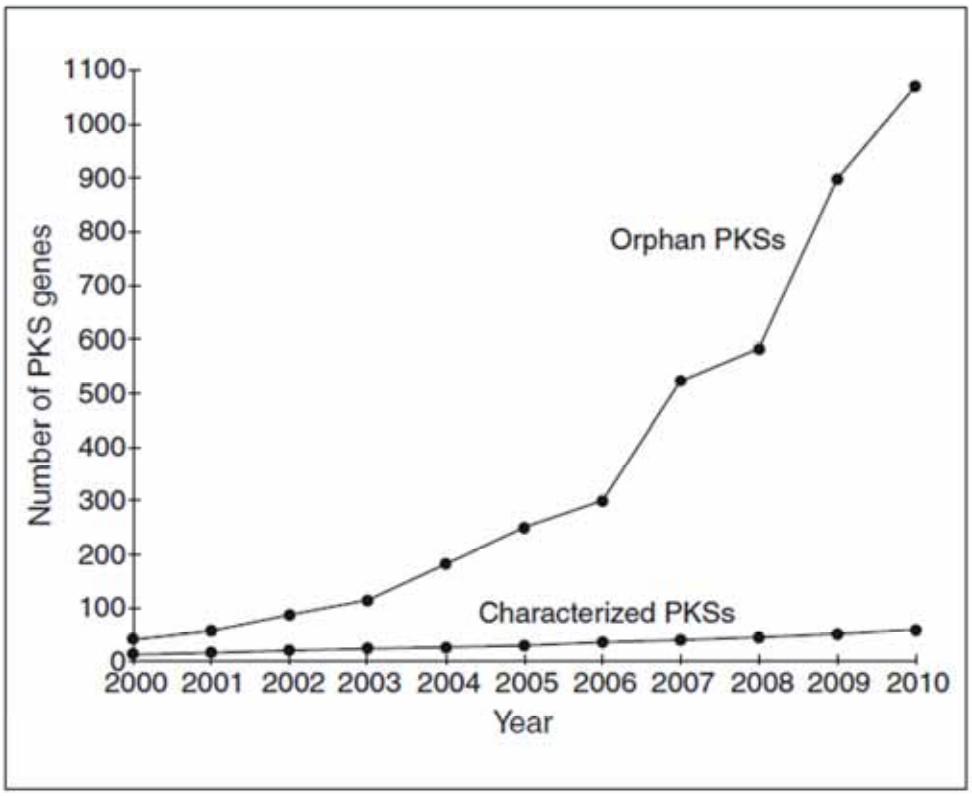


Figure 2. Distribution of *pks* genes between 2000 and 2010.

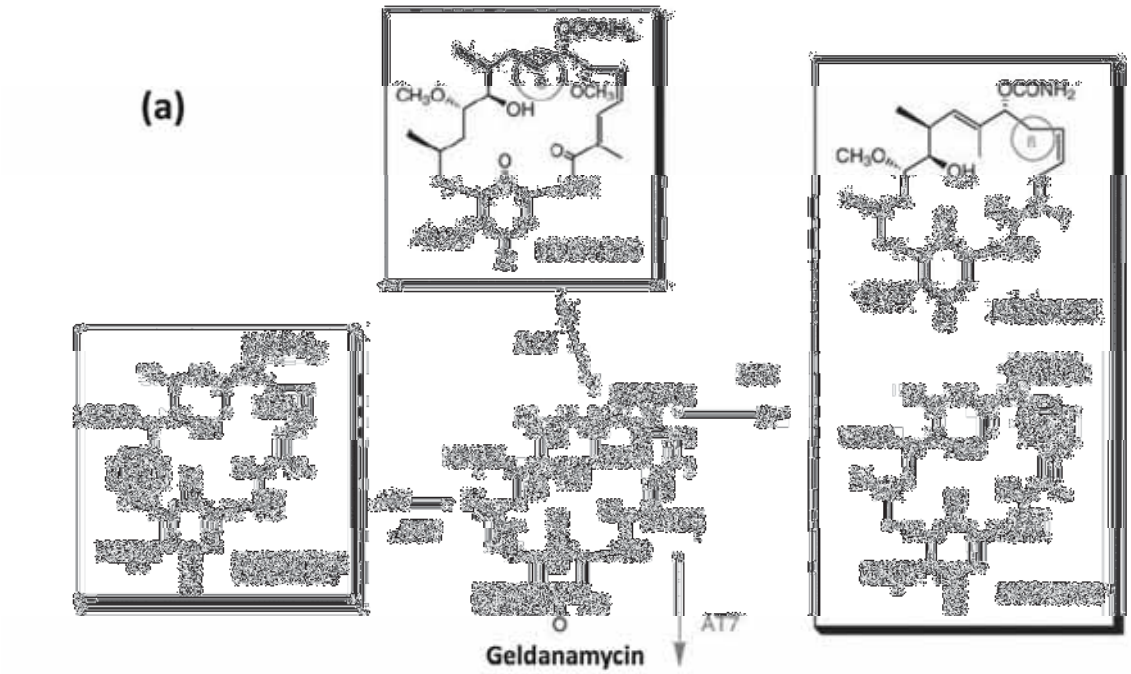
Comparison of the number of orphan PKSs and characterized PKSs. In contrast to characterized PKSs, no polyketide product has been attributed to orphan PKSs. The number of orphan PKSs has greatly increased with new genomic techniques in the early 21th century. **From Wong et al., 2012.**

These emerging technologies were applied to understand the biosynthesis of chemically well-characterized polyketides such as nargenicin [4] and erythromycin [5], and pointed to a biosynthetic route to the reduced aliphatic polyketide. In this, each successive addition of a building block is immediately followed by reductive modification of the newly-formed keto group (so-called 'processive' biosynthesis), as opposed to construction of a polyketone chain followed by reduction.

3. NEW CHALLENGES FOR POLYKETIDE DRUGS

Discovering analogues of widely used polyketides with improved pharmaceutical properties or totally new structures with novel properties has always been a challenge. Efforts in medicinal chemistry have focused on two main strategies, (1) the identification of new compounds from the biosphere or (2) the semi-synthetic modification of existing ones. The latter strategy has proven to be valuable for the development of today best-selling drugs such as azithromycin (erythromycin-derived), simvastatin (lovastatin-derived) or doxycyclin (tetracyclin-derived). However, these semi-synthetic drugs are restrained to the same mechanism of action than their parental compound and, in the case of antibiotics, are often subjected to previously-developed bacterial resistance.

Identification of new naturally occurring compounds has recently become a much simpler approach. Indeed, development of high-throughput sequencing techniques and whole genome sequencing allowed the mapping of genes coding the enzymes responsible for the biosynthesis of polyketides: the polyketide synthases (PKSs). The genomic era allowed researchers to accumulate hundreds of *pks* genes for which the resulting polyketide is to-be-assigned are now available (**Figure 2**). However, such a data-mining analyze is time-consuming and requires to precisely know the complex biosynthesis mechanism used by the concerned PKS in order to confidently attribute a structure to the polyketide.



(b)

Compound	Hsp90 K_d (nM)
Geldanamycin	670
17-AAG	1300
KOSN1631 (6-desmethoxy)	5000
KOSN-1630	3000
KOSN1859 (8-desmethyl)	5200
KOSN1877 (14-desmethyl)	660
KOSN1558	1000
KOSN1559	16

Figure 3. Structures and binding affinities of biosynthesized geldanamycin analogues.

(a) Genetically engineered geldanamycin analogs are enclosed in rectangles, their structure modifications compared to geldanamycin are encircled. Those modifications arose from exchange of gene parts coding the acyl-transferase activity (AT) responsible for the incorporation of specific building blocks during the biosynthesis. (b) Measured affinities of geldanamycin and analogs for Hsp90. **Modified from Patel et al., 2004**

Alternatives to the so-called semi-synthetic modification approach to develop new compounds have arisen from the growing biochemical knowledge on polyketide biosynthesis. These new strategies consist either in exploiting existing PKS machineries, or genetically engineering the *pks* genes. The first strategy is exemplified by the whole-cell biosynthesis of simvastatin [6], a compound normally synthesized by a chemical modification of the lovastatin polyketide. The authors successfully used the substrate versatility of the PKS responsible for the biosynthesis of the lovastatin to directly produce the simvastatin in an engineered *E. coli* strain. This method represents an interesting alternative to the expensive and time-consuming chemical modification. The latter strategy is commonly referred to as "combinatorial biosynthesis". In addition to classic point mutations targeting amino acids of interest, it requires design of patchwork PKS genes from two or more genes spliced together. Using this strategy, enzymatic activities from a given PKS could be added or exchanged to another PKS. First reports of successful applications to develop new compounds were published in the mid-1990s [7], [8] with the synthesis of a few novel triketide lactones. A decade later, encouraging advances have been made [9], [10], [11], which permitted, for example, the biosynthesis of more efficient geldanamycin analogues anti-cancer drugs (**Figure 3**). In 2005, with the arising of generic approaches to design synthetic *pks* genes [12], directed engineering of modular PKSs had resulted in the production of a total of more than 200 new polyketides, but key challenges remain before the potential of combinatorial biosynthesis can be fully realized. Indeed, these low efficient and labor-intensive experiments cannot be brought up to high-throughput industrial drug discovery. Also, in all the different enzymatic activities catalyzed by PKSs, only the ones responsible for the incorporation, but not modification, of the building blocks have been successfully engineered. Besides, PKSs chosen so far to design patchworks could be considered as "case studies" since they often share high similarity and are therefore easier to manipulate, but do not offer a wide variety of new biosynthesized polyketides.

All the recently developed strategies to identify or create new compounds have been hindered by the lack of precise knowledge concerning the molecular and structural features that govern the biosynthesis mechanism of polyketides. Such knowledge would hopefully allow a more rational design of various polyketides by combinatorial biosynthesis, and help the development of improved methods to

identify naturally occurring polyketides from uncharacterized *pks* genes sequences. Therefore, in parallel to the development of new tools for genetic engineering [13] and heterologous expression [12] to efficiently biosynthesize polyketides, research is now focusing on deciphering the structure and function of the complex machineries that are PKSs.

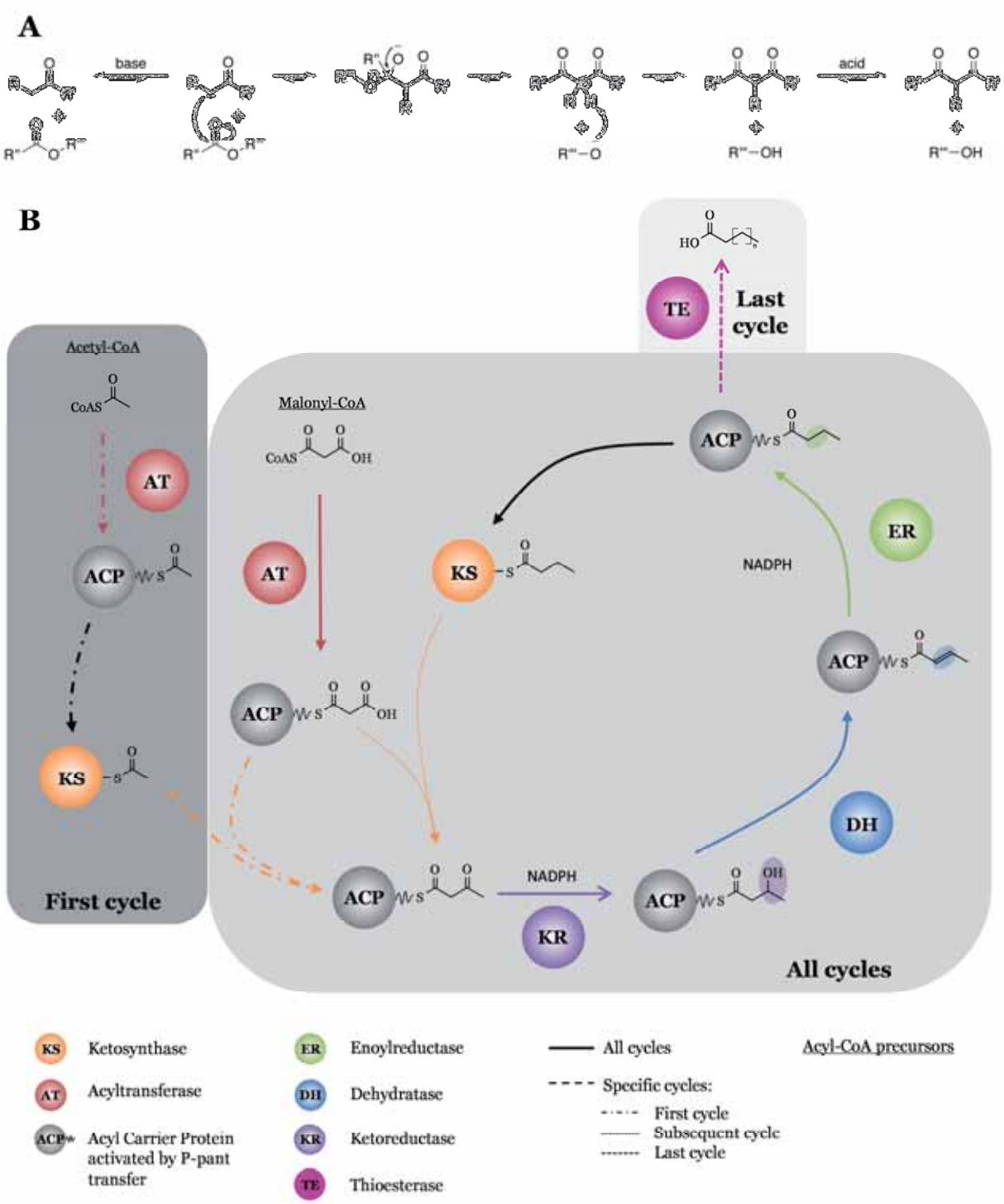


Figure 4. Fatty acid biosynthesis.

(A) Mechanism of the Claisen condensation. Acid/base reactions are catalyzed by the ketosynthase (KS) in the context of fatty acid biosynthesis. (B) Except for the thioesterase (TE), growing intermediates linked to the enzymes are from the first cycle. Fatty acid synthases then use iterative cycles to extend the acyl (represented by the dotted lines. At the last cycle, the acyl is released from the ACP as a free acid (or an ester) by the TE.

B. POLYKETIDE SYNTHASES

1. POLYKETIDE BIOSYNTHESIS

Polyketide synthases are the enzymes responsible for the biosynthesis of polyketides. In order to provide a wide variety of polyketides and to adapt to different environments, these enzymes have evolved to different types. To better characterize and classify polyketide synthases, it is appropriate to digress into the field of fatty acid biosynthesis. Indeed, polyketide and fatty acid biosynthesis are highly related, both in the nature of the chemistry used in the chain extension and in the character of the enzymes that are used for chain assembly. Interestingly, the earliest publications addressing the origins of fatty acids and polyketides actually appeared in the same issue of the *Journal of the Chemical Society* in 1907 [2], [15].

a) Fatty Acid Biosynthesis

GENERAL REACTION OF THE FATTY ACID BIOSYNTHESIS

Fatty acids are assembled from C2 units by head-to-tail linkage using an iterative mechanism. A starter acyl unit (usually acetyl) is attached via a Claisen condensation (**Figure 4A**) to a "building block" (which will now be referred to as an "elongation unit"), *i.e.* a malonyl unit. This condensation reaction results in a β-keto ester, which is then successively modify to give a saturated chain. Several enzymes are involved in this pathway and can be grouped into two classes depending on their function: the condensation enzymes and the modification enzymes. All of these

represented activities are either coded as domains in a multi-domain protein or as individual enzymes assembling into a complex.

THE REACTION SEQUENCE

The condensation reaction takes place with the help of three different enzymes (**Figure 4B**). First, the acyltransferase (AT) recruits an acetyl (from acetyl coenzyme-A) as the starter unit and transfers it to the acyl carrier protein (ACP) which in turn loads it into the cysteine thiol of the KS. Then, the AT recruits an elongation unit from an acyl-coenzyme-A precursor and transfers it to the ACP. The condensation reaction is then catalyzed by the KS and results in a β-keto ester bound to the ACP. The modification reactions, which will now be referred as "β-carbon processing", also involves three enzymes: a ketoreductase (KR), a dehydratase (DH) and an enoylreductase (ER). The β-keto ester first undergoes a reduction by the KR to yield a hydroxyl, and then is dehydrated by the DH resulting in a carbon-carbon double bond formation. Finally the growing fatty acid is reduced again by the ER to yield a saturated carbon chain longer than the original by two methylene units. The product is released as a free acid or an ester by a thioesterase (TE).

THE ROLE OF ACP AND ITS ACTIVATION

The ACP has a central role in fatty acid biosynthesis. It is often referred to as a "shuttle" because it carries the acyl to be extended and brings it to the other enzymes to realize condensation and β-carbon processing. But the acyl is not carried by a thiol in the primary chain of the ACP, instead it is the terminus of a flexible arm added to the protein in a post-translational modification. This ACP activation is catalyzed by phosphopantetheinyl transferases (PPTases) that use coenzyme-A (CoA) and a divalent cation as cofactors. It consists in the covalent linkage of a 4'-phosphopantetheine moiety from CoA to a conserved serine residue of the ACP [16]. The 4'-phosphopantetheine (P-pant) carries the growing chain and delivers it to the various enzymes responsible for chain extension. This post-translational modification has been shown to be essential to fatty acid biosynthesis, polyketide biosynthesis and

non-ribosomal peptide biosynthesis [17], [18]. The activation process is detailed in the chapter dedicated to PPTases (**section I.B.2**).

THE DIFFERENT FAS SYSTEMS

All the represented activities that lead to fatty acid biosynthesis can be genetically and structurally coded in different manners depending on the function to achieve and mainly on the type of organism. Generally, in bacteria the fatty acid synthase consists of a complex composed by a set of discrete proteins that can be isolated, we refer to it as the FAS-II system. In mammals, on the other hand, all the enzymes are covalently linked to form a large multifunctional protein: FAS-I. However, in some organisms, these two systems can be simultaneously expressed and serve different functions. Interestingly, a eukaryotic FAS-II system is also present in mitochondria of eukaryotic cells (for review [19]). While FAS-I mainly produce saturated palmitate and reaction intermediates, FAS-II is more versatile as it can synthesize products with variable chain lengths and different degree of unsaturation or hydroxylation.

b) Polyketide Biosynthesis

THE DIFFERENT TYPES OF PKS

Polyketide synthases can also be divided into type-I and type-II, similarly to the FAS systems. Indeed, type-I and type-II PKS types correspond to a unique polypeptide chain bearing all enzymatic activities or multiple discrete enzymes, respectively. In general, type-I PKSs biosynthesize macrolides, *e.g.* erythromycin, or long aliphatic chains, whereas type-II PKSs often biosynthesize aromatic molecules produced by iterative cycles, *e.g.* tetracycline (**Figure 1**).

Type-I PKSs can be further sub-divided into two categories:

- Iterative type-I which synthesize a polyketide by iterative cycles. At the end of a cycle the extended acyl is transferred back to the KS to realize another cycle, just like FAS does (**Figure 4**).
- Modular type-I, which are consecutive PKSs organized in modules, each module being responsible for the condensation of an elongation unit with the intermediate (equivalent to a cycle). Instead of being subjected to iterative condensation cycles, the acyl is streamed through the different modules until it is released by a thioesterase activity, usually included in the terminal module. PKSs can be mono-modular (one module per PKS), or multi-modular (several modules per PKS).

In addition to type-I and type-II, PKSs can also exist as type-III, which are self-contained enzymes that form homodimers and catalyze the priming, extension, and cyclization reactions iteratively to form polyketide products. They do not use an ACP and hence, do not require post-translational activation by PPTase. Resulting polyketides are often small molecules comprising one or two aromatic cycles Best known examples are flavonoids that constitute the major pigments in plants and are synthesized by chalcone synthases type-III PKSs. Due to their divergence with type-I PKSs, the primary focus of this manuscript, type-III PKSs will not be described. However, excellent reviews are available on the subject: [20], [21], [22].

Together with non-ribosomal peptide synthases (NRPSs), type-I PKS and type-I FAS form the megasynthases class of enzymes. Megasynthases are large proteins containing all the catalytic activities required for the biosynthesis of their respective products on a single polypeptide chain. Catalytic activities are borne by domains separated by linkers on the polypeptide chain. These enzymes also have in common the use of a carrier protein (CP) domain, named acyl carrier protein (ACP) for FASs and PKSs, and peptidyl carrier protein (PCP) for NRPSs because of the nature of the carried intermediate. All these carrier proteins must be activated by the transfer of a P-pant moiety by PPTases. NRPSs also possess condensation and modification domains, but due to the peptidic nature of their substrates these domains are not homologous to FASs and PKSs domains, and thus will not be discussed.

Despite their architectural and functional similarities, FAS and PKSs also possess striking differences in their composition. While type-I FASs necessarily

include KS, AT, DH, ER, KR, ACP and TE domains to carry out their function, a distinction between essential and optional domain exists in PKSs:

- Essential (or core) domains are responsible for condensation: KS, AT and ACP
- Optional domains in different combinations: KR, KR-DH, or KR-DH-ER, that gradually process the previously formed β-keto ester to give a hydroxyl, a carbon-carbon double bond, or a saturated chain, respectively
- The thioesterase activity can be borne by a domain or a trans-active enzyme
- Additional domains can also be included in PKSs such as methyltransferase and sulfhydrolase

Partial β-carbon processing of the intermediate, involving KR and DH domains, will then allow the presence of hydroxyl and keto substitutions or the formation of a permanent carbon-carbon double bond with all associated stereochemistries. In addition to the variability in domain composition, AT domains from PKSs cover a wide range of possible elongation units and a given AT domain can sometimes even accept different elongation units.

The possible organization of PKSs into modules permits a great number of combinations in terms of domain composition, substrate specificity and selectivity for each round of condensation/β-carbon processing. This overall flexibility in PKSs allows the synthesis of a wide variety of polyketides. Besides, accounting for the potential action of trans-acting enzymes to further modify the resulting polyketide, possibilities for the nature of the final product are then almost infinite. For these reasons, type-I PKSs and especially modular ones represent very interesting targets with regard to biotechnological aspects such as combinatorial biosynthesis. Indeed, the presence of all domains on the same polypeptide chain makes PKSs easier to manipulate.

ERYTHROMYCIN BIOSYNTHESIS: A PARADIGM OF MODULAR PKSs

The biosynthesis of erythromycin has been extensively studied since the discovery of this macrolide polyketide in 1949 by Eli Lilly and co-workers (U.S. patent

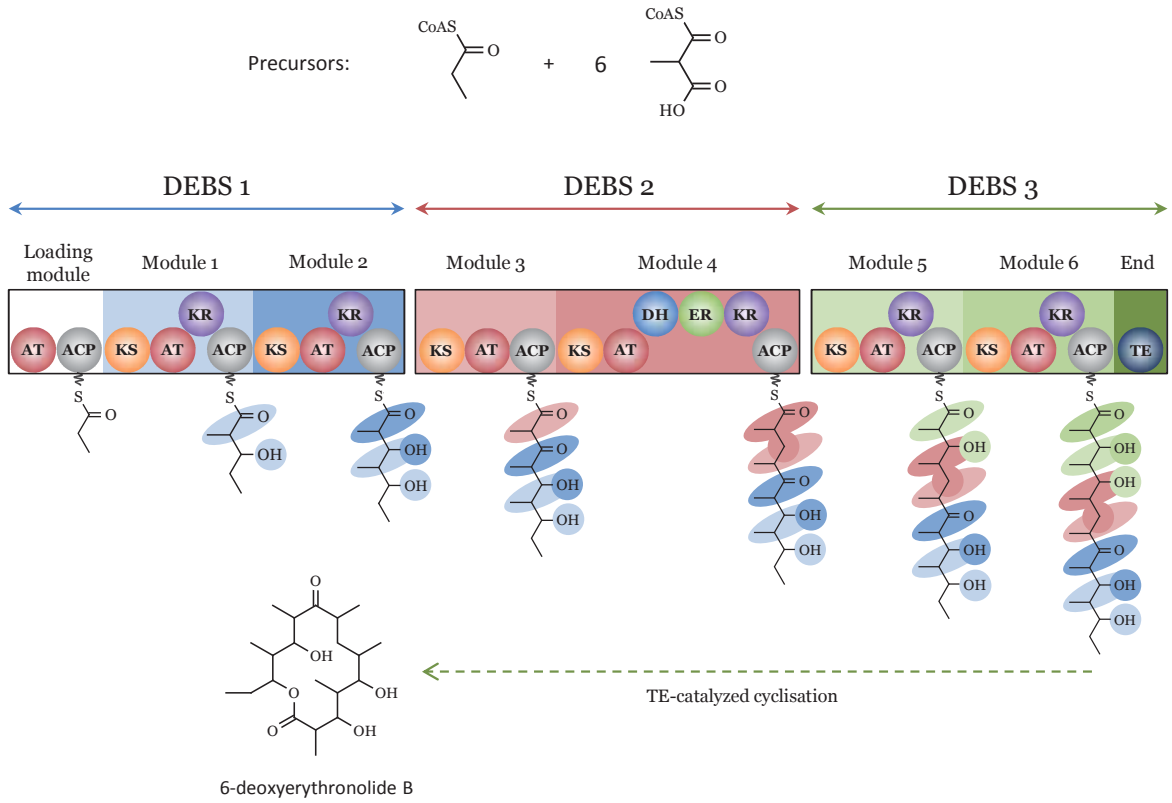


Figure 5. 6-deoxyerythronolide B synthase: the modular polyketide synthase paradigm.

Modules are encoded by three independent polypeptide chains DEBS1, DEBS2 and DEBS3 but act in concert to grow the polyketide chain to catalyze the formation of 6-deoxyerythronolide B (6-dEB). First, the AT domain of the loading module recruits a propionyl moiety and transfers it to its cognate ACP. The 6 subsequent modules then catalyze 6 rounds of condensation/reduction using malonyl as the elongation unit. Depending on each module composition, the ketone of the previously incorporated malonyl (or propionyl for module 1) is either left intact (module 3), reduced to a hydroxyl (modules 1, 2, 5 and 6) or reduced to a hydrogen-saturated carbon (module 4) as depicted by color conservation between modules and modifications. The TE domain finally catalyzes the cyclisation and release of the chain to yield 6-deB. 6-deB will then be further enzyme-modified to give erythromycin.

2,653,899; 1953). It is now well characterized and represents the paradigm for the understanding of the structure and biochemical function of modular polyketide synthases.

Eryrthromycin biosynthesis can be divided into two phases. In the first phase the PKSs catalyze sequential condensations of one unit of propionyl-CoA and six units of methylmalonyl-CoA to give 6-deoxyerythronolide B (6-deB), the first enzyme-free intermediate. In the second phase, 6-deB is modified by a series of "tailoring" enzymes which include hydroxylases, glycosyltransferases, and methyltransferases. Though the modifying steps are essential to produce the active antibiotic erythromycin, the rest of this chapter will focus on the biosynthesis catalyzed by modular PKSs, the 6-deoxyerythronolide B synthases (DEBS).

6-deoxyerythronolide B biosynthesis is catalyzed by 3 DEBS subunits (**Figure 5**). Each homodimeric DEBS subunit contains two 160–200 kDa protein modules, each responsible for a single round of polyketide chain extension and β-carbon processing. Within each module are present several catalytic domains that are analogous in structure, function, and organization to the corresponding FAS components. All six DEBS modules contain the three core domains but their composition in optional β-carbon processing domains is variable. Additionally, at the N-terminus of the most upstream module is a loading didomain (AT-ACP) that primes the KS domain of module 1 with the propionyl starter unit. Finally, cyclization and release of the 6-dEB is controlled by a thioesterase (TE) domain located at the C-terminus of the most downstream module.

OUTSTANDING ISSUES

As highlighted above, the DEBS example shows the striking collinearity that exists between the organization of the domains encoded by the PKSs genes and the order of the biochemical reactions that generate the polyketides. Indeed, the chain length, substitution pattern, and oxidation level of the initially generated linear 6-dEB, are the direct consequence of both the number of DEBS modules and the domain composition of these modules. Nonetheless, such straightforward correlations have proven to be elusive for many modular PKSs-based biosynthesis mechanisms. Presence in individual PKS modules of inactive β-carbon processing

domains [23], [24], [25], *trans*-acting domains [26], [27], [28], [29], multiple AT or ACP domains [30], [31], "broken modules" [32], [33], as well as module skipping [34], [35] and stuttering [36], [37] hinder the establishment of a clear relationship between *pks* gene sequences and biochemical function. Also, unusual AT domain specificity toward elongation units such as chloroethylmalonyl-CoA [38] or hexylmalonyl-CoA [39] eventually adds further complexity to the process.

In addition to these "abnormalities" in PKS modules, there still are a number of outstanding issues about polyketide biosynthesis that mainly concern:

- Elongation unit selectivity by AT domains
- Recognition of the growing polyketide by KS domains.
- Substrate specificity of the β-carbon processing domains.
- Stereochemistry induced by β-carbon processing of the β-keto ester
- Domain-domain interactions required for both stabilizing PKS structure and mediating ACP specificity toward the other domains
- Module-module interactions that allows efficient and error-prone streaming of the intermediate to the next module.
- Overall structural architectures of isolated modules (mono-modular PKS) or iterative PKS that could be different depending on their domain composition.
- Overall structural architecture of multiple modules on multi-modular PKS (DEBS-like)
- Interactions with enzymes acting in *trans*
- Shuttling with other metabolite biosynthesis pathways

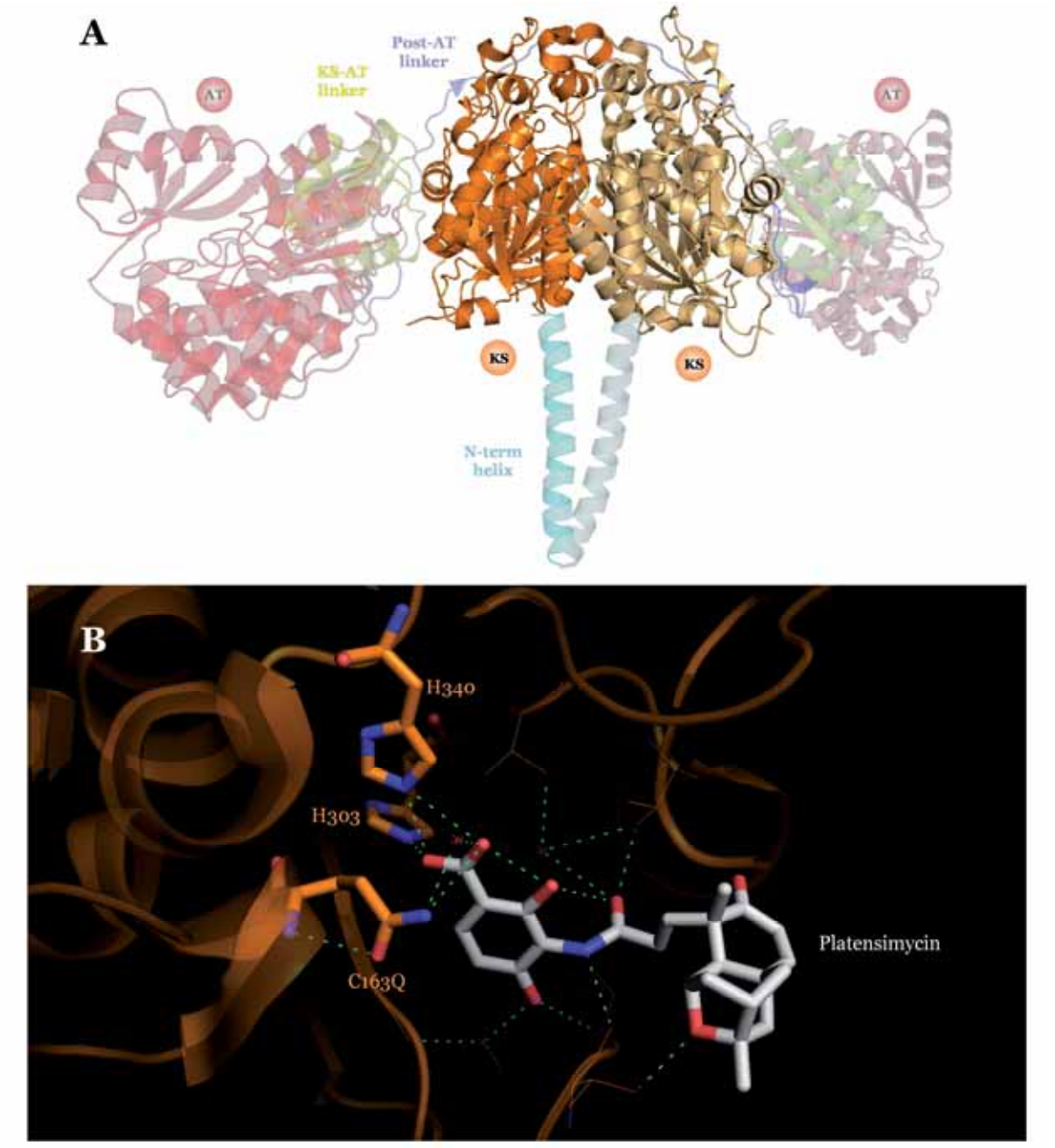


Figure 6. Architecture and ligand binding of the KS domain.

(A) Cartoon representation of the crystal structure of the KS-AT didomain from DEBS module 5 showing the role of the KS dimer in the architecture of the condensation domains (PDB code: 2HG4). (B) Interaction network between the inhibitor platensimycin and *E.coli* FabF. Catalytic histidines and cysteine (mutated to glutamine) are shown in sticks (PDB code: 2GFX).

c) Review of Structural and Functional Data on Polyketide Synthases

(1) CONDENSATION DOMAINS

In the last two decades, efforts have focused on both structural and functional studies of PKSs-mediated biosynthesis and yielded great advances in the understanding of these complex machineries. X-ray or NMR structures of each individual domain from PKSs are now available and some issues highlighted in the previous section have been at least partially solved. Yet, structures of multiple domains on a single polypeptide chain are almost inexistent though they are essential to decipher the complex mechanism required for polyketide biosynthesis. For that matter, biochemical and structural information on sequence- and structure-related FASs remain invaluable for PKSs studies. Furthermore, in addition to the functional analysis of isolated domains, recent results on linkers, inactive domains, docking domains or dimerization motifs indicate an essential role of these structural constituents in the structure and function of PKSs and are therefore mentioned in this chapter.

THE KETOSYNTHASE DOMAIN

Among PKS domains, the ketosynthase domain is the most conserved. All KS domains, either from PKSs or FASs, adopt a thiolase fold (**Figure 6A**). This fold consists in alternative layers of α-helices and β-sheets to form an α/β/α/β/α architecture. Several 3D structures of KS-AT didomains are available from type-I PKS [40], [41] and type-I FAS [42].

KS is a homodimer composed of ~400-450 residue-long monomers with an extensive dimer interface stabilized by a pair of hydrogen-bonded, antiparallel β-strands, which creates a 14-stranded β-sheet spanning both monomers. It is thought to stabilize dimerization of the entire megasynthase [43]. KS is responsible for the

chain extension of the growing polyketide, and thus features a substrate-binding tunnel divided into two halves: an acyl-binding region to bind the ACP-acyl to be extended, and a P-pant-binding region to bind the elongation unit. The P-pant-binding region stretches from the enzyme surface to the active-site cysteine and is relatively well conserved, reflecting its universal role in binding the P-pant moiety. The reactive cysteine is positioned on a "nucleophilic elbow" such that the positive dipole of the α-helix decreases its pKa and increases its nucleophilicity [44]. In addition to this cysteine (from a TA**C**SSS motif), the catalytic machinery also comprises two histidine residues (from EA**H**GTG and KSNIG**H**T motifs). The acyl-binding region, on the other hand, can vary significantly between different KS domains. While type-I FAS domains are highly specific toward saturated acyl chains [44], type-I modular KS domains have a wider range of substrate specificities that vary in length from diketide to decaketide. Also, some PKS KS domains have been shown to possess specificity with regard to different β-carbon status [45]. Indeed, the acyl-binding pockets of PKS KS domains are amphipathic and allow hydrogen-bonding interactions with the carbonyl groups of the growing polyketide chain [40], [41] whereas FAS KS domains have a hydrophobic, narrow pocket of a suitable size to specifically accommodate their corresponding fatty acyl substrates [46]. Hence, substrate specificity may rather arise from selective ACP recognition, as seen in DEBS modules [47].

A putative condensation mechanism has been described based on a comprehensive structural study of the FabF enzyme, a type-II KS, either inhibitor-bound (**Figure 6B**) or covalently linked to a dodecanoic acid. In both cases, the catalytic cysteine had been mutated to a glutamine [48]. First, a trans-thioesterification reaction would take place between the acyl-ACP and the reactive cysteine inside the substrate tunnel with the help of an oxyanion hole. Then, the ACP bearing an elongation unit would dock to the acylated KS which generate, by a decarboxylation process, an enolate that could finally attack the thioester carbonyl to elongate the polyketide. This putative mechanism is based on a single comprehensive study and needs to be confirmed or at least completed. The lack of (physiological) substrate-bound KS structures hinders the understanding of the detailed reaction mechanism of decarboxylation, acyl transfer, and KS-ACP interaction. It is also noteworthy that other inhibitors of KS activity, such as cerulenin and thiolactomycin,

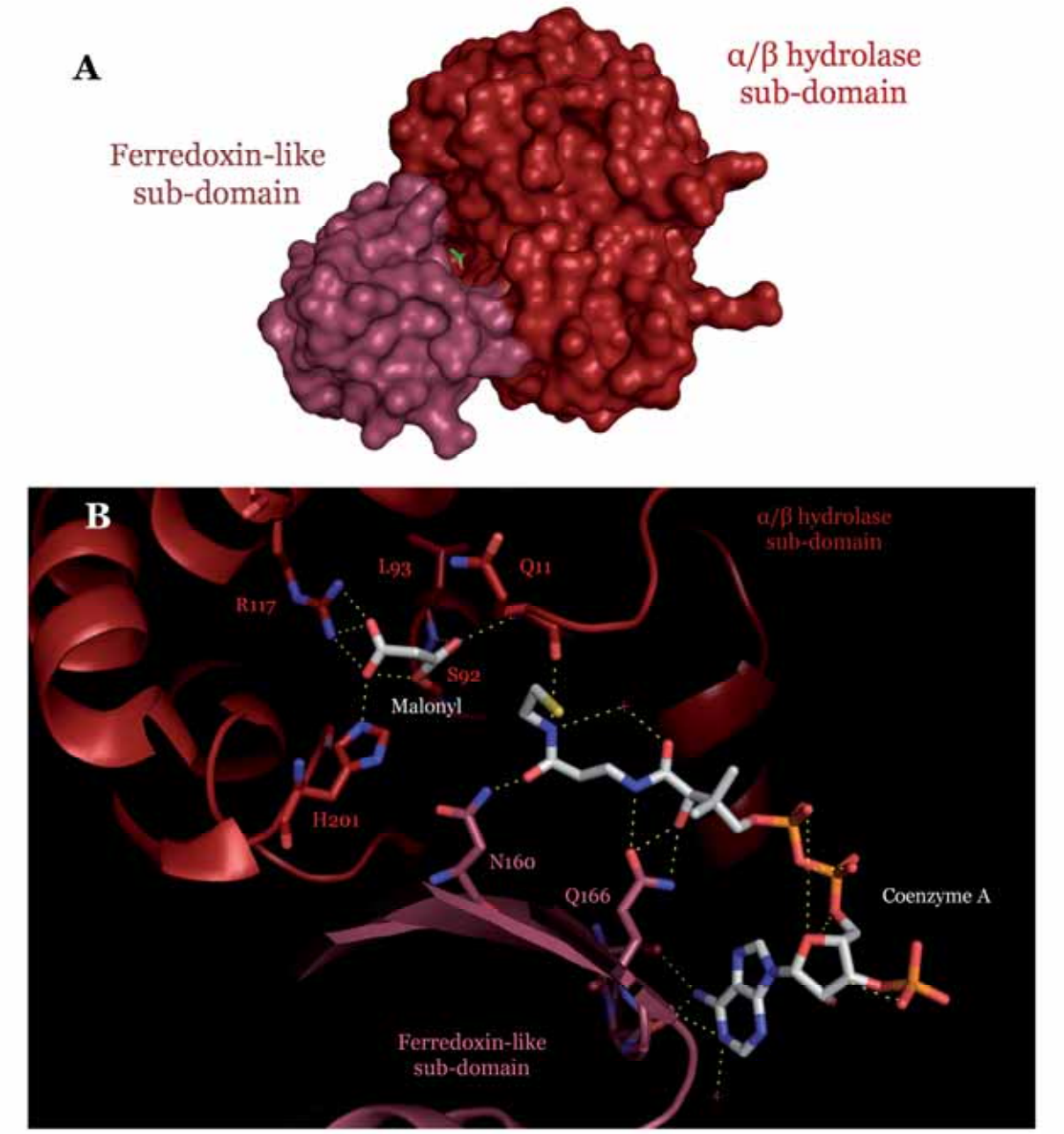


Figure 7. Acyltransferase fold and substrate binding.

(A) Surface representation of the AT domain from DEBS module 3. Bound acetate, colored green, lies at the bottom of the open substrate cavity formed by the two subdomains (PDB code: 2QO3). (B) Interaction network between malonyl-CoA substrate and AT domain. Covalently bound malonyl and stabilized CoA are represented in white sticks. The catalytic S92 as well as stabilizing residues are represented in sticks and annotated (PDB code: 2G2Z).

identified almost half a century ago, have been structurally characterized [49], [50] but are still not clinically used.

THE ACYLTRANSFERASE DOMAIN

The acyltransferase is a ~300 residue domain (**Figure 7A**) connected to the KS domain through a ~140 residue "KS-AT" linker**.** ATs contain two subdomains: a ~240 residue catalytic subdomain that possesses an α/β hydrolase fold and ~60 residue ferredoxin-like subdomain. All reported structures are highly similar (rmsd <2 Å). These include DEBS modules 3 [41] and 5 [40], and human FAS [42] as didomains together with KS domain but also FabD, the FAS-II acyltransferase from *Escherichia coli* [51], and the malonyl-CoA:ACP transacylase from *Streptomyces coelicolor* [52].

AT domain catalyzes the transfer of a starter/extender unit to its cognate ACP, it is responsible for the selection of this unit and is then often referred to as the "gatekeeper" of polyketide biosynthesis. Transfer to ACP occurs by a ping-pong bi-bi mechanism through a covalently bound acyl-intermediate to the catalytic serine. Once the catalytic serine has been activated by the neighboring histidine, starter/extender unit is recruited and it results in formation of an acyl-AT. Such acyl-enzyme complex has been observed when soaking FabD with malonyl-CoA (**Figure 7B**). In addition to confirming the role of an invariant arginine (R117 in FabD) in interacting with the substrate carboxylate, the acyl-FabD structure has also greatly helped in understanding the catalytic mechanism [53]. In FabD, the catalytic serine S92 attacks the malonyl-CoA thioester carbonyl and forms an intermediate stabilized in the oxyanion hole mediated by Q11 and L93. Then, the imidazole side chain of H201 positions the hydroxyl function of S92 for subsequent nucleophilic substrate attack by the free thiol group of the P-pant carried by the ACP domain. Interactions with the P-pant moiety from CoA are also described and involve the carbonyls of S163, P164 and G165, and the side chains of N160 and Q166 (**Figure 7B**).

Several AT domains which possess specificity toward malonyl- and methylmalonyl-CoA have been structurally and functionally characterized and correlated motifs for the selection of the extender unit have emerged. Two sequence motifs lying in close proximity to the catalytic residues, GH**S**XG and XAX**H,**

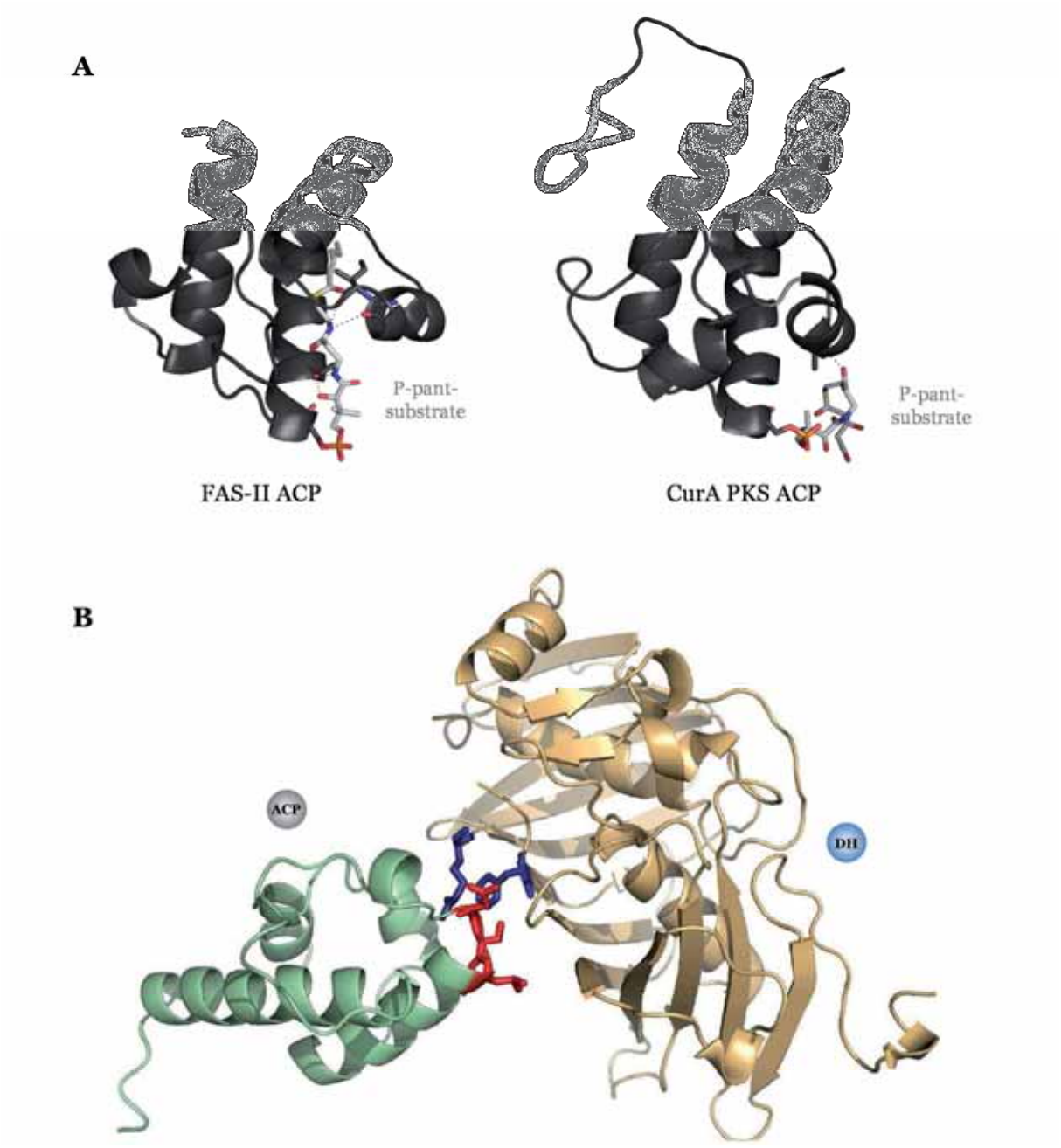


Figure 8. Acyl-ACP conformations and ACP-DH docking model.

(A) Comparison between type-II FAS (PDB code: 2FAC) and type-I PKS (PDB code: 2LIW) ACP-acyls. In FAS-II, the acyl is buried into a central cavity and stabilized by several hydrogen bonds whereas in type-I PKS, the acyl is stacked against the ACP structure via a single hydrogen bond. (B) Figure. Docking model of the DH/ACP interaction with putative residues engaged in the interaction represented in sticks. **From Anand *et al.*, 2012**

(catalytic residues are in bold), have been identified to specifically bind either malonyl-CoA (GH**S**QG and HAF**H**) or methylmalonyl-CoA (GH**S**(I/V)G and YAS**H**). Unfortunately, these motifs are not universally conserved and it is sometimes very difficult to guess which extender unit is selected by analyzing the amino acid sequence only. Also, some ATs are specific for substrates other than malonyl- and methylmalonyl-CoA and no general pattern could be assigned to these domains yet. Nonetheless, the arginine that forms a salt-bridge with the substrate carboxylate in extending AT domains is often replaced by a non-polar amino acid in loading domains which are selective for shorter substrates. For example, the AT domain from the DEBS1 loading module (**Figure 5**) contains a tryptophan instead of an arginine and is selective for acetyl-CoA.

While the catalytic mechanism has been defined in good details thanks to structural information, assigning the substrate specificity on a sequence basis only, still remains a challenging task. Indeed, substrate specificity is likely to be a combinatorial result of different structural elements that interact throughout the entire protein fold, rather than an influence of a limited number of residues. Thus, additional structures of native or mutated AT domains in complex with their substrate may be required to clear this specificity problem out.

THE ACYL CARRIER PROTEIN DOMAIN

ACP covalently shuttles starter/extender units and growing intermediates to the other domains. This ~80-residue long domain consists in a four-helix bundle and is usually located at the C-terminal end of a PKS module. NMR structures are available for two modular type I PKS: DEBS module 2 [54] and curacin PKS CurA [55]. Despite low sequence identity between ACP domains, the overall fold and the motif “DSL” at the end of helix α2 which contains the catalytic serine are always conserved. However, the position of the bound acyl may change, as illustrated by either crystal structures of acyl-ACPs from type-II FASs [56] with fatty acyl groups pointing toward the inner cavity, or NMR structures of acyl-ACPs from rat type-I FAS [57] and type-I PKS CurA [55] where the bound substrates are rather stacked against the helical bundle (**Figure 8A**).

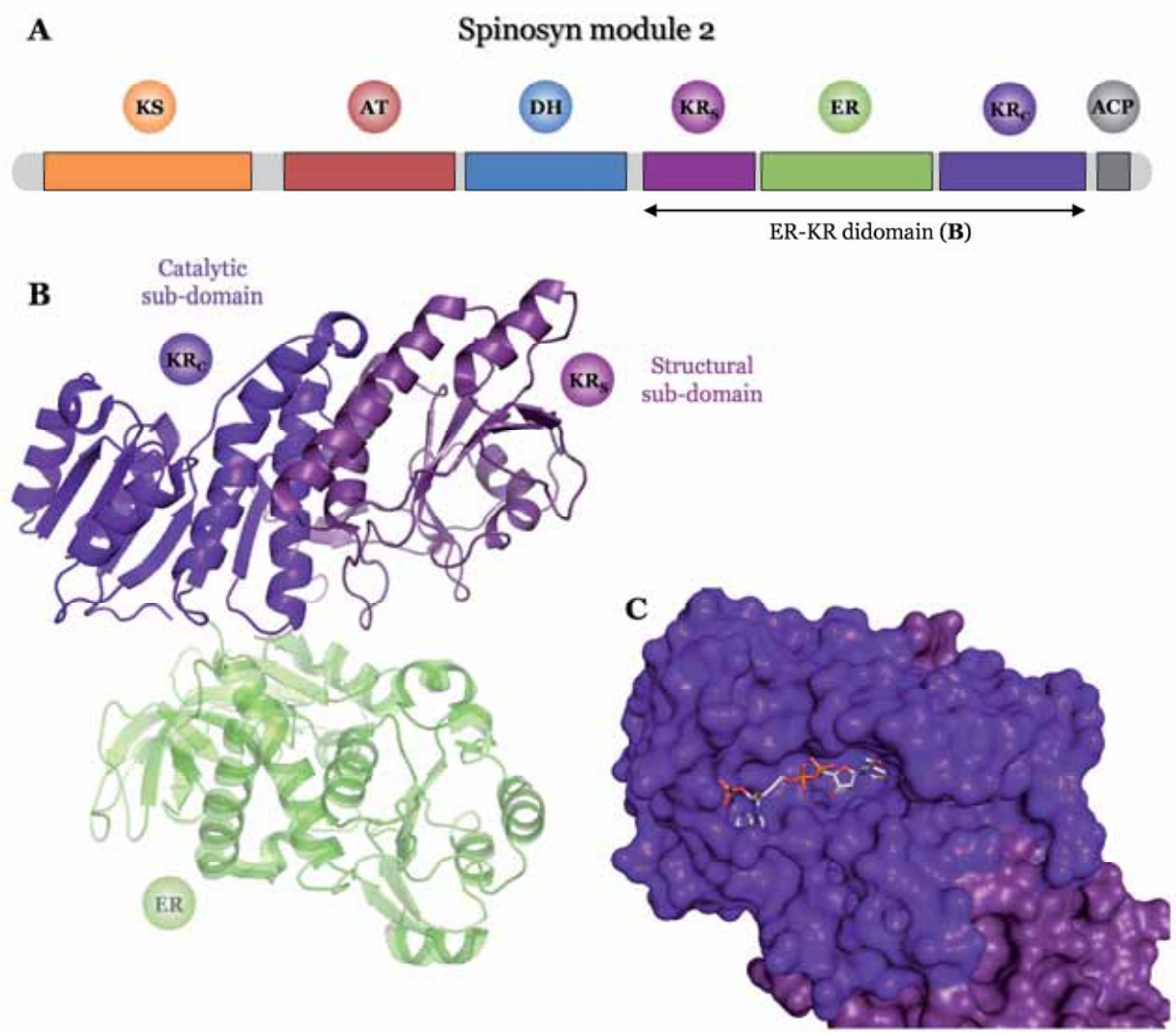


Figure 9. KR structural and catalytic sub-domains structure.

(A) Domain organization of the spinosyn module 2 showing sequence "splitting" of the KR domain as a result of the insertion of the ER domain between its two sub-domains. (B) Crystal structure of the didomain ER-KR from spinosyn module 2 confirming spatial proximity between the two KR sub-domains (PDB code: 3SLK). (C) Binding of NADPH in the solvent-exposed cavity of the KR catalytic sub-domain. NADPH is represented in sticks, sub-domain coloring is identical to A (PDB code: 2FR0).

Since ACP domain acts as a shuttle, it must communicate with all other domains. Characterization of the interactions that mediate communication between ACP and other domains is critical to understand how the corresponding activities are regulated. Nonetheless, to date, no complex involving a condensation or a β-carbon processing domain and an ACP domain from type-I PKS has been experimentally determined. Instead, modeling studies have been carried out where ACP docking poses are proposed for KS-AT didomain [58], DH and KR domains [59]. These studies have allowed the identification of putative residues involved in ACP interaction/recognition. For example, the DH/ACP (from DEBS module 4) modeling study [59] has allowed confirmed the role of two putative interacting residues on DH, previously reported as hindering dehydration activity of acyl-ACP when mutated (**Figure 8B**) [60]. Another example is the successful engineering of a modified PKS resulting in a module that catalyzes two rounds of condensation instead of one [47]. This very promising work is based on mutations in the ACP domain that have been guided by KS-AT/ACP modeling [58]. According to these docking studies, ACP interactions with other domains are very labile and seem to be mediated by a little set of residues that are not universally conserved among PKSs and might then not account for a universal interaction mechanism.

(2) BETA-CARBON PROCESSING DOMAINS

THE KETOREDUCTASE DOMAIN

The ketoreductase domain is monomeric and ~450 residue long. It comprises a ~200 residues "structural" (non-catalytic) sub-domain and a ~250 residues catalytic sub-domain, both of which possessing a Rossmann-like fold. KR belongs to the SDR (short-chain dehydrogenase/reductase) superfamily and is structurally and functionally related to the FAS-II FabG ketoreductase. Within both FASs and type I PKSs containing a full set of β-carbon processing domains, the structural sub-domain (named KRs in PKSs or ΨKR in FASs) is distantly located from the catalytic sub-domain in the primary structure (**Figure 9A**) though these are very close in the tertiary structure (**Figure 9B**). The structural sub-domain seems to stabilize the

catalytic subdomain, thereby enabling the entire KR domain to be located away from the 2-fold axis of the megasynthase [46].

The KR domain is responsible for the reduction of the β-ketone from the elongation unit incorporated by the upstream module using NADPH as a co-factor. It also sets the stereochemistry of the resulting hydroxyl group. KRs can be divided into three types: A-type and B-type depending on the stereochemistry of the generated β-hydroxyl, or C-type if they are inactive. Modules that contain others β-processing enzymes in addition to KR usually contain a B-type KR. Some KRs can also epimerize the α-substituent and are then further divided into subtypes 1 and 2 based on the orientation of the α-substituent thus yielding A1-, A2-, B1- and B2- subtypes. High-resolution crystal structures are available for each of these KR-types isolated from type-I PKS [61], [62], [63], [64]. Fingerprints have been identified to distinguish between the different KR types based on their amino acid sequence. The most obvious one is a B-type characteristic "LDD motif" that allows to differentiate B-type KRs from A-type KRs, in which this motif is absent. More intuitively, C-types lack either the catalytic tyrosine or the NADPH binding site that are present in A- and B-types. Remarkably, in some iterative type I PKSs, the KR switches between A-type and B-type depending on the intermediate it is reducing [65].

While almost all the reported structures are composed of a binary complex with the co-factor NADPH (**Figure 9B**), no structure of a ternary complex involving a bound polyketide has been deposited in the PDB yet. However, putative catalytic residues have been identified and consist in a tyrosine (from a YAAAN motif), a lysine, and a serine. The hydroxyls of serine and tyrosine would be in charge of activating the polyketide β-ketone. When the NADPH attacks this β-ketone, the lysine-stabilized tyrosine would give its proton and then generate a β-hydroxyl polyketide. Other residues are also thought to play a role in binding the polyketide. Interestingly, a few point mutations on isolated KRs are sufficient to switch from a KR type to another, but when these mutants are introduced back into a full-length PKS, the programmed polyketide is not produced [61], [66].

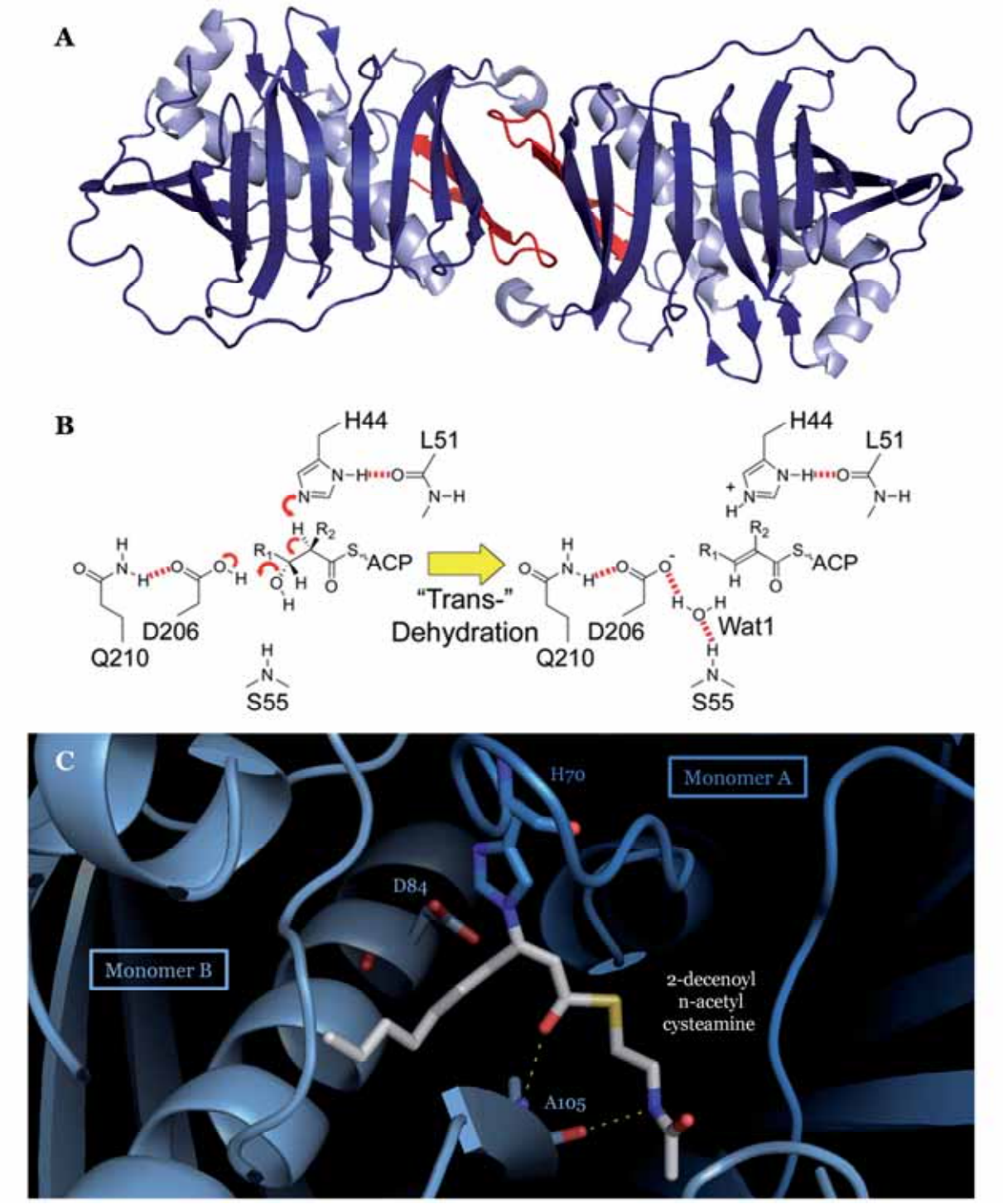


Figure 10. DH fold, dimerization and inhibitor binding.

(A) DH dimer from DEBS module 4, α-helices are colored light blue, and β-strands are colored dark blue. The first 20 residues participating in the linear dimerization are colored red (PDB code: 3EL6). (B) Proposed mechanism of *trans* dehydration for DEBS module 4 DH. (C) Covalent binding of 2-decenoyl-n-acetylcysteamine to the catalytic histidine in the DH substrate tunnel. The catalytic dyad and the stabilizing residue are shown in sticks, the inhibitor is represented in white sticks. To the exception of A105, no other interacting residues have been identified.

THE DEHYDRATASE DOMAIN

Dehydratase domain catalyzes the dehydration of a polyketide intermediate to yield a double bond between the α- and β-carbons. Interestingly, when isolated from a type-I PKS it is also able to catalyze the reverse reaction, even if this function is apparently not used in the context of full-length PKSs. Hydrated product might be counter-selected by downstream KS.

A few 3D structures of DH domains from type-I PKSs are available: the DH domain from DEBS module 4 [60] and DH domains from the curacin PKSs [67]. They all possess a double hotdog fold (**Figure 10A**) similar to the homodimers of FAS-II FabA, each FabA monomer being a single hotdog [68]. For this reason, DH domain monomers from type-I FASs and type-I PKSs are thought to arise from ancestral gene duplication of type-II FASs. However, while FabA contains two active sites – one by monomer – DH domains from FASs or PKSs only contain a single one.

Dehydratase activity requires a catalytic dyad composed by an aspartic acid and a histidine. It is noteworthy that in some cases, the aspartic acid is replaced by a glutamic acid. The aspartic acid is thought to donate a proton to the β-hydroxyl, and the histidine is thought to abstract one, thus catalyzing the elimination of a water molecule and generating a carbon-carbon double bond (**Figure 10B**). Most generated products bear a *trans* double bond but when DH domains are preceded by A-type KR domains it results in the formation of a *cis* double bond [67]. Some DHs are also known to possess an isomerase activity that shifts a double bond from the α- and β-carbons to the β- and γ-carbons [69]. However, sequence alignments of DHs have not permitted to identify patterns that distinguish between DHs that (1) catalyze the formation of *cis* versus *trans* double bonds, (2) serve as epimerases, or (3) are nonfunctional.

Details about P-pant-bound substrate stabilization and the dehydration mechanism are still needed. For example, conserved active site residues have been identified but no function has been attributed to them yet [60]. A structure of a binary complex between FAS-II FabA and the mechanistic inhibitor 3-decynoyl-N-acetylcysteamine is available though. However, this ligand presents a *cis* double bond, it is not a DH substrate since it covalently binds the catalytic histidine and therefore probably adopts a biased conformation (**Figure 10C**). Another

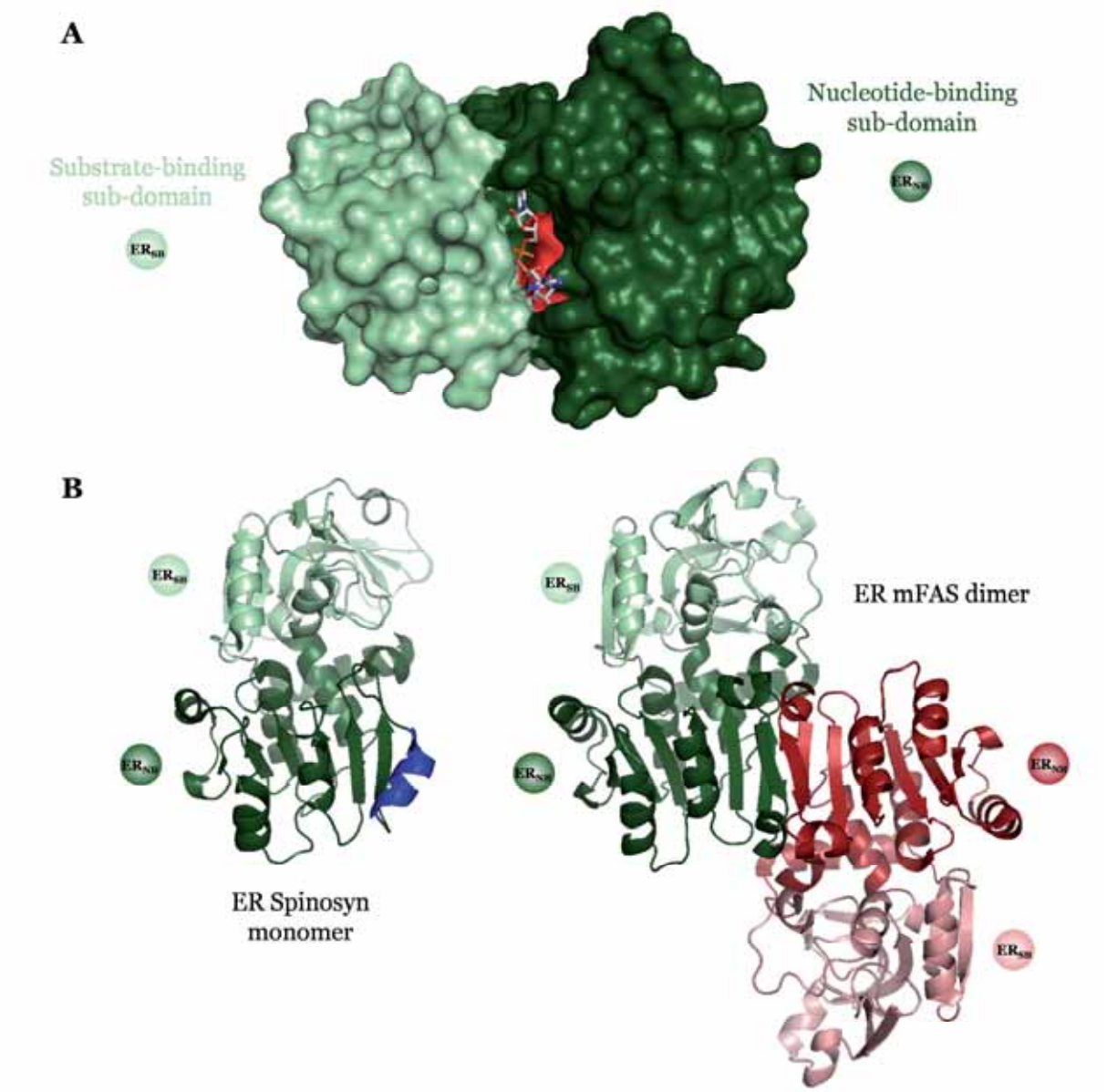


Figure 11. ER domain fold, NADPH binding and oligomerization state.

(A) ER folds into two sub-domains creating a cavity in which NADPH binds through a conserved GGVGMA motif shown in red (PDB code: 3SLK). The substrate binding sub-domain is colored in light green, and the nucleotide binding sub-domain is colored in dark green (B) Comparison between the ER domain from porcine FAS (PDB code: 2VZ8) that dimerizes through two anti-parallel β-strands and the ER domain from the spinosyn PKS module 2 in the same orientation. In the spinosyn ER, the dimerization is made impossible by the presence of an α-helix (colored in blue) upon the β-strands.

controversial aspect concerns the dimerization interface of DH domains inside PKSs. Indeed, all reported structures from isolated DH domains describe a linear dimer (**Figure 10A**) whereas the DH dimer in the mammalian FAS-I is bent ~60° [46]. Interestingly, the dimerization interface of PKS DH domains is mainly built by the ~20 N-terminal residues which are not present in the mammalian FAS DH (**Figure 10A**).

THE ENOYLREDUCTASE DOMAIN

ER belongs to the acyl-CoA reductase family of the medium-chain dehydrogenase/reductase (MDR) superfamily and is structurally related to the bacterial quinone oxidoreductase [70]. Its active site is located in a cleft between the ~140-residue nucleotide-binding sub-domain and the ~170-residue substrate-binding sub-domain (**Figure 11A**). It employs NADPH to stereoselectively reduce *trans*-α,β-double bonds generated by DHs.

Structural determination of the $NADP^+$-bound ER domains from porcine FAS [46] and spinosyn PKS module 2 [71] have helped the identification of putative catalytic residues. They reveal an invariant lysine-aspartate pair ~6 Å from the NADPH although the aspartate is not identically positioned in the two structures. An additional tyrosine has been proven to play a role in setting an *S* configuration on the α-carbon (when substituted) but is present in only half of the ER domains from modular PKSs [71]. The spinosyn ER domain shows a decreased activity on a short substrate-analogue when Y421 and K422 are mutated to alanines. On the contrary, D444 does not seem to participate to the catalytic mechanism of the spinosyn ER domain, since the catalytic activity is not altered when this residue is mutated to an alanine. However, it may participate in the FAS catalytic mechansim, since it is strictly conserved among mammalian FASs [46]. ER domains bind NADPH in a deep cleft located between the subdomains (**Figure 11A**). A pyrophosphate-binding motif GGVGMA has been identified, along with conserved residues lysine and arginine that form salt bridges with the adenine ribose phosphate. Eventually, a ternary complex involving a bound polyketide intermediate is missing to (1) determine which residues are involved in stabilizing the substrate and (2) decipher the catalytic mechanism.

One striking difference between mFAS and spinosyn PKS structures is the oligomerization state of the ER domain (**Figure 11B**). Indeed, as seen in mFAS, MDR enzymes usually dimerize through hydrogen bonds and form a 12-stranded β-sheet spanning the whole dimer. Instead, the spinosyn PKS ER domain crystallizes as a monomer, a state already observed for the *trans*-acting lovastatin ER [72]. In addition, interaction with KR domain is very different to the one described in the porcine FAS structure. Also, in multi-modular type-I PKSs, it is thought that dimerization of ER domains would hinder movement of the cognate ACP which is already restricted by the downstream module [71]. This feature is obviously not relevant in FASs or mono-modular type-I PKSs whose mobility is not restricted in such manner.

(3) OVERALL ARCHITECTURE OF FATTY ACID SYNTHASES

No structure of a type-I PKS is yet available but clues on the overall architecture of PKS can be derived from structural information on mammalian FAS. Indeed, (1) domain organization is preserved between the two megasynthases, (2) both mFAS and type-I PKSs are homodimeric [73], [74], and (3) sequence and especially structure analysis of individual domains reveal higher similarity between mFAS and PKS than between mFAS and bacterial FAS-II. In addition, there is evidence for a common evolution between mFASs and PKSs. An active methyltransferase (MT) domain from a common ancestor of PKSs and mFASs may be at the origin of both the non-functional MT domain found in mFAS, and the active ones found in several PKSs [75], [76]. mFAS can then be considered as a single PKS module specialized for iterative fatty acid synthesis [46]. Although no structure of a full-length type-I PKS is available, didomains from PKS have been structurally determined. These structures show a similar architecture to mFAS but also some differences in the organization of the domains. This chapter will then summarize structural data coming from mFAS (both crystal [46] and electron microscopy [77]) and from mono- and multi-domain structures of PKSs [40], [41], [60], [67], [71].

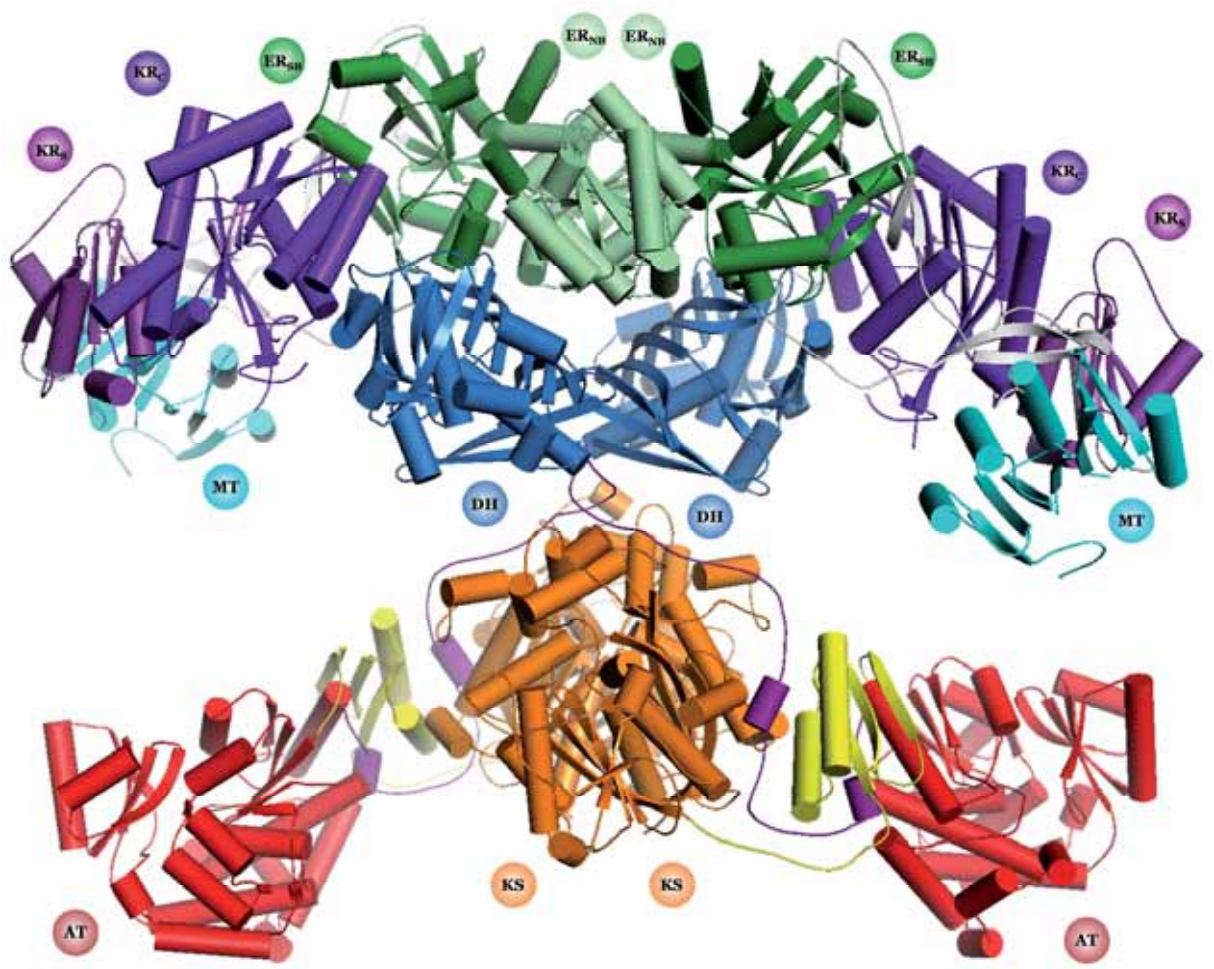


Figure 12. Representation of the porcine FAS crystal structure.

mFAS crystal structure showing overall architecture and domain-domain interactions. ACP and TE domains are absent from the crystal structure. Helices are represented as cylinders for clarity. Domain coloring (by sequence order): KS is orange, KS-AT linker is yellow, AT is red, post-AT linker is purple, DH is blue, ΨME is cyan, ΨKR (KR_S) is violet purple, ER substrate binding sub-domain is forest green, ER nucleotide binding sub-domain is pale green, KR_C is purple blue, linkers are white.

As seen in **Figure 12**, mFAS assembles into a homodimer resembling an "X" shape where the lower legs of the "X" are represented by the KS and AT domains, and the upper legs are represented by the β-carbon processing domains DH, ER and KR [46], [77]. ACP and TE domains are not clearly defined in density and thus were not positioned in the overall architecture. Two additional non-catalytic domains are found in the β-carbon processing part of the structure: a pseudo-methyltransferase (referred as ΨME) and a non-catatlytic KR (referred as ΨKR, similar to the "structural" KR sub-domain described in several structures of PKSs). mFAS homodimer is stabilized by the dimerization interfaces of KS and ER domains, and to a lesser extent, by the DH domain and the linker between AT and DH domains. These interactions are responsible for the 5400 Å² dimer surface contact area.

The condensation part seems to be particularly well conserved between mFAS and PKSs [40], [41], [46] (**Figures 6A and 12**). Indeed, in both cases, KS mediates homodimerization at the 2-fold axis of the megasynthase and is connected through a structured linker to AT which is located on the periphery. The only difference is the presence of a helical insertion in the KS-AT linker in PKS that does not influence the relative positions of the KS and AT domains. The structure continues with a ~20 residue linker which account for the connection between the condensation and the β-carbon processing part of the megasynthase. This feature is well conserved between mFAS and PKSs.

To date, only the structure of mFAS [46] describes the organization of the full β-carbon processing part (DH-ER-KR) from a megasynthase. However, a structure of an ER-KR didomain from the module 2 of the spinosyn PKS is available and show important differences in the organization of the domains compared to mFAS [71]. On the one hand, in mFAS, the ER domain dimerizes along the 2-fold axis of the megasynthase and forms interactions with both DH and KR domains. On the other hand, ER-KR didomain structure from the spinosyn PKS, reveals a monomeric ER domain (**Figure 11B**) and a completely different interface between ER and KR domains (**Figure 9B**). These preliminary data show that different architectures may be possible among the β-carbon processing part of megasynthases. It has also been suggested that the mFAS architecture would not be compatible with the multi-modular PKS mechanism [71]. Indeed, in multi-modular PKSs, ACP domain is tethered to a dimeric domain, either a downstream KS or a terminal TE (in contrast

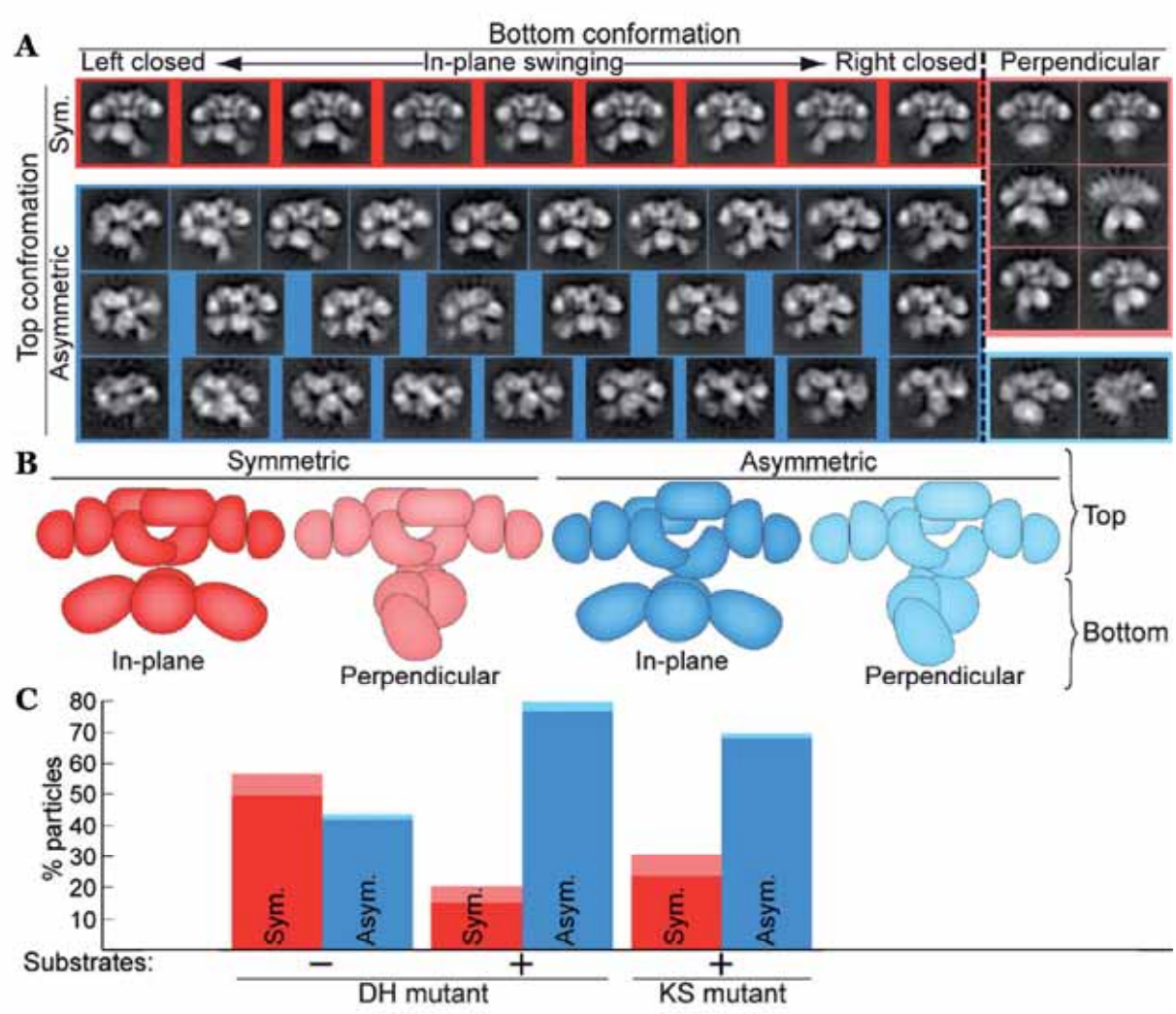


Figure 13. Rat FAS conformations and correlation with catalytic activities.

(A) Classification of electron microscopy single particles into 44 classes. Further categorization has been applied for symmetric (red) and asymmetric (blue) particules with either perpendicular (faded colors) or in-plane configurations of the lower part of the FAS structure. (B) Cartoon representation of each category colored according to A. (C) Percentage of particles for each category and for different FAS versions with or without substrates. The strongest correlation between a conformation (asymmetric) and a represented activity is observed for the DH mutant. **Adapted from Brignole *et al.*, 2009**

to the monomeric TE in mFAS) and is then less mobile than its mFAS counterpart, a feature that would not allow it to access all the β-carbon processing domains. This ACP-tethering is obviously not observed in mono-modular PKSs, which can then possibly adopt the same architecture than mFAS.

In any case, ACP domain, with help of the flexible KR-ACP linker and its post-translationally added P-pant arm, seems to possess enough mobility to reach other domains active sites without significant architecture reorganization. However, some of these sites are distantly located and are not present on the same face of the mFAS "X" architecture (**Figure 12**). Additionally, earlier studies had provided biochemical evidences indicating that the ACP domains from FASs [78], [79] and PKSs [73], [80], [81] were able to make functional contacts with the KS and AT domains from both monomers of the megasynthase. Thus, overall architecture flexibility is required for the ACP to efficiently shuttle the substrate to the other domains in both mFAS and PKSs.

Such flexibility has been studied in the case of the rat FAS using single particle electron microscopy and a combination of different catalytic mutations and substrates [77]. It reveals the pliability of FAS that result in multiple conformations (**Figure 13A**). These conformations have been classified using three main movement criteria: (1) in-plane rotation between the top and bottom portions of the structure, (2) re-organization resulting in progressive asymmetry in the upper half of the structure, and (3) off-plane rotation between the top and bottom portions. This work suggests a direct correlation between conformation and specific enzymatic activities. It provides direct evidence that FAS can adopt a conformation in which the lower part of the structure is rotated 80-100° relative to the upper part of the structure. Thus, rotation of ~90° in either direction would allow ACP domain to interact preferentially with KS or AT domains depending on the rotation direction and would restore the stable "X" shape conformation seen in the crystal structure of porcine FAS. Using inactive mutants of the KS and DH activities and their respective substrates, the different classes of conformations can also be correlated to the condensation and dehydration activities (**Figure 13**). Furthermore, rotation motions are apparently coordinated with rearrangement of β-carbon processing domains, possibly resulting in asynchronous chain extension and β-carbon processing reactions between the two

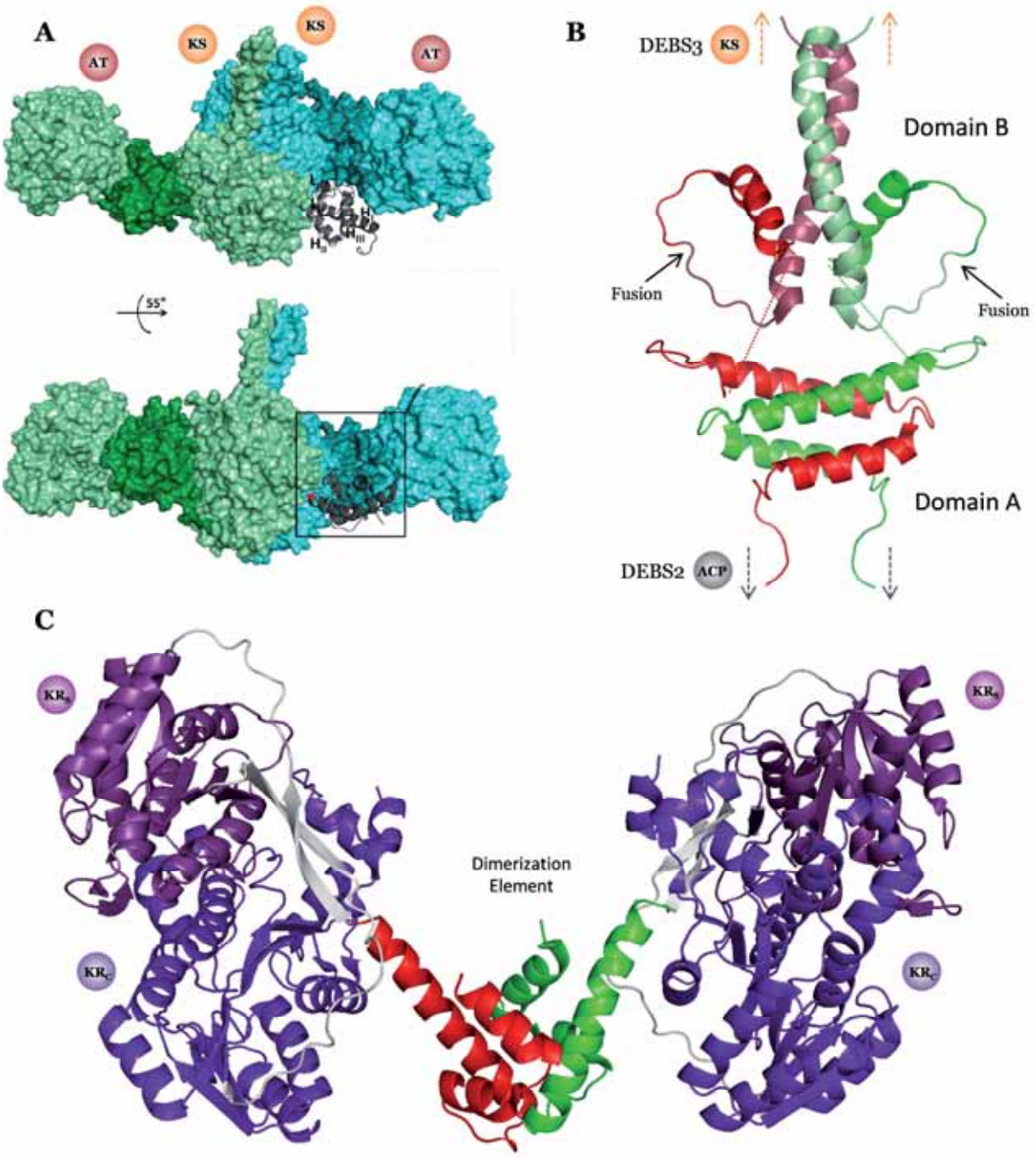


Figure 14. Linkers and structural motifs.

(A) KS-AT/ACP docking model from DEBS module 5 showing a possible binding mode for the ACP domain which occurs via the KS-AT linker (dark green and dark cyan). ACP domain is represented in black cartoon. **Adapted from Kapur *et al.*, 2010** (B) Schematic representation of juxtaposed NMR structures of domains A and B from DEBS modules 2-3 docking domains fused together (PDB codes: 1PZQ, 1PZR). Monomer A is red and monomer B is green. Bright colors are for module 2 and fade colors are for module 3 which are normally coded on two different polypeptides. Engineered protein fusions between modules 2 and 3 are indicated by dark arrows. (C) Crystal structure of a dimeric DE-KR from spinosyn PKS module 3 (PDB code: 4IMP). DEs are colored green and red depending on the monomer. KR_S are colored pink and KR_C are colored violet.

reaction chambers formed by the two monomers. FAS flexibility has been estimated to contribute ~20% to the overall rate of fatty acid biosynthesis [78].

(4) LINKERS AND OTHER STRUCTURAL MOTIFS

Linkers have long been considered as semi-flexible tethers that hold adjacent domains in proximity [82]. In contrast, KS-AT linker has been recently referred to as "not just a swinging arm" [40]. Indeed, along with post-AT linker, KS-AT linker plays an important structural role in fixing the relative positions of the KS and AT domains using both hydrophobic and hydrogen bonding interactions. The KS domain, the KS-AT linker, the AT domain, and the post-AT linker are packed together in an apparently rigid arrangement that precludes significant movement of the KS and AT domains relative to one another. Based on recent biochemical [83] and molecular docking [58] studies, KS-AT linker and post-AT linker have been proposed to play a crucial role in ACP binding for subsequent KS and AT activities in both FASs and PKSs (**Figure 14A**).

Some linkers are also specific to PKSs. This is the case of sequences located at the N- (~35-residue long) and C-termini (~90-residue long) of modular PKSs. These linkers have been demonstrated to form compatible pairs that play a role in the specificity of inter-modular interactions between modules from different polypeptides and have been named "docking domains" [82]. In addition to the partially specific recognition between ACP (module n) and KS (module n+1), these docking domains provide specific interactions in order to stream the growing polyketide to the correct module. NMR studies have helped in identifying the module specificity behind these docking domains [84], and the solution structures of compatible docking domains from DEBS modules 2 and 3 has also been solved [85].(**Figure 14B**). 6 helices provided by the C-terminal docking domain of the dimeric DEBS module 2, and 2 helices provided by the N-terminal docking domain of the dimeric DEBS module 3 assemble together into a complex. The structure reveals that all eight helices are required to stabilize the homodimeric PKS structure.

Other structural motifs are thought to play a role in stabilizing some PKSs dimers. As observed in mFAS, apart from the KS domain, the β-carbon processing domains DH and ER are able to mediate dimerization of the megasynthase. In PKSs, when the β-carbon processing part is composed by the monomeric KR domain only, the KS domain remains the sole domain able to mediate dimerization. In this context, a motif has been identified in ~50% of these KR-reducing PKS modules that would aid KS domains in dimerizing the PKS and has been therefore called a "dimerization element" (DE). The DE from the spinosyn PKS module 3 has been structurally determined (**Figure 14C**) [86]. A model representing a module containing the KS-AT-KR-ACP domains from a multi-modular PKS has also been constructed, with help of the KS-AT didomain [40] and ACP domain [54] structures. Interestingly, the previously reported structures of the DH domain [60] (**Figure 10A**) and the ER-KR didomain [71] (**Figure 9B**) from multi-modular PKSs are compatible with this model. Therefore, a similar architecture for multi-modular PKSs possessing a full β-carbon processing part was envisioned by the authors [71]. They also propose that a putative interaction between the KS domain and the structural KR domain (KR_S) would explain the absence of this DE in the remaining ~50% of the KR-reducing PKS modules [86].

2. ACP ACTIVATION AND 4'-PHOSPHOPANTETHEINYL TRANSFERASES

Polyketide as well as fatty acid and non-ribosomal peptide intermediates, are not covalently bound to an amino-acid residue on ACPs but are instead tethered via a thioester linkage on the post-translationally added P-pant moeity. This residue extends the reach of the CP to present the polyketide intermediates it shuttles to the other domains. The PPTase enzyme superfamily, responsible for the transfer of P-

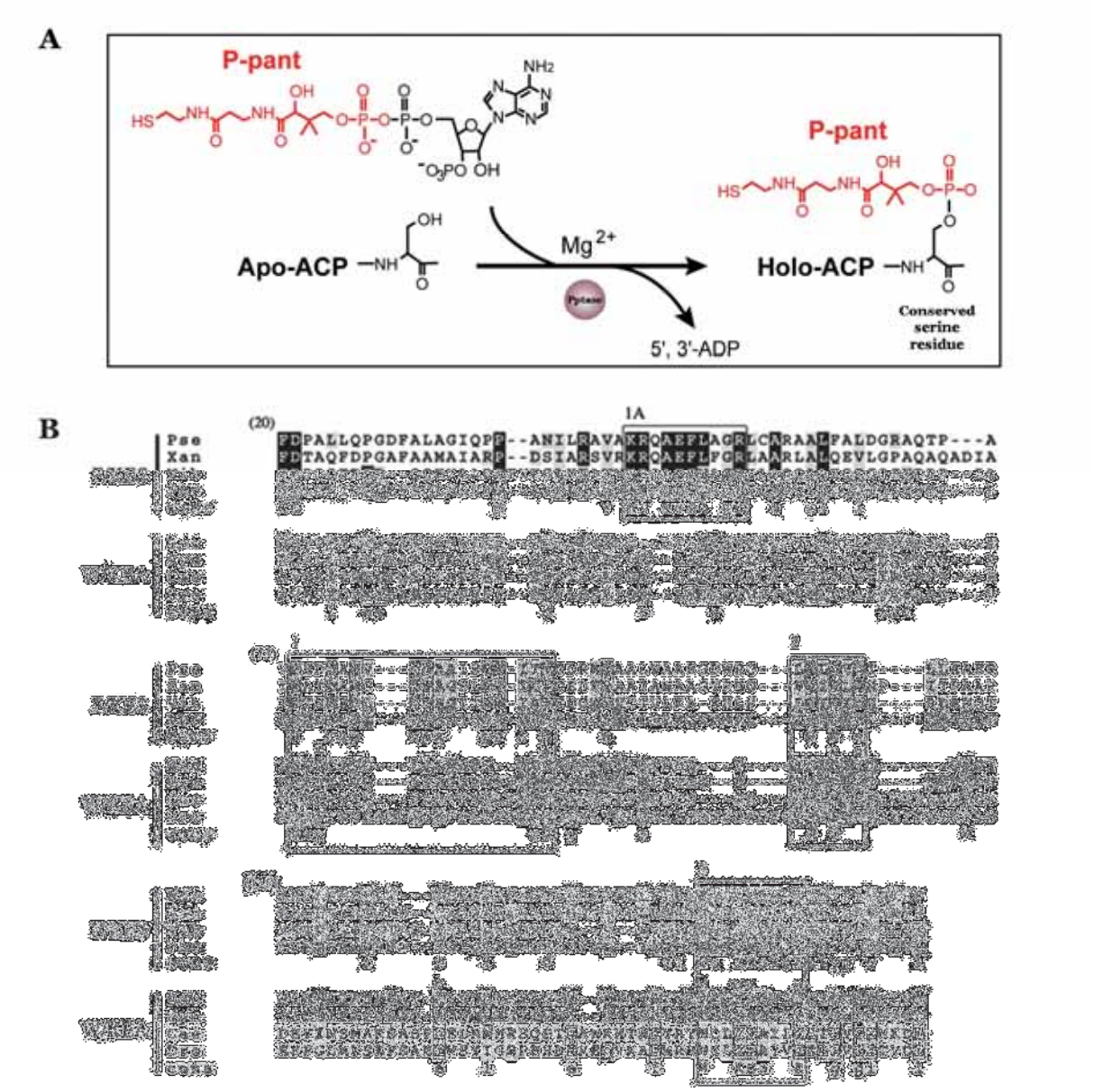


Figure 15. P-pant transfer and group II PPTases sequence motifs.

(A) Reaction of P-pant transfer catalyzed by a PPTase with CoA and Mg2+ as co-factors. P-pant is transferred onto a conserved serine residue on an ACP thus turning it from its apo form into its holo form. (B). Box shade alignment of "Sfp-like" PPTase family representatives. Black shading displays identical residues, and gray shading depicts similar residues. Two subfamilies are shown. The F/KES (top alignment) and W/KEA (lower alignment) sequences are separated with a consensus (cons) line shown below each subfamily. Sequences include the following: Pse, *Pseudomonas aeruginosa*, AAG04554; Xan, *Xanthomonas albicans*, AAG28384; Vib, *Vibrio cholerae*, AAD48884; Pho, *Photorhabdus luminescens*, AAK16071; Bac, *Bacillus subtilis*, P39135; Syn, *Synechocystis PCC6803*, BAA10326; Cae, *Caenorhabditis elegans*, A89451; Dro, *Drosophila melanogaster*, AAM11059. PPTase motifs are boxed and numbered. Numbering for the Sfp PPTase from *Bacillus subtilis* is shown in brackets, and an asterisk indicates residues implicated in stability or activity roles. **Figure and legend from Copp et al., 2009**.

pant to CPs, has been described in the mid-1990s by the Walsh group who has later greatly contributed to their characterization [17].

a) Transfer of the 4'-Phosphopantetheine Moiety

PPTases catalyze the transfer of the essential prosthetic P-pant from CoA to an invariant serine residue contained within the conserved sequence motif Gx(D/H)**S**(L/I)(D/K) of all CPs [87] (**Figure 15**). The transfer results in a phosphoester linkage of the P-pant moiety onto the solvent-exposed serine and provides a 20 Å flexible arm to the CP. It allows the thioester attachment of the growing intermediate at the opposite side of the flexible arm to present it to the other catalytic domains.

Along with CoA, a divalent metal ion is required for the transfer. To date, three different cations have been found to allow the transfer of P-pant. Mg2+ and Ca2+ ions have been identified in structures of PPTases produced in *E. coli* and have been demonstrated as essential for the *in vitro* transfer of P-pant [88], [89], [90]. In the case of the *Bacillus subtilis* Sfp, Mn^{2+} allows for an enhanced catalytic activity [91] compared to Mg^{2+} [88]. However, whether the metal ion is necessary for binding of CoA remains unclear [92].

b) PPTases Classification

PPTases can exist as distinct proteins or be part of the megasynthase they activate [93]. They also possess different specificity toward CPs. Enzymes that activate primary metabolism ACPs (fatty acid biosynthesis) are referred as Acp synthases (AcpS) whereas PPTases involved in secondary metabolism (polyketide and non-ribosomal peptide biosynthesis) are Sfp-like enzymes. However, some PPTases

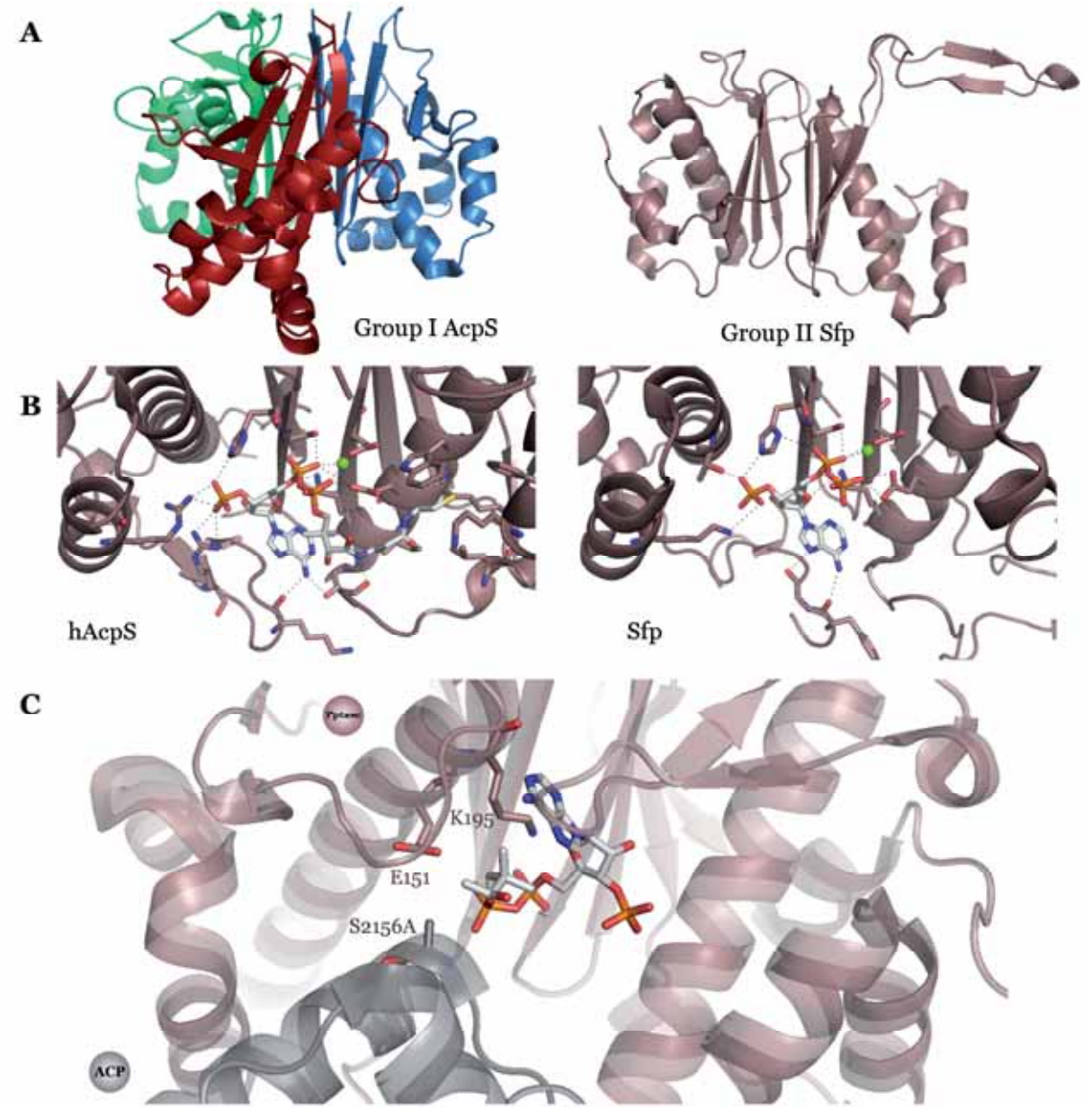


Figure 16. PPTase groups I and II, CoA-binding and hAcpS/hACP complex.

(A) Trimeric arrangement of Group I AcpS from *B.subtilis* (PDB code: 1F7T) compared to the pseudo-dimeric fold of *B.subtilis* Sfp (PDB code: 1QR0) showing a similar organization of the (pseudo-)dimer. (B) CoA-binding in hAcpS (left panel) and Sfp (right panel). Stabilizing residues and CoA are represented in sticks, Mg^{2+} is represented as a green sphere. In hAcpS, the P-pant wass buried in a hydrophobic cavity whereas it wass not visible in the Sfp crystal structure.(C) Structure of the ternary complex involving human PPTase AcpS (hAcpS), CoA and FAS-I human ACP (hACP, PDB code: 2CG5). hAcpS is colored brown, hACP is colored gray. CoA and catalytic residues are depicted in sticks. It is noteworthy that the ACP serine has been mutated to alanine in order to inhibit P-pant transfer and stabilize the complex.

show broader specificity and can activate CPs from both FASs and PKSs [94]. Despite the absolute requirement for PPTases in a wide range of important biosynthetic pathways, these enzymes have remained elusive for a long time due to their low sequence identity and lack of proximity to their respective biosynthetic clusters. The absence of characterized PPTases has hampered many efforts to produce polyketide and nonribosomal peptide products in heterologous host systems utilizing *E. coli*, largely due to the inability of *E. coli* PPTases to activate foreign substrates.

We are now able to identify and classify PPTases on the basis of sequence motifs and structural features [95]. PPTases partition into three major groups, depending on their length, structural architecture and CP substrate specificity. Group I is represented by bacterial transferases of ~120 residues, acting on the ACP of bacterial FASs. Structure determinations of AcpS from *B. subtilis* [90] and *Streptococcus pneumoniae* [96] reveal a trimeric arrangement of an α/β fold domain containing a $\beta_1\alpha_3\beta_2$ motif (**Figure 16A**). The active site is located in the cleft between the different subunits, resulting in three active sites per trimer.

Group II PPTases are exemplified by the monomeric ~240 residues Sfp protein from *B. subtilis*, which transfers a P-pant moiety on the non-ribosomal PCP domain of bacterial surfactin synthetase [88], [91]. This group can be identified using specific sequence motifs (**Figure 15B**). It may have evolved from the AcpS family by gene duplication [89] with subsequent diversification into two sub-families [87]. Indeed, the structure determination of Sfp [97] reveals a two-domain enzyme with intrinsic pseudo 2-fold symmetry, however, with similar domain architecture as the trimeric AcpS (**Figure 16A**). Sfp is particularly interesting for its broad specificity toward both CoA-derivatives and CPs. Indeed, it is able to transfer prosthetic groups from non-physiological substrates such as acetyl-CoA or benzoyl-CoA [88] and has been early used for polyketide or non-ribosomal peptide production in heterologous hosts [98], [99].

Finally, Group III PPTases comprise transferases that are an integral part of the heteromeric yeast and fungal FAS, and probably share the same trimeric arrangement than Group I PPTases [100]. However, no structure of a group III PPTase is yet available although the structural determination of the fungal FAS from *Thermomyces lanuginosus* brought light onto its possible location inside the megasynthase [101].

c) Coenzyme A Binding and Catalytic Mechanism

In Sfp [97], CoA is bound in a cleft formed by the two domains, and the structure reveals basic principles of Mg^{2+} and CoA binding. CoA is bound in a bent conformation within a pocket delineated at the bottom by residues from strands β2 and β4. No defined conformation is observed for the pantetheinyl moiety of CoA (apart from the β-phosphate) in the electron density map probably because it does not interact with Sfp. This feature is in good agreement with the broad-substrate specificity of Sfp, since the pantetheinyl part can be replaced or fused to an acyl without disturbing the CoA-derivative binding in the Sfp pocket [88]. In contrast, the ADP moiety and the pyrophosphate are stabilized by a myriad of polar interactions involving a dozen residues from Sfp (**Figure 16B**). Both α- and β-phosphates participate in the complexation of an Mg^{2+} ion whose additional ligands are the carboxylate groups of Asp107, Glu109 and Glu151 and one water molecule. Interestingly, this CoA binding mode is not universal among group II PPTases. Indeed, in contrast to Sfp, the pantetheine moiety is buried in a hydrophobic cavity in the human PPTase (hAcpS) structure [92] (**Figure 16B**), that probably enhances the specificity toward CoA and prevents CoA-derivatives from binding.

The structure of a ternary complex involving hAcpS, the ACP domain of human FAS (hACP), and CoA has been determined and a detailed mechanism describing the substrate binding and catalytic process has been proposed [92] (**Figure 16C**). In this complex the hACP catalytic serine had been mutated to an alanine (S2156A) to inhibit P-pant transfer and facilitate the observation of the hAcpS-hACP complex. The catalytic mechanism would start with the abstraction of a proton from the hydroxyl group of the hACP catalytic serine S2156 by the neighboring E151 from hAcpS. The resulting hydroxylate would then carry out a nucleophilic attack on the CoA β-phosphate, and after protonation by K195, the pyrophosphate is cleaved, and the products dissociate from the enzyme. The putative role of E151 and K195 as key acid/base catalysts is also strongly supported by the observed significant

loss of activity for the associated alanine mutant proteins for both residues. This proposed mechanism for human PPTase is similar to the one suggested for Sfp [91] but significantly different from the group I AcpS mechanism, where a metal-ion-activated water molecule is proposed to be responsible for the abstraction of the proton from the acceptor hydroxyl group [90].

PPTases are potential drug targets since they are required to enable any PKS-, FAS- or NRPS-based biosynthesis. They have been shown to be essential for the survival of numerous organisms [96], [102], [103], [104].

3. PKS ROLE IN PRODUCING ORGANISMS

Complex polyketide biosynthesis involving several PKSs and other enzymes to further modify, transport and possibly excrete polyketides, is energy demanding. For example, to produce erythromycin [105], *Saccharopolyspora erythraea*, devotes some 60 kb of its DNA to make this macrolide from propionate units by an assembly-line process involving no less than 28 active sites arranged along three PKSs, followed by hydroxylation and glycosylation involving 18 further proteins, and then has to protect its ribosomes from the highly specific toxicity of the antibiotic by an equally specific methylation of a site on the ribosomal RNA . Evolution has then mostly selected polyketide-synthesizing complex machineries that represent a competitive advantage for their producing organisms [106], [107]. Indeed, the ability to produce erythromycin gives *S. erythraea* a strong selective advantage in nutrient poor environments by poisoning neighboring organisms.

Polyketides possess a wide range of activity including inhibition or killing of other microorganisms but also toxicity to multicellular organisms, hormone-like roles, metal transport, virulence of pathogens, photo-protection, coloration, extra-cellular communication and probably even more activities that remain to be characterized [107].

The actinomycetes, a diverse order of Gram-positive bacteria which include *S. erythraea*, synthesize most of all known polyketides. These are often secondary metabolites that give a selective advantage to these non-motile bacteria over other species [106], [108], [109]. Inside the actinomycetes order, *Mycobacterium tuberculosis* (*Mtb*) has chosen another path since it does not produce any secondary metabolites, instead, *Mtb* uses polyketides to enhance its virulence. Indeed, an important part of the *Mtb* virulence is mediated by polyketide-derived lipids or at least lipids whose biosynthesis involves PKSs. In a context of growing appearance of multidrug- and even totally drug-resistant *Mtb* strains, this feature makes PKSs very promising targets for the development of new drugs against this wide-killer bacterium.

C. TUBERCULOSIS

1. TUBERCULOSIS: THE DISEASE

a) History and State of the Art

"Tuberculosis" (TB) is a very old disease, whose history is closely related to that of mankind. Its causative agent, *Mycobacterium tuberculosis*, have killed more persons than any other microbial pathogen [110]. It is presumed that an early progenitor of *Mtb* was probably contemporaneous with early hominids in East Africa, three million years ago, and co-evolved with them [111]. TB was documented in Egypt, India and China as early as 5,000, 3,000, and 2,300 years ago, respectively. Mycobacterial DNA was detected in a human skeleton from the Iron Age (400-230 BC) found in the United Kingdom [112]. Around 460 BC, Hippocrates used the term "phthisis" to describe the wasting away of patient's body caused by the most widespread pulmonary disease in Ancient Greece. The beginning of the 17th century coincided with the arrival of an epidemic known as the "Great White Plague" (for the extreme paleness of people suffering from TB) in Europe that continued for the next 200 years. At that time, precise pathological and anatomical descriptions of the disease began to appear. In his "*Opera Medica*" (1679), Franciscus Sylvius de la Böe was the first to identify the presence of tumor-like nodules in the lungs and other areas of the patients, the so-called "tubercles". But it was not until 1834 when Johann Lukas Schölnlein established the current name "tuberculosis" to describe diseases with tubercles.

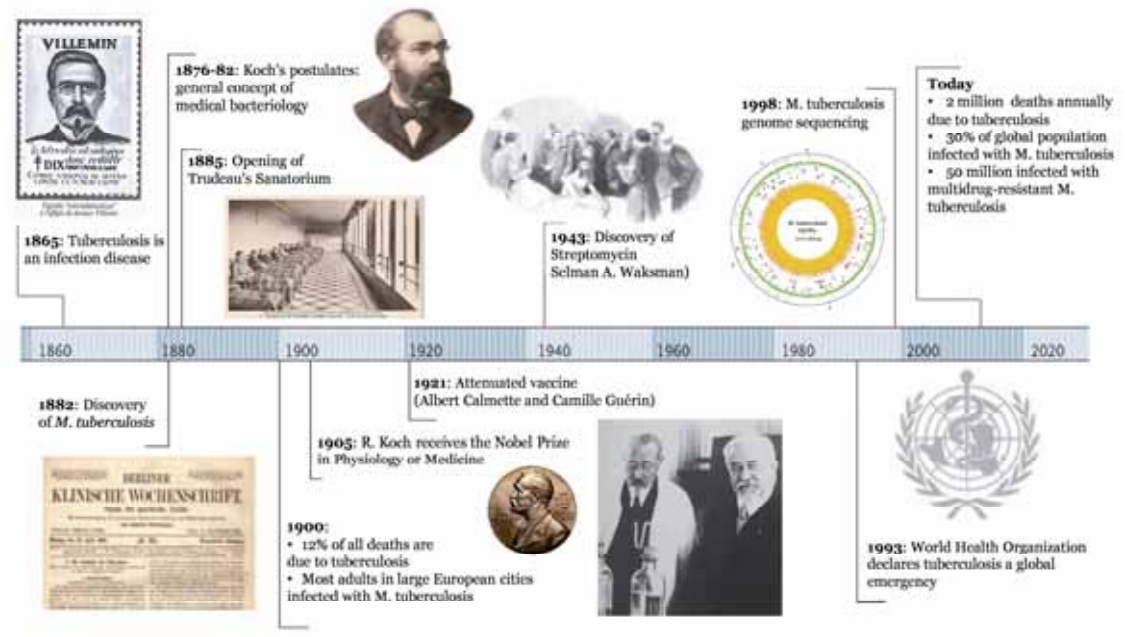


Figure 17. History of tuberculosis.

Adapted from Kaufmann et al., 2005.

In 1865, the French military surgeon Jean Antoine Villemin demonstrated that tuberculosis could be transmitted from humans or cattle to rabbits (**Figure 17**). In the light of this revolutionary evidence, he postulated that a specific microorganism caused the disease. On March 24th 1882, Robert Koch made his famous presentation "*Die Äetiologie der Tuberkulose*". He demonstrated the infectious nature of TB. The introduction of sanatorium cures in the late 19th century provided the first widely practiced approach to anti-TB treatment. In 1854, Hermann Brehmer presented his medical dissertation "*Tuberculosis is a curable disease*" and then opened an in-patient hospital where patients were continuously exposed to fresh air. He emphasized a regimen of rest, rich diet, and exercise with encouraging results [113]. This became the model for all subsequent sanatoria. With Edward Jenner's successful discovery, showing that infection with cowpox would give immunity against smallpox in humans, many doctors placed their hopes on the use of *Mycobacterium bovis* – the agent that causes bovine TB – for the development of a vaccine against human TB. From 1908 until 1919, Albert Calmette and Camille Guérin serially passed a pathogenic strain of *M. bovis* 230 times, resulting in an attenuated strain called "Bacille Calmette-Guérin" (BCG). The resulting strain was avirulent, but still able to stimulate the immune system of different animal models, thus protecting them against TB infection. BCG was first administered to humans in 1921 and it was still widely used until the beginning of the 21th century [114]. Then, in the middle of World War II, antiobiotics represented the final breakthrough in the treatment of the bacterium. In 1943, streptomycin, a compound with antibiotic activity, was purified from *Streptomyces griseus* by Selman A. Waksman, who obtained, in 1952, the Nobel Prize in medicine for this discovery. When this therapy was used on TB patients a considerable improvement was observed. The discovery of a series of new anti-TB drugs: para-aminosalicylic acid (1949), isoniazid (1952), pyrazinamide (1954), ethambutol (1962) and rifampicin (1963) further enhanced the treatment. Unfortunately, this initial optimism was soon tempered by the development of the first resistant strains to anti-TB drugs, although it was later demonstrated that cocktails treatment involving several drugs could overcome this problem [113].

General improvement in public health and success of antibiotic therapy resulted in a pronounced reduction of infection and death rates in Europe and in the USA. In the 1980s, elimination of TB was thus envisioned to 2010 [CDC, A strategic plan for the elimination of tuberculosis in the United States]. Nevertheless, around

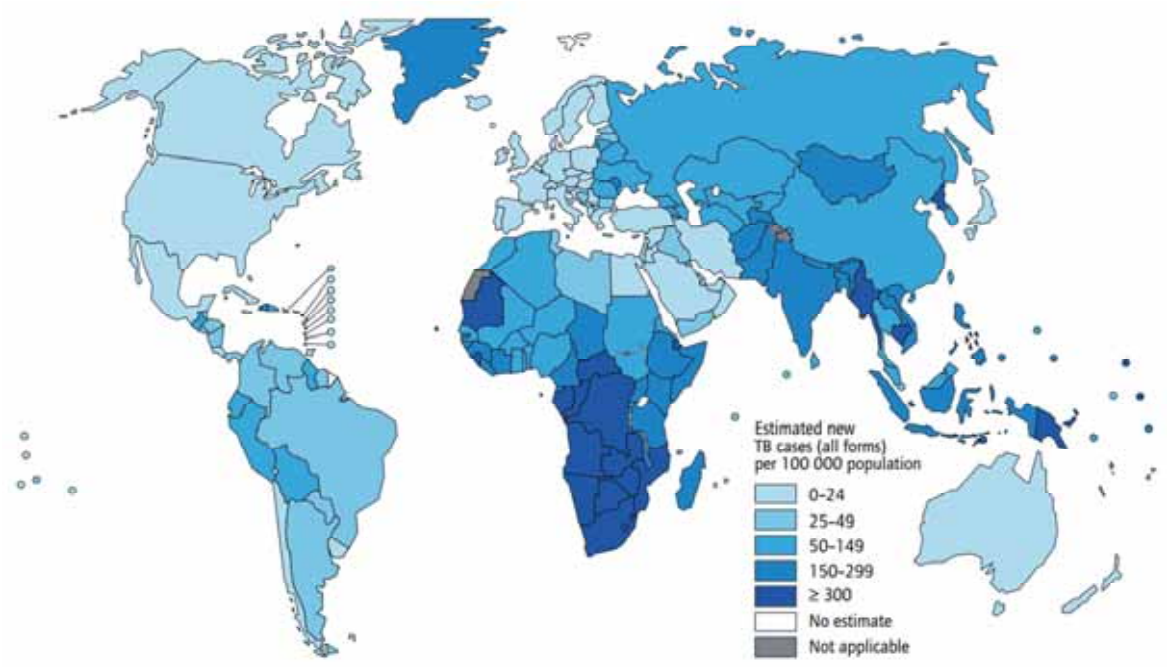


Figure 18. Estimation of new TB cases per 100000 population per year.

More than 80% of cases are reported in Asia (59%) and Africa (26%). The 5 most concerned countries are India, China, South Africa, Indonesia and Pakistan. **From WHO, 2012**

1985, cases of TB began to rise again in industrialized countries. The decline in TB control activities – TB programs had become loose in industrialized countries because the disease was considered close to elimination – and the acquired immunodeficiency syndrome (AIDS) epidemic were the major factors in the reemergence of TB. The World Health Organization (WHO) declared TB a global health emergency in 1994 [WHO, TB: a global emergency).

TB is today the second major cause of death from any infectious disease, and furthermore, the first cause of death coming from a unique pathogen. In 2012, 8.7 million new cases of TB occurred and 1.4 million people died, and it is estimated that one third of population is infected with *Mtb* [WHO, Global Tuberculosis Report, 2012]. The worldwide tuberculosis incidence rates are estimated to have peaked in 2004 and to have decreased at a rate of less than 1% per year since that time. Most cases are in Asia and Africa, with smaller proportions of cases in the eastern Mediterranean region, European region, and the Americas (**Figure 18**). Tuberculosis remains a disease of poverty that is inextricably associated with overcrowding and under-nutrition.

b) Prevention and Treatment

(1) PREVENTION

The primary worldwide mean of fight against tuberculosis, whose first use occurred in 1921, was the BCG vaccine. It has been given to 4 billion people so far and to more than 90% of the children in the world today, making it the most widely used vaccine in the world. Nevertheless, immunity it induces to TB, decreases after about ten years. While it is effectively treating disseminated forms of TB (meningitis and military) in childhood, its protection against contracting pulmonary TB in both childhood and adulthood is rather inconsistent (**Table 2**) and it has done little to contain the tuberculosis pandemic [115]. However, an effective vaccination regimen

Study	Vaccine Efficiency
Norway	81%
Denmark (school)	94%
Ireland (school)	82%
Canada (Indians)	81%
United Kingdom	75%
United Kingdom (Asians)	49%
USA, Chicago (infants)	72%
USA, New York (infants)	7%
Korea, Seoul	74%
Argentina, Buenos Aires	73%
USA, Georgia-Alabama	16%
Israel (children)	38%
South Africa (miners)	62%
Australia, Queensland	41%
Puerto Rico	29%
Haiti	80%
Burma, Rangoon	38%
Thailand, Bangkok	74%
India, Madanapalle	20%
Indonesia, Jakarta	37%
Columbia, Cali	16%
Kenya, Kisumu	22%

Table 2. BCG vaccine efficiency in different regions of the world.

Adapted from Fine, 1995.

would be the most efficient way to control the epidemic and this strategy has not been abandoned. New vaccine candidates are indeed being tested (for review [116]). Nowadays, the most powerful treatment remains by far, the antibiotic therapy. It has been used since the early 1950's and further developed with the discovery of novel classes of antibiotics and the generation of drug cocktails.

(2) FIRST-LINE TREATMENT

Current TB treatment requires a combination of drugs and obeys to two phases: an initial intensive phase of treatment (2 months) with isoniazid (INH), rifampicin (RIF), pyrazinamide (PAZ) and ethambutol (EMB) designed to kill actively growing and semi-dormant bacilli; followed by a continuation phase (4 months) with INH and RIF only, to eliminate most persistent bacilli and reduce the number of relapses [WHO, Treatement of tuberculosis, 2009]. The drugs features are presented in **Table 3**. Interestingly, isoniazid targets the type-II FAS enoyl-reductase which is a functional homolog of type-I PKS ER domains. This illustrates the druggability of this family of enzymes.

Treatment completion is critical to reduce development of acquired drug resistance, therefore, trained personnel supervises that each dose of medication is properly taken by the patient. This strategy is known as directly observed treatment short-course (DOTS) and was promoted as the official policy of the WHO in 1991 [WHO, What is DOTS?].

(3) RESISTANCE AND SECOND-LINE TREATMENT

Another major challenge is the ever-increasing resistance to anti-TB drugs. *Mtb* strains that possess drug resistance have been classified into three categories. Multidrug-resistant TB (MDR-TB) refers to TB caused by *M. tuberculosis* isolates that are resistant to the most effective drugs, INH and RIF. Infection by MDR strains

Drugs	Molecular targets	Biological consequences	Genes implicated in resistance	Frequent side effects
Isoniazid	FAS-II ER	Inhibition of long chain fatty acids and mycolic acids biosynthesis	*katG* (catalase required for activation of isoniazid), *inhA* (promoter and ORF)	Mostly hepatotoxicity
Rifampicin	ARN polymerase β-subunit RpoB	Inhibition of transcription	*rpoB*	Mostly hepatotoxicity
Pyrazinamide	Ribosomal protein S1 RpsA	Inhibition of *trans*-traduction	*rpsA*, *pncA* (pyrazindamidase required for activation of pyrazinamide	Hepatotoxicity, arthralgia
Ethambutol	Arabinosyl transferases EmbB, EmbC	Inhibition of cell wall biosynthesis	*embB*, *embC*	Arthralgia, optic neuritis

Table 3. First-line drugs used in TB treatment

Drugs	Molecular targets	Biological consequences	Genes implicated in resistance	Frequent side effects
Aminoglycosides (kanamycin, amikacin, streptomycin)	ARN 16S	Inhibition of translation	*rrs* (ARN 16S), *rpsL* (ribosomal protein S12), *gidB* (ARN16S methyltransferase)	Ototoxicity, kidney damage
Cyclic peptides (capreomycin, viomycin)	ARN 16S, ARN 23S	Inhibition of translation	*rrs*, *tlyA* (ARNr methyltransferase)	Ototoxicity, kidney damage
D-cycloserine	Alr (D-alanine racemase), Ddl (D-ala-D-ala ligase)	Inhibition of peptidoglycan biosynthesis	*alr* (D-alanine racemase)	Psychiatric disorders, psychosis convulsions
Fluoroquinolones (ciprofloxacine, ofloxacine)	ADN gyrase	Inhibition of DNA transcription, replication and reparation	*gyrA*, *gyrB* (ADN gyrases), *lfra* (efflux pump), *mpfa* (pentapeptide repeats protein)	Psychiatric disorders, tendon toxicity
Thioamides (ethionamide, prothionamide)	InhA (FAS-II ER)	Inhibition of long fatty acids and mycolic acids biosynthesis	*inhA* (FAS-II ER)	Agranulocytosis, hepatitis, vasculitis, and thrombocytopenia

Table 4. Second-line drugs used in drug-resistant TB treatment

Figure 19. XDR-TB cases report.

XDR-TB infections have been encountered in 84 countries in 2012. **From WHO, 2012.**

is longer to treat (18-24 months) and requires second-line drugs, which are more expensive, less potent and have more adverse side-effects, including hepatitis, depression, and hallucinations (**Table 4**). Extensively drug-resistant (XDR-TB) infection can only be treated with the most efficient second-line drugs, fluoroquinolones and injection drugs such as kanamycin and capreomycin. In 2010, 58 countries had confirmed at least one case of XDR-TB [WHO, Multidrug and extensively drug-resistant TB] (**Figure 19**). Finally, even more alarming is the emergence of deadly "Totally drug-resistant" TB (TDR-TB), defined as XDR-TB additionally resistant to treatment with any fluoroquinolones and any of the second-line anti-TB injection drugs. Although it has been observed in India, Iran, and Italy [117], [118], TDR-TB is poorly documented and so far, remains untreatable.

(4) NEW TREATMENTS

First-line treatment is efficient on non-resistant TB and, except for pyrazinamide, side effects are rather mild and rare [119]. If properly taken, second-line treatment is also efficient on resistant TB but induces side-effects on 70% of treated individuals and forces 56% to withdraw at least one drug from the treatment regimen [120]. While ototoxicity is the most frequent side effects among second line drugs, more crippling ones may appear such as psychiatric disorders, central nervous system damage and leukopenia. Drugs toxicity and treatment duration have become a problem for the DOTS program, as they decrease patient compliance who then tend to quit their regimen thus increasing the possibilities for the bacterium to develop new resistances. Drugs that would shorten the treatment course, possess minor side-effects, and help treat XDR- and TDR-TB are then urgently required. While most of the drugs that compose the arsenal of the TB treatment were discovered during the 1950's and the 60's, new drugs, derived from known compounds, have entered clinical trials in the last decade and might partially respond to this emergency [121] (**Figure 20**). However, there is still a crucial need for new drugs targeting new pathways, thus implying a fundamental comprehension of the mechanisms used by *Mtb* to survive, develop and infect host.

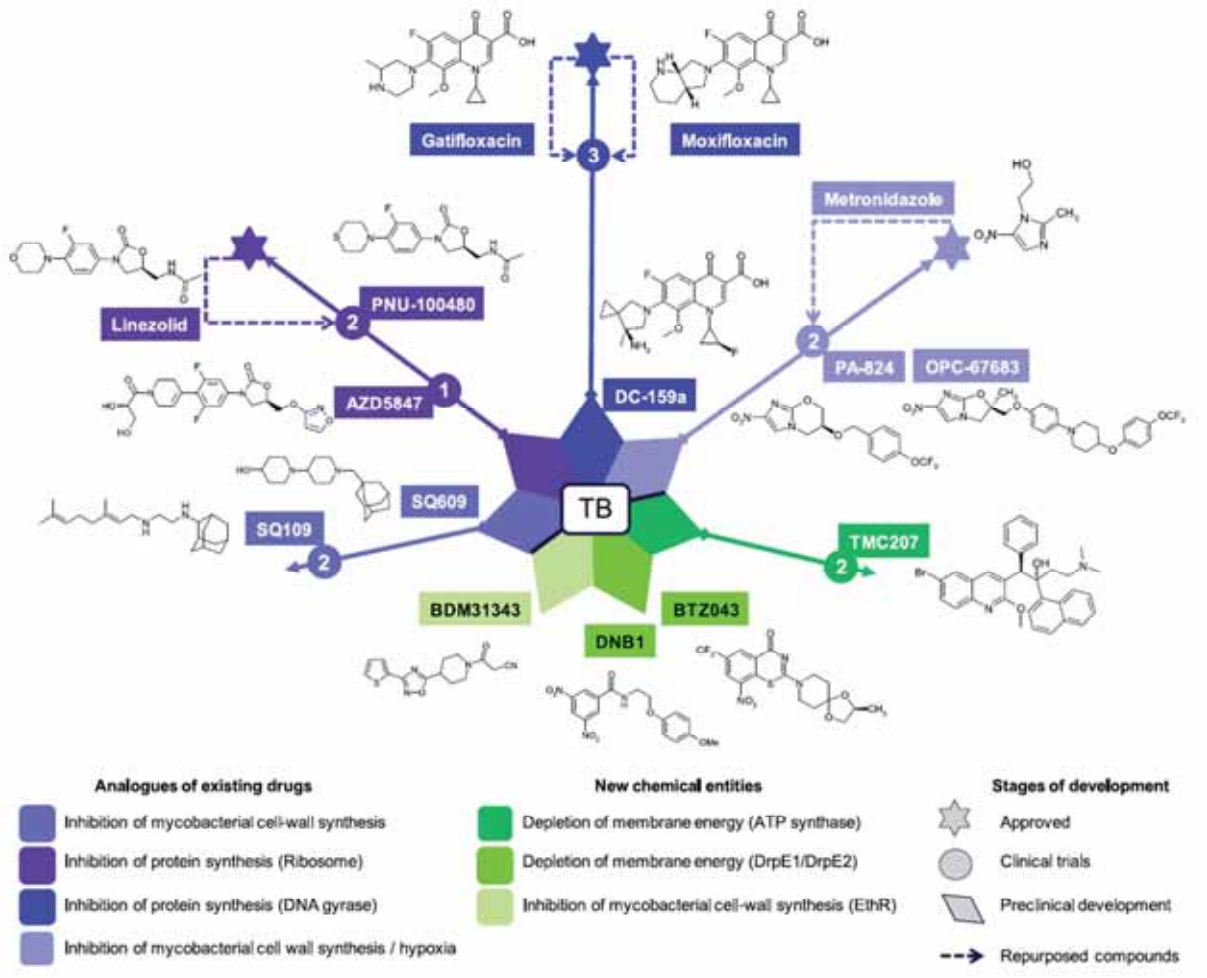


Figure 20. TB drugs in development and their targets.

Numbers 1, 2 and 3 indicate in which clinical trial stage is the tested compound. **From Villemagne *et al.*, 2012.**

2. TUBERCULOSIS: THE BACTERIUM

a) Main Features

Within the G+C rich gram-positive *Actinobacteria* phyllum, the genus *Mycobacterium* belongs to the order of *Actinomycetales,* suborder of *Corynebacterineae*, a taxon characterized by a multilayered cell envelope extremely rich in uncommon lipids [122]. To date, more than 100 species of mycobacteria have been described [123], mostly environmental saprophytes. Phylogenetic analyses and phenotypic characteristics, such as growth rate and pigmentation, can be used to classify these species. The classical distinction between rapid and slow growing species is based on the ability of strains to develop visible colonies in less or more than 7 days, respectively. Interestingly, the slow growing mycobacteria comprise most of the pathogenic species, including *M. tuberculosis*. The various etiologic agents of TB are clustered in the *M. tuberculosis* complex (MTBC), which currently includes seven species: *M. tuberculosis*, *M. africanum*, *M. canettii*, *M. bovis* (primarily a pathogen of cattle), *M. caprae* (a pathogen of goats), *M. microti* (a pathogen of voles) and *M. pinipedii* (a pathogen of seals and sea lions) [124]. Mycobacteria grouped in MTBC are characterized by 99.9% similarity at the nucleotide level and identical 16S rRNA [125]. Phenotypically *M. tuberculosis* is an aerobic bacillus, catalase and nitrate reductase positive, non-motile and nonsporulated of 2-5 μm in length and 0.2-0.5 μm in width.

The first complete genome sequenced was that of *M. tuberculosis* H37Rv [126], the reference laboratory strain in TB research. It comprises 4,411,532 bp with a G+C content of 65.6%. The last re-annotation identified 4,044 genes thought to encode 3,994 proteins and 50 stable RNAs [127]. The most striking feature in the *Mtb* genome is the complex repertoire of genes involved in lipid metabolism which represents about 9% of the coding capacity of the genome. These biosynthesized lipids are encountered in the atypical envelope of mycobacteria.

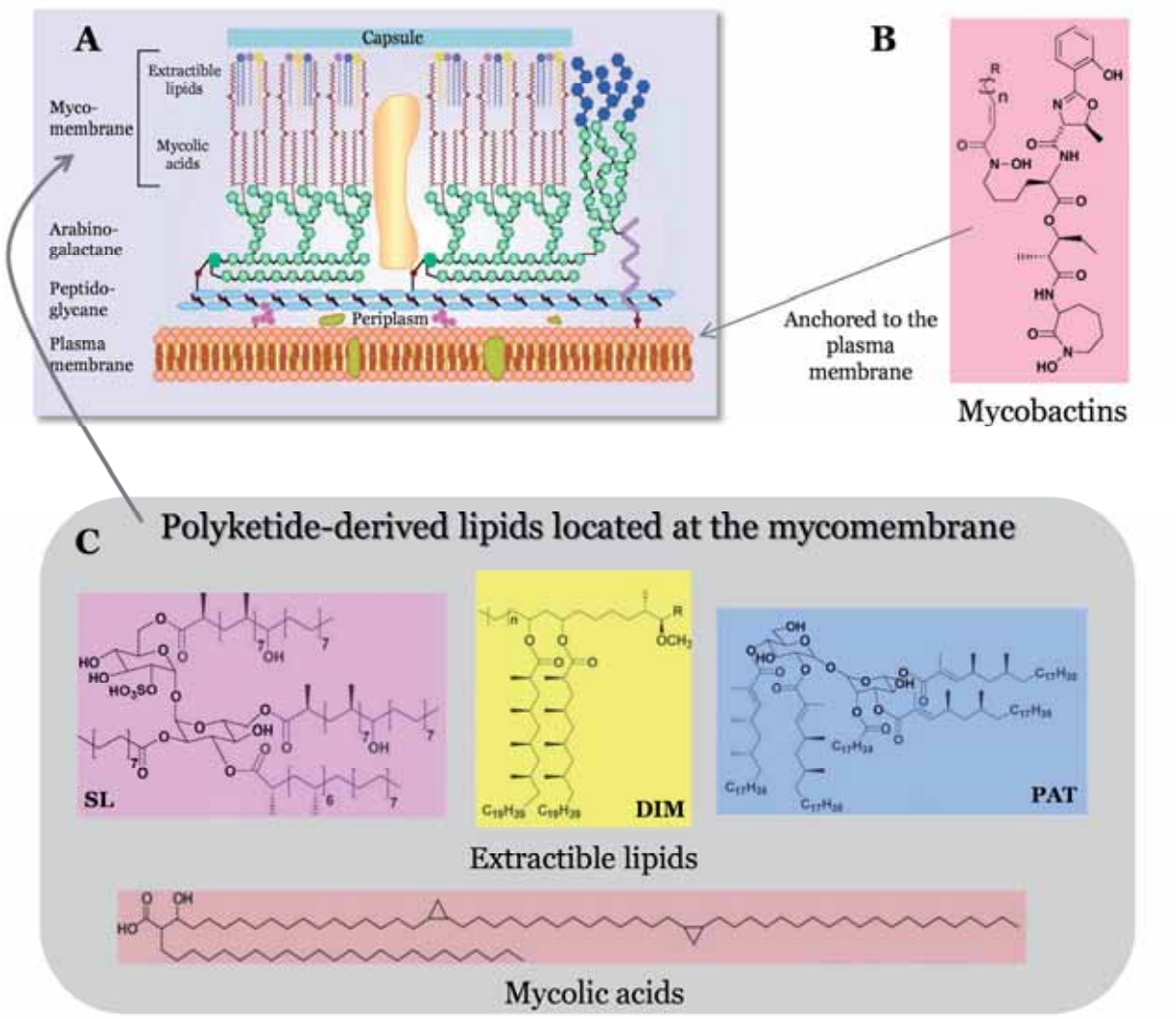


Figure 21. The mycobacterial cell envelope and the structures of its main polyketide-derived molecules.

(A). Schematic representation of the mycobacterial cell envelope. (B) Structure of the mycobactins anchored to the plasma membrane. (C) Structures of the polyketide-derived lipids located at the mycomembrane. SL: sulfolipid; PAT: polyacyltrehalose; DIM: phtiocerol dimycocerosates. **Adapted from Gokhale *et al.*, 2007**

b) The Mycobacterial Cell Envelope

The envelope of *M. tuberculosis* differs considerably from cell walls of both Gram-positive and Gram-negative bacteria. Although they are structurally more closely related to Gram-positive bacteria, mycobacteria do not fit into this category as the molecules attached to the cell wall are distinctively lipids rather than proteins or polysaccharides. The high lipid content of the envelope makes it a natural barrier to most antibiotics and chemical agents. Frequently, they do not retain the crystal violet and appear as "ghosts" after Gram staining. Mtb can rather be observed using the acid fastness staining technique.

The mycobacterial envelope is unique, both in molecular composition and in the architectural arrangement of its constituents. Schematically, the cell envelope of the bacterium, from the inside to the outside, comprises a typical bacterial plasma membrane, a cell wall and an external capsule, which contains proteins, polysaccharides and lipids [128] (**Figure 21**). Recently, new electron microscopy techniques have enabled the identification of a periplasmic space [129]. The cell wall consists of a peptidoglycan (PG) covalently attached to the hetero-polysaccharide arabinogalactan (AG), which is in turn esterified by very long-chain (C60–C90) fatty acids called mycolic acids. The outer layer of the cell wall presents various non-covalently attached lipids and glycolipids esterified with multimethyl-branched longchain fatty acids [130], [131]. These complex lipids are restricted, with few exceptions, to pathogenic mycobacteria; therefore, it has been thought that they may play important roles in the virulence of *Mtb*. In *Mtb*, these lipids include the phthiocerol dimycocerosates (DIM) and the closely related phenolic glycolipids (PGL, not present in all *Mtb* strains), the trehalose ester families that include sulfolipids (SL), diacyltrehaloses (DAT) and polyacyltrehaloses (PAT), and the family of mannosyl-β-1-phosphomycoketides (MPM). Interestingly, the biosynthesis pathways of all these pathogenicity-related lipids, involve PKSs.

PKS	Polyketides	Associated molecules	*M. tuberculosis* expresssion
Pks15/1	*p*-Hydroxyphenylalkanoic acids	Phenoglycolipids (PGL)	Mostly Asian strains
Pks2	(hydroxy)phtioceramic acids	Sulfolipids (SL)	YES
Pks3/4	Mycolipenic and mycosanoic acids	Di- / Poly-acyltrehalose (DAT/PAT)	YES
Pks5	Polymethyl-branched fatty acids	Lipooligosaccharides	NO
Pks6	Unknown	Unknown	Unknown
Pks7	Unknown	Unknown	Unknown
Pks8/17	Monomethyl-branched unsaturated fatty acids	Unknown	Unknown
Pks9	Unknown	Unknown	Unknown
Pks10, 11 and 18	Acyl-α-pyrones	Unknown	Unknown
Pks12	Mycoketides	Mannosyl-β-1-phosphomycoketides (MPM)	YES
Pks13	Mycolic acids	Mycolic acids and trehalose mono- / di-mycolates (TMM/TDM)	YES
Pks16	Unknown	Unknown	Unknown
Mas	Mycocerosic acids	Phtiocerol dimycocerosates (DIM) and PGL	DIM: YES, PGL: mostly Asian strains
PpsA-B-C-D-E	Phtiocerol/Phenol-phtiocerol	Phtiocerol dimycocerosates (DIM) and PGL	DIM: YES, PGL: mostly Asian strains

Table 5. *Mycobacterium tuberculosis* polyketide synthases.

3. POLYKETIDE-DERIVED LIPIDS AND VIRULENCE

Sequencing of *Mycobacterium tuberculosis* H37Rv revealed the existence of 23 ORF potentially coding PKSs [126]. However, several ORF code non-functional enzymes, while others code functional enzymes in only a few *Mtb* strains, and yet others code enzymes that are functional *in vitro* but the corresponding polyketides have never been isolated from mycobacteria. These data are summarized in **Table 5**. In this chapter, only well characterized polyketide-derived lipids (**Figure 21)** involve in *Mtb* virulence are discussed. In addition, the cell wall iron transporters mycobactins, which are also synthesized by PKSs, will be discussed. Structures of the following molecules are presented in.

a) Mycolic Acids and Derivatives

Mycolic acids are essential to the survival of mycobacteria [132] and hence their biosynthesis is the target of widely used drugs such as isoniazid and ethambutol. These α-branched and β-hydroxylated long-chain fatty acids indeed play a major role in the envelope structure. They are located in the inner membrane of the mycomembrane and are covalently attached to the AG, on one side, and interact with other lipids in the outer membrane part, on the other side. In addition to their essential structural properties, mycolic acid derivatives also play a role in the virulence of *Mtb*. Esterified forms of trehalose called trehalose monomycolates (TMM) and trehalose dimycolates (TDM) specifically interact with host cell receptors and induce a pro-inflammatory response from the host immune system [133]

The complex biosynthesis of mycolic acids involves multiple pathways. Briefly, type-I FAS first synthesizes fatty acid precursors, type-II FAS elongates other precursors coming from FAS-I and finally the key step resides in the condensation of the fatty acid chains to yield mycolic acids. This condensation is realized by a

specialized PKS called Pks13, which is essential to mycolic acids biosynthesis [134]. Interestingly, Pks13 is composed by the core KS and AT domains, but also two ACP domains in accordance with the two fatty acid chains to condensate, and a TE domain thought to release the final product. One fatty acid chain is selected by the AT domain [135], and is then transferred onto the C-terminus ACP, whereas the other one is directly loaded on the N-terminus ACP by a dedicated fatty acid AMP ligase (FAAL), FadD32 [136]. Such loading of precursors by a dedicated enzyme is a common feature of several PKSs in *Mtb*.

b) Sulfolipids

Sulfolipids (SL) are multiacylated trehalose-containing lipids that belong to the extractable lipids group of the mycobacterial envelope. SL interact non-covalently with mycolic acids in the outer membrane and their expression is restricted to *Mtb* and *Mycobacterium. canetti*. They are present in the cell envelope of *M. tuberculosis* under different forms. The main form is tetra-acylated with a palmitic acid, a phtioceranic acid and two hydroxyphthioceranic acids. Other forms can be found, especially diacylated form which is an intermediary in the synthesis of the tetra-acylated one. The role of SL in the virulence of *Mtb* was a matter of debate because of the absence of any phenotype *in vivo*, although these lipids seemed to be involved in the survival of the bacillus within macrophages as observed using *in vitro* experiments [137]. However, a recent study has demonstrated a role of these lipids in the virulence of *M. tuberculosis* both *in vivo* in a mouse model, and *in vitro* in human macrophages [138].

The biosynthesis of SL starts by the transfer of a sulfate onto a trehalose molecule, catalyzed by the sulfotransferase Sft0 [139]. In parallel, the polyketide synthase Pks2 synthesizes the methyl-branched phtioceranic and hydroxyphthioceranic chains [140]. The acyltransferases PapA1 & PapA2 are then responsible for the transfer of the acyl chains onto the trehalose-2-sulfate which generate a diacylated SL [141], [142]. Chp1 is responsible for the transfer of the two

remaining acyl chains, and Mmpl8 and Sap are thought to be responsible for the transfer of the SL to the mycomembrane [143].

c) Diacyl Trehalose and Polyacyl Trehalose

Like sulfolipids, diacyltrehaloses (DAT) and polyacyltrehaloses (PAT) are extractable lipids only expressed by *M. tuberculosis* & *M. canettii.* DAT & PAT are lipids constitutionally composed by a trehalose, acylated by a palmitic acid. Other acylations may occur to form a variety of lipids. Thereby, two forms of DAT exist: DAT1 composed by a palmitic acid and a mycosanoïc acid, and DAT2 composed by a palmitic acid and a mycolipanolic acid. DAT & PAT are not involved in the structure of the mycobacterial envelope but seem to play a role in the membrane permeability. It has been demonstrated, both *in vitro* [144], [145] and *in vivo* [138], that DAT & PAT are implicated in the virulence of *M. tuberculosis*, in particular by inhibiting T lymphocytes proliferation. They are also involved in the bacterial survival within human macrophages [146].

The biosynthesis pathway of DAT & PAT remains unclear. PapA3 is thought to transfer a palmitic acid onto a trehalose molecule to form a trehalose-2-palmitate [147]. PapA3 also seems to be involved in the acylation of the trehalose-2-palmitate with the product of the polyketide synthases Pks3/4 to form a diacyltrehalose [148]. Then, this lipid may be exported to the surface of the mycobacterial envelope by Mmpl10 [148], or used as an intermediary in the polyacyltrehalose biosynthesis pathway.

d) Mycobactins

Mtb synthesizes two structurally related siderophore compounds named mycobactin and carboxymycobactin to bind iron from the external medium. The

main functional difference between them is that carboxymycobactin is secreted into external medium whereas mycobactin is associated to the plasma membrane. In host cells, iron concentration is 10^{11} to 10^{12} fold lower than the concentration required for *Mtb* to survive. Mycobactins have been demonstrated to play a major role in such environments [149], [150]. Indeed, *Mtb* mutant strains deficient for the biosynthesis of these compounds hardly multiplicate in these conditions. Also, when either mice or isolated macrophages are infected by *Mtb,* the transcription level of genes coding the biosynthesis pathway of mycobactins is increased [151], [152].

Mycobactins are mixed "polyketides/peptides" compounds and thus are biosynthesized by PKSs and NRPSs. Genes encoding these enzymes are located in the *mtb* 24kb cluster, and include three NRPSs (*mtbB*, *mtbE* and *mtbF*) and two PKSs (*mtbC* and *mtbD*) [126], [153].

e) Phtiocerol Dimycocerosates

(1) A MAJOR ROLE IN MYCOBACTERIUM TUBERCULOSIS VIRULENCE

As early as 1974, it was recognized that a DIM-deficient spontaneously arising variant of the laboratory strain H37Rv was attenuated in a guinea pig model of infection [154]. Two years later, it was shown that the *in vivo* survival of an avirulent *Mtb* strain was enhanced by coating the bacteria with cholesterol oleate and purified DIM [155]. Yet, it took a quarter century to clearly establish a link between DIM biosynthesis and virulence. Indeed, in the late 90's, two independent signature-tagged transposon mutagenesis studies led to the isolation of *Mtb* mutants with a severe growth defect in mice [156], [157]. These transposons have inserted into the large chromosomal locus where enzymes responsible for biosynthesis and transport of DIM are coded, thus leading to a defect in DIM production or translocation. Furthermore, strains deficient in DIM are markedly attenuated for growth in the

lungs of infected mice and in macrophages [156], [158], [159]. DIM have been shown to be responsible for a strong decrease in phagosome acidification [160], and for resistance against both reactive nitric intermediate (RNI) toxic radicals produced by macrophages [159], and INF-γ-dependent host cell mechanism [161]. A more comprehensive analysis of the effect of DIM presence in *Mtb* has revealed that DIM participates in receptor-dependent phagocytosis through a mechanism involving host plasma membrane reorganization [162]. DIM would insert in the plasma membrane to facilitate phagocytosis and inhibit phagosome acidification, thus creating a protecting niche for *Mtb*. Additionally, DIM seem to play an important role in the impermeability of the mycobacterial cell wall since a defect in these lipids lowers mycobacteria resistance to SDS by a 100 fold [158]. Interestingly, in a recent study on the roles of SL, DAT/PAT and DIM in *Mtb* virulence, it has been shown that a hierarchical order exists between these lipids, in which DIM have arisen as the prevalent one [138].

While Euro-American *Mtb* strains, including H37Rv, only produce DIM, highly virulent Asian strains such as *Beijing* strains, produce both DIM and phenoglycolipids (PGL), a structural variant of DIM. PGL production has then been associated with hyper-virulent phenotypes in mice models, since it inhibits several pro-inflammatory cytokines (TNFα, IL-12, IL-6) implied in the regulation of *Mtb* infection [163], [164]. In human macrophage infection, PGL induces the production of IL-4 and IL-13, which are cytokines that deactivate phagocytes. In accordance to this feature, human monocytes incubated with lipid extract not containing PGL, produce IL-12 [165]. Additionally, PGL is also implicated in dissemination and virulence of *Mtb* in the context of central nervous system infection in rabbit meningitis tuberculosis [166].

(2) STRUCTURE AND BIOSYNTHESIS

DIM are composed of a C33-C41 long β-diol chain called phthiocerol, esterified on C9 and C11 by mycocerosic acids which are C27-C31 long methyl-branched chains. As shown in **Figure 22**, different forms of DIM exist depending on (1) phtiocerol and

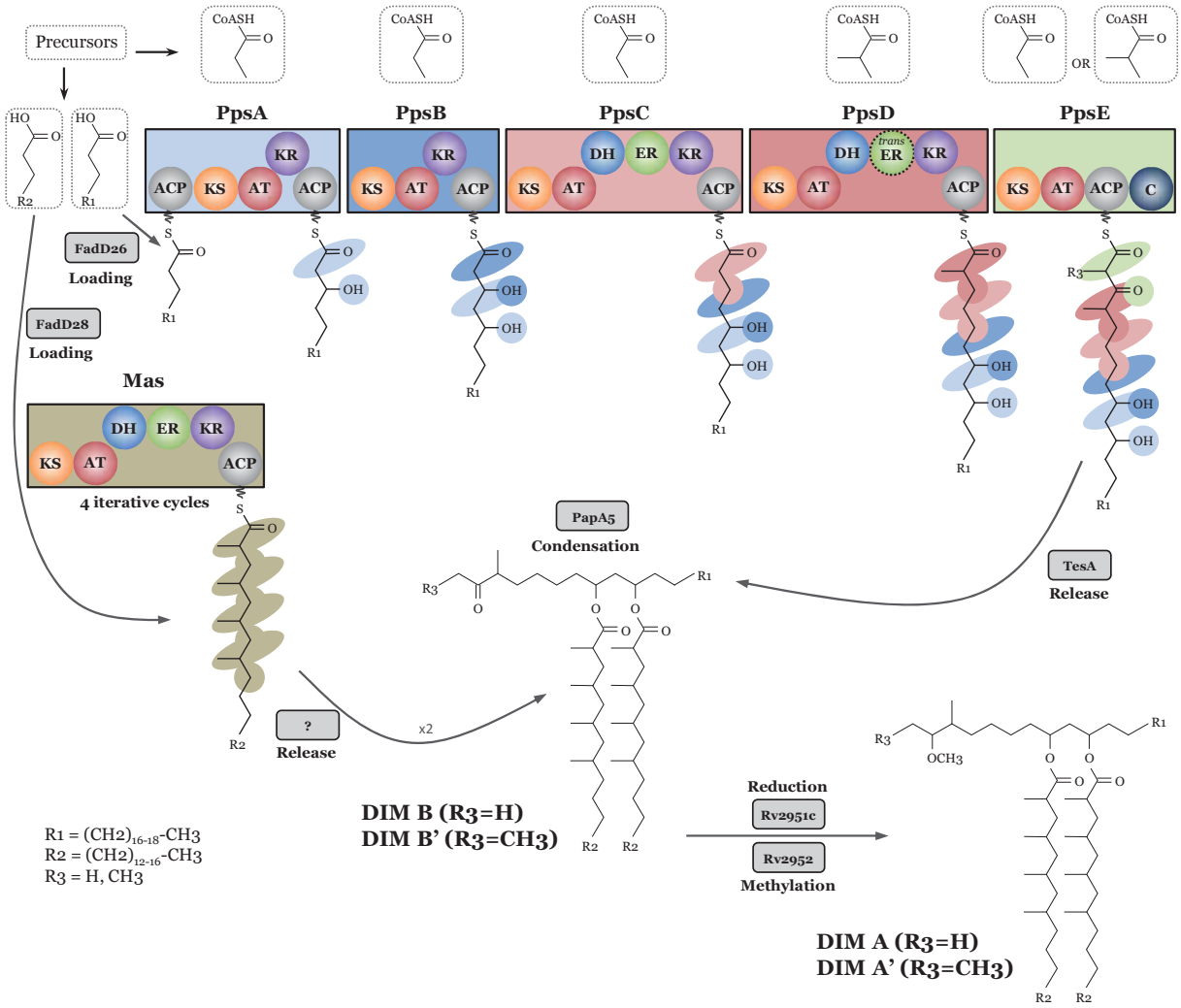


Figure 22. Phthiocerol dimycocerosate biosynthesis.

Condensation of two mycocerosic acids and a phtiocerol result in DIM B. Further reduction and methylation steps of the phtiocerol ketone yield DIM A (or DIM A'), which is the major compound in *Mtb*. PKSs and their domains are colored in accordance with Figures 5 and 6. Precursors are embedded in dashed rectangles. In the case of PpsD there is no ER domain, instead, the *trans*-acting enoylreductase Rv2953 catalyzes the reduction of the carbon-carbon double bond.

mycocerosic acids chain length, (2) presence of a methoxyl (DIM A, major form) or a hydroxyl group instead of the ketone on C3 (DIM B, minor form), (3) presence of an additional methyl moiety on C1 (DIM A' or DIM B'). PGL additionally include a *p*-hydroxyphenylalkanoic acid linked to sugars. In *Mtb*, the major form is known as "PGL-tb" and the sugar part corresponds to a 2,3,4-tri-*O*-methyl L-furanosyl α(1→3) L-rhamnopyranose α(1→3) 2-*O*-methyl L-rhamnopyranose.

Enzymes responsible for the production of DIM and PGL are coded in a 70kb cluster referred to as the "DIM-PGL" cluster. These include six PKSs for DIM biosynthesis (**Figure 22**) and an additional one for PGL biosynthesis. Other enzymes are responsible for precursor recruitment and transfer, polyketide condensation and modification, and transport to the external part of the mycomembrane (not shown in Figure 22). PpsA, PpsB, PpsC, PpsD and PpsE are type-I mono-modular PKSs (**Figure 22**, compared to DEBS multi-modular PKS in **Figure 5**), which catalyze the formation of the phthiocerol moiety (or phenol-phtiocerol moiety for PGL biosynthesis). Mas is a type-I iterative PKS which catalyze the formation of mycocerosic acids [167]. Pks15/1 is a type-I iterative PKS implied in PGL biosynthesis that it responsible for the elongation of *p*-hydroxybenzoic acid (formed by the Rv2949c pyruvate lyase [168]) to *p*-hydroxyphenylalkanoic acid with malonyl-CoA as elongation unit [169]. A 7bp deletion in the *pks15/1* gene is responsible for the defect in PGL for several Euro-American laboratory strains (H37Rv, Mt103, CDC1551, Erdman) [169].

Phtiocerol biosynthesis starts with the transfer of a C20-C22 fatty acid precursor (or *p*-hydroxyphenylalkanoic acid in the case of PGL biosynthesis) on the ACP domain of PpsA by the FadD26 FAAL [170] (**Figure 22**). PpsA catalyzes the condensation of the precursor with a malonyl-CoA elongation unit and reduce the β-ketone with the help of its KR domain [167]. The growing polyketide is then transferred to the ACP domain of PpsB to undergo the same treatment. Next, PpsC and PpsD elongate it using a malonyl- and methylmalonyl-CoA, respectively. They also catalyze the complete reduction of the previously incorporated β-ketones using their full set of β-carbon processing domains (KR, DH, and ER). Finally PpsE catalyzes the condensation with either a malonyl- or a methylmalonyl-CoA without further modification, to form the phthiocerol chain.

In parallel, mycocerosic acid biosynthesis starts with the transfer of a C16-C20 by FadD28 FAAL [170] onto the ACP domain of Mas, which then iteratively elongates and processes the polyketide, with methylmalonyl-CoA as elongation unit [167]. Resulting mycocerosic acids are then transferred onto the (phenol-)phtiocerol moiety by the PapA5 acyltransferase [171]. This enzyme has been shown to be essential to DIM biosynthesis and since it interacts with PpsE, it has been suggested that it would be responsible for the DIM release from the C-terminus condensase domain (C) of PpsE [172]. Finally, the condensation product (DIM B) can be reduced on the (phenol-)phtiocerol C3 by Rv2951c [173], and then methylated by Rv2952 [174] to yield DIM A (or PGL-tb precursor). In the case of PGL, additional glycosyltransferases Rv2962, Rv2958c, Rv2957 and the methyltransferase Rv2959c are required to add and methylate sugar moieties, respectively, [174] to yield PGL-tb.

(3) TRANSPORT

DIM and PGL biosyntheses require cofactors such as NADPH and ATP which are only present in the cytosol, thus suggesting that it occurs in the cytoplasm. DIM transfer in the outer part of the mycomembrane is poorly understood, nonetheless, several enzymes have been shown to participate to DIM transport. For example, it has been demonstrated that Mmpl7 is essential to the presence of DIM in the *Mtb* envelope [156], [158]. In addition, two-hybrid screening and GST pull-down experiments have shown that Mmpl7 interact with PpsE. Mmpl7 being a resistance nodulation cell division (RND) permease transporter, authors have proposed a model in which the whole biosynthetic machinery would be recruited by Mmpl7 in close proximity to the cytoplasmic membrane to facilitate transfer [175]. DrrC, a subunit of the ABC (ATP Binding Cassette) transporter also involving DrrA and DrrB, has also been shown to be implicated in DIM transport, since an *Mtb* strain mutated on *drrC* produces DIM, but these are not present in the envelope [158]. Finally, the lipoprotein LppX would also be involved in DIM transport. Indeed, inactivation of *lppX* in *Mtb,* induces an accumulation of DIM in both cell wall and plasma membrane, but not in the mycomembrane [176].

All the studies revealing a potential role of transporters in DIM transfer have been realized on strains that do not produce PGL. However, structural similarities between these lipids suggest that they are transported by the same proteins. This hypothesis has gained further credibility with the demonstration of the essentiality of *mmpl7* to the presence of PGL-tb in the envelope of the PGL-producing HN878 strain [163].

f) Activation of the Mycobacterial Polyketide Synthases

In *Mtb*, two putative PPTases have been identified: Rv2553c (AcpS) which would be a 130 residue-long group I PPTase, and Rv2794c (PptT) which would be a 227 residue-long group II PPTase involved in mycobactin biosynthesis [126], [153]. In a study using the fast-growing *Mycobacterium smegmatis*, both AcpS and PptT have been shown to be essential for the growth of the bacterium [103]

AcpS is responsible for the activation of ACPs of both FAS systems (FAS-I ACP domain [104] and FAS-II AcpM [177], [178]). Its solution structure has been determined [179] and has confirmed the trimeric arrangement with an α/β fold of the protein, a typical feature of group I PPTases. The ACP binding site is a flexible loop that would allow the enzyme to adapt to both the FAS-I ACP domain and AcpM.

PptT is assumed to be responsible for the activation of all the PKSs and NRPSs in *Mtb* [103], [153]. Indeed, *in vivo* activation of PpsA, PpsB, PpsC, PpsD, Mas and Pks13 [103], and *in vitro* activation of MbtB and MbtE [153] have been demonstrated. PptT is therefore required for the biosynthesis of sulfolipids, DAT & PAT, mycobactins, phtiocerol dimycocerosates, and mycolic acids [103]. Using an *Mtb* conditional expression mutant of PptT, our collaborators in the Guilhot's group at the Institut de Pharmacologie et de Biologie Structurale in Toulouse, have recently established that PptT is required for mycobacterial survival in various environments, including those encountered in macrophages, and during the various phases of infection in an animal model. More specifically, they have demonstrated that PptT is

required for the replication of *Mtb* both *in vitro* and *in vivo* during the acute and chronic phases of infection in immunocompetent mice [104]. The impaired survival of *Mtb* in mice during the chronic phase was probably not solely due to disrupted mycolic acids biosynthesis, since in chronic tuberculosis, mycobacteria are thought to have low level of replication and therefore mycolic acid synthesis is probably low. As antibiotics affecting the synthesis of mycolic acids have only poor bactericidal activity during the chronic phase of infection in mice, the requirement of PptT for *Mtb* replication may be the consequence of additive effects. Indeed, PptT is required for the expression of mycobactins, DIM, PGL, SL, DAT/PAT which are all involved in *Mtb* virulence. Moreover, DIM also contribute to the permeability barrier role of the mycobacterial cell envelope [158], and their absence increase *Mtb* sensitivity to the bactericidal activities of phagocytes [159].

II. RESULTS

A. PROJECT I: STUDY OF A TYPE-I MONO-MODULAR POLYKETIDE SYNTHASE

1. PROJECT

a) Significance

Polyketides have been shown to be of great value in numerous pharmaceutical applications (see **section I.A.1**), and interest is rising on the design and production of new polyketide compounds. Enzymes responsible for their biosynthesis, the polyketide synthases, are complex molecular machineries that consist in a set of catalytic domains designed to work in synergy in order to produce a large panel of possible compounds. New approaches implying the modification of PKSs to synthesize new polyketides require a precise understanding of these complex machineries to gain efficiency. In an attempt to better characterize polyketide biosynthesis and address some outstanding issues (described in section **I.B.1.b**) about the structural features that determine the function of PKSs, we have decided to undertake a functional and structural study on PpsC, a mono-modular type-I PKS.

Tuberculosis is the second greatest killer disease worldwide, after AIDS, due to a single infectious agent [WHO, 2011]. The low BCG vaccine efficiency, in addition to the emergence of multi-drug resistant strains contaminating nearly half a million people per year underlies the failure of the actual therapies. Now that the

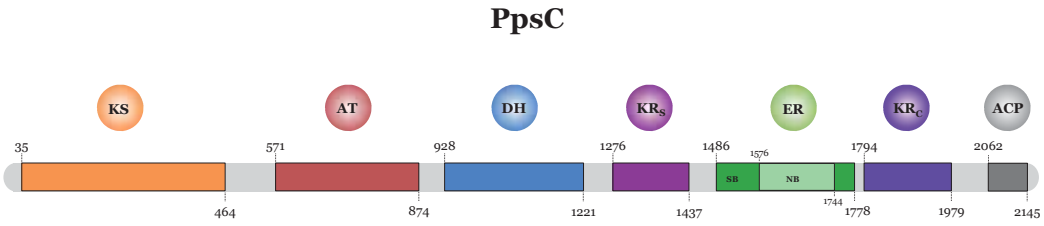


Figure 23. PpsC domains boundaries.

Bioinformatic boundaries of the PpsC domains. KS: ketosynthase ; AT: acyltransferase ; DH: dehydratase ; ER: enoylreductase ; KR_S: ketoreductase structural sub-domain ; KR_C: ketoreductase catalytic sub-domain ; ACP: acyl carrier protein.

determinants of *Mtb* pathogenicity begin to be precisely defined, a novel approach consisting in the rational design of *Mtb* drugs has arisen. In this context, several studies have shown that some polyketide-derived lipids greatly contribute to *Mtb* virulence. More specifically, DIM have been shown to be required for *Mtb* to possess its fully virulent phenotype and therefore their biosynthesis pathway represents a promising target candidate for drug design. Such approach being greatly facilitated by the knowledge of the tridimensional structure of the protein target, we have been interested in determining the high-resolution structure of a PKS essential for DIM biosynthesis.

The aim of this project is to structurally and functionally characterize the polyketide synthase PpsC from *Mtb*. Among the six PKSs that participate in DIM biosynthesis, PpsC has been chosen for three main reasons. First, it is a modular PKS, the most promising class of PKSs for the production of new drug-active polyketides. Second, between all PpsA-E PKSs, PpsC is the only one to possess a full set of β-carbon processing domains (**Figure 23**) and thus represents a model for the study of this cluster of enzymes. We were also interested in deciphering the overall architecture of a PKS in order to understand how the domains assemble together and compare it to the structurally- and functionally-related mFAS. Finally, at the beginning of this project, PpsC, along with PpsB and Mas, were the only enzymes that had been demonstrated as strictly essential to DIM biosynthesis [180], [181], an indispensable character for rational drug design approaches.

It is noteworthy to mention that our group is also interested in the characterization of the PpsA and Mas enzymes. The functional and structural study of these enzymes, in addition to PpsC, is part of a global project (XPKS-MYCO) involving several groups in our institute (IPBS, Toulouse, France), whose focus is the characterization of PKSs essential for the virulence of *Mtb*

b) Strategy

Due to their length and flexibility, PKSs are thought to be very difficult to crystallize in their entire form. Moreover, no structure of a full-length PKS has ever been reported. Structural information on PKSs has then arisen from protein fragments representing one or rarely two domains of a given enzyme. Taking this into account, we have decided to study individual fragments representing one or more domains of PpsC.

PpsC is a 231 kDa mono-modular PKS that possess six catalytic domains: KS, AT, DH, ER, KR and ACP, separated by linkers (**Figure 23**). Our strategy first relied on identifying soluble fragments representing one or more domains of the enzyme using the recently developed *domain trapping* strategy. These fragments were then subjected to a specific protocol developed to produce and purify them in large amounts. Pure and stable fragments were then selected for crystallization trials and further characterized using enzymatic tests and biophysical methods.

2. APPLICATION OF THE DOMAIN TRAPPING METHOD

In order to structurally and functionally characterize PpsC, our first goal has been to identify soluble fragments, representing each catalytic domain. The most commonly used approach to this problem is limited proteolysis. However, this method has several drawbacks including the low diversity in fragment space and possible issues concerning the solubility of the fragments. Not to mention the many other limitations inherent to this methodology, achieving it for every six domains that compose PpsC would have been too much time-consuming and labor-intensive.

Instead, we have chosen to use the *domain trapping* strategy, recently developed by J.D. Pédelacq in collaboration with researchers from the Los Alamos National Laboratory. This methodology relies on two independent technologies which aim at identifying suitable protein fragments for downstream structural and functional characterization. Briefly, randomly-cut gene fragments of the gene coding the protein of interest are cloned and expressed in two separate vectors for *in-vivo* selection of (1) in-frame fragments using an insertion dihydrofolate reductase (DHFR) vector and (2) soluble fragments using the previously developed split-GFP technology [182], [183]. Finally, an inverse polymerase chain reaction (iPCR) step is used to create domain-focused sub-libraries of soluble fragments.

This methodology has first been used on the p85α subunit of the phosphoinositide-3-kinase to demonstrate its efficiency. It has then been successfully applied to the polyketide synthase PpsC, allowing us to obtain libraries of *E. coli*-expressed soluble fragments centered on each catalytic domain of the enzyme. These fragment libraries represented the starting point of my work on this project. Some of the results I have obtained for PpsC, *i.e.* the crystallization of the AT, DH, and ER domains, have been included in the following publication and will also be discussed in the next chapter:

J.-D. Pedelacq, H. B. Nguyen, S. Cabantous, B. L. Mark, P. Listwan, C. Bell, N. Friedland, M. Lockard, A. Faille, L. Mourey, T. C. Terwilliger, and G. S. Waldo, "Experimental mapping of soluble protein domains using a hierarchical approach.", *Nucleic acids research*, vol. 39, no. 18, p. e125, Oct. 2011.

Nucleic Acids Research, 2011, 1–11
doi:10.1093/nar/gkr548

Experimental mapping of soluble protein domains using a hierarchical approach

Jean-Denis Pedelacq[1,2,*], Hau B. Nguyen[3], Stephanie Cabantous[4,5,6], Brian L. Mark[7], Pawel Listwan[3], Carolyn Bell[3], Natasha Friedland[3], Meghan Lockard[3], Alexandre Faille[1,2], Lionel Mourey[1,2], Thomas C. Terwilliger[3] and Geoffrey S. Waldo[3,*]

[1]CNRS; IPBS (Institut de Pharmacologie et de Biologie Structurale), 205 route de Narbonne, F-31077 Toulouse, France, [2]Université de Toulouse; UPS, IPBS, F-31077 Toulouse, France; [3]Bioscience Division, MS-M888, Los Alamos National Laboratory, Bikini Atoll Rd, SM30, Los Alamos, NM 87545, [4]INSERM UMR1037-Cancer Research Center of Toulouse, [5]Université de Toulouse, [6]Institut Claudius Régaud, 31052 Toulouse Cedex and [7]Department of Microbiology, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

Received April 12, 2011; Revised May 26, 2011; Accepted June 15, 2011

ABSTRACT

Exploring the function and 3D space of large multidomain protein targets often requires sophisticated experimentation to obtain the targets in a form suitable for structure determination. Screening methods capable of selecting well-expressed, soluble fragments from DNA libraries exist, but require the use of automation to maximize chances of picking a few good candidates. Here, we describe the use of an insertion dihydrofolate reductase (DHFR) vector to select in-frame fragments and a split-GFP assay technology to filter-out constructs that express insoluble protein fragments. With the incorporation of an IPCR step to create high density, focused sublibraries of fragments, this cost-effective method can be performed manually with no *a priori* knowledge of domain boundaries while permitting single amino acid resolution boundary mapping. We used it on the well-characterized p85α subunit of the phosphoinositide-3-kinase to demonstrate the robustness and efficiency of our methodology. We then successfully tested it onto the polyketide synthase PpsC from *Mycobacterium tuberculosis*, a potential drug target involved in the biosynthesis of complex lipids in the cell envelope. X-ray quality crystals from the acyl-transferase (AT), dehydratase (DH) and enoyl-reductase (ER) domains have been obtained.

INTRODUCTION

Over the past 10 years, the *Mycobacterium tuberculosis* Structural Genomics Consortium (http://www.doe-mbi.ucla.edu/TB/), a large-scale center funded by the National Institutes of General Medical Sciences (NIGMS), has cloned more than 1400 protein targets for cell-based production in *Escherichia coli*. Only half of the proteins expressed have been produced in a soluble form. Other structural genomics initiatives also confirmed this step to be one major bottleneck in structural biology. Screening approaches for improved folding and stability of protein targets have brought new insights into solving structures of single-domain proteins (1,2). However, structure determination of multidomain proteins has been found to be more challenging due to their larger size and increased instability. In addition, domain boundaries are not always straightforward to predict (3). In this respect, high-throughput approaches of generating libraries of truncated DNA fragments (4) combined with a colony filtration immunoblot using antibody detection of tagged constructs (5,6), the detection of a fluorescent fused GFP phenotype (7,8) or a fused C-terminal biotin acceptor peptide (9) have proven to be successful in identifying constructs potentially amenable to functional and structural characterization. Unfortunately, all these approaches lack a filtering strategy to effectively eliminate DNA fragments, which do not encode authentic protein domains. Instead, fully automated strategies have been implemented to effectively screen the thousands of clones and pick a few good candidates (10).

*To whom correspondence should be addressed. Tel: +33 5 61 17 54 11; Fax: +33 5 61 17 59 94; Email: Jean-Denis.Pedelacq@ipbs.fr
Correspondence may also be addressed to Geoffrey S. Waldo. Tel: +505 667 8161; Fax: +505 665 3024; Email: waldo@lanl.gov

Eliminating frame-shifted fragments from DNA libraries has become a key step to addressing the construction of expression plasmid libraries that produce protein domains in-frame with the target gene. Existing systems involve expressing fragments as N-terminal fusions to murine dihydrofolate reductase (mDHFR) (11), kanamycin (12) or β-lactamase (13). However these technologies yield false positives originating from translation initiation at internal ribosome binding sites (IRBS). A number of bipartite selection systems have also been developed in an attempt to overcome these limitations. In these systems, the DNA sequence of interest is inserted between the two halves of the reporter, which are both required to give an observable phenotype (14–18).

In this article, we describe a novel approach that uses a two-body *E. coli* dihydrofolate reductase (DHFR) (19) scaffold for selecting in-frame DNA sequences from a random library of a fragmented gene, combined with the split-GFP technology (20) to identify soluble candidates. We used the regulatory subunit p85α of the class I_A phosphoinositide 3-kinase (PI3K) as a benchmark to validate our method. We then tested it onto the polyketide synthase PpsC from *M. tuberculosis*. This 230 kDa mega-synthase plays a key role in the virulence of this microbial pathogen through the synthesis of phtiocerol dimycocerosates, a family of lipids located in the cell envelope. With the incorporation of an inverse polymerase chain reaction (PCR) step to increase population density in fragments within domains, X-ray quality crystals from the acyl-transferase (AT), dehydratase (DH) and enoyl-reductase (ER) domains have been obtained.

MATERIALS AND METHODS

Gene cloning and fragmentation

The *p85α* and *Ppsc* genes from *M. tuberculosis* were cloned into the NdeI/BamHI and NdeI/SpeI sites of a pET26b plasmid (Novagen, Madison, WI, USA), respectively, and PCR amplified using gene-specific primers (Supplementary Data 2). DNA fragmentation conditions of the *p85α* gene were optimized using small aliquots of concentrated PCR products incubated with a serial 2-fold dilutions of a DNase I stock solution at 1 U/µl (Invitrogen, Carlsbad, CA, USA). Best condition corresponded to a 24-fold dilution from a 1 µl stock solution of DNase I with 20 µl of 10 mM Tris–Hcl pH = 7.4 and 3 µl of 10 mg/ml Bovine Serine Albumine (BSA). Cleaned PCR product of 50 µl was mixed with 6 µl of 0.5 M Tris–Hcl pH = 7.4 and 1 µl of 100 mM $CoCl_2$. Both solutions were pre-incubated in a PCR block at 15°C for 5 min before mixing. Two libraries were generated by adding 6 µl of the 24-fold DNase I solution to the PCR mixture: a 250–400 bp DNA library (small size) and a 400–750 bp DNA library (large size) with incubation times of 5 and 2 min, respectively. Digestion reactions were stopped by adding 650 µl of PB buffer before cleaning through a Qiaquick PCR purification column (Qiagen Inc. USA, Valencia, CA, USA). In the case of *Ppsc*, small size (400–850 kb) and large size (850–1650 kb) DNA libraries were obtained from 160 µl of cleaned PCR product using a HydroShear device from Genomics Solutions (Ann Arbor, MI, USA) applying 25 cycles at speed codes 8 and 13, respectively. Extremities of the fragments were polished using 3′-5′-exonuclease activity of Vent polymerase (New England Biolabs, Beverly, MA, USA) at 72°C for 20 min. Double-stranded DNAs were resolved by agarose gel electrophoresis and visualized by ethidium bromide staining. A slab of gel containing DNA fragments with desired size was then excised and recovered with a QIAquick gel extraction kit (Qiagen Inc. USA, Valencia, CA, USA). DNA fragments designed for the screening of BCR domain constructs were amplified using gene-specific primers (Supplementary Data 2) and ligated into the NdeI/BamHI of pTET ColE1 GFP 11 vector.

Construction of insertion DHFR library

Blunt fragments were ligated in a StuI-digested insertion DHFR (iDHFR) pET vector (Supplementary Data 1) for 12 h at 16°C, and ligated plasmids were transformed into electro-competent *E. coli* DH10B cells (Invitrogen, Carlsbad, CA, USA) to increase efficiency. Starting with 5 × 106 clones/library, transformed cells were plated onto Luria-Bertani (LB) agar plates containing 35 µg/ml kanamycin, allowing the *E. coli* cells lawn to grow overnight at 37°C. Overnight colonies from lawns of 3 × 105 clones, estimated by dilution plates, were washed off and used for plasmid preparation prior to transformation into chemically competent *E. coli* BL21 (DE3) Tuner cells. Following overnight growth at 32°C on LB/kanamycin medium, cells were diluted in LB containing 20% glycerol to $OD_{600\,nm} = 1.0$ for −80°C freezer stocks. Forty microliter of the 1.0 OD freezer stock were used to seed a 3 ml LB/kanamycin culture. Cells were propagated until $OD_{600\,nm} = 0.5$ was reached, and then induced with 20 µM IPTG for an additional 2–3 h. Cells were diluted in 1 ml LB to $OD_{600\,nm} = 2.0$, yielding 10 × 108 cells/ml, and plated on LB medium containing 6 µg/ml trimethoprim (TMP) and 20 µM IPTG. To compare the colony-forming unit (CFU), numbers in the presence or absence of TMP, cells were further diluted to 1/16000 and plated out on two LB/agar plates containing 20 µM IPTG in the presence or absence of TMP. All plates were incubated overnight at 32°C.

Inverse PCRs

Recovered iDHFR libraries were diluted for plasmid preparation. NdeI/BamHI and NdeI/SpeI restrictions sites were used to release fragments from *p85α* and *Ppsc*, respectively. Gel extracted and cleaned inserts were ligated into their corresponding digested pTET ColE1 GFP 11 vector. Inverse PCRs were performed following the protocol by Hoskins and colleagues (21). For each *p85α* and *Ppsc* targeted domain, phosphorylated forward and reverse primers were designed (Supplementary Data 2). Briefly, 100 µl inverse PCRs (IPCRs) were conducted with Phusion DNA polymerase (Finnzymes) according to the manufacturer's instructions. Following a self-ligation with T4 DNA ligase in a 100 µl volume at 16°C overnight and a digestion with DpnI enzyme at 37°C for 2 h 30 min, ligated pTET ColE1 GFP 11

vectors containing targeted inserts were transformed into chemically competent BL21 (DE3) pET GFP 1–10 cells.

Solubility screens using the split-GFP assay

In vivo solubility screenings were performed as previously described (22). Briefly, cells were grown to saturation in LB containing 35 μg/ml kanamycin and 75 μg/ml spectinomycin, and diluted in 20% glycerol to $OD_{600\,nm}$ = 1.0 for −80°C freezer stocks. Frozen cells were thawed at 0°C, 400-fold diluted (twice) in LB and plated onto a nitrocellulose membrane with selective LB-agar containing the same antibiotics (approximately 3000 colonies). After overnight growth at 32°C, the membrane was transferred onto a pre-warmed plate containing 250 ng/ml AnTet for 2 h, and rested back onto its original LB-Kan-Spec plate for 1 h. Following induction with 1 mM IPTG at 37°C for 1 h, the induced colonies were illuminated using an Illumatool Lighting System (LightTools Research), equipped with a 488 nm excitation filter. An ensemble of 96 clones with decreasing levels of *in vivo* fluorescence intensities were picked for each library. Columns 1–3 of the tissue culture plate correspond to only bright clones, columns 4–9 present a range in fluorescence intensity levels from medium bright to faint and columns 10–12 only correspond to very faint clones. As a control, a total of 96 clones were picked randomly by hand, transferred to 96-well plates and grown before sequencing. All the clones were used as starter cultures on 96-well tissue culture plates for *in vitro* complementation split-GFP assays using our in-house automated, high-throughput, liquid-handling platform (23).

Identification of fragments boundaries

Individually picked clones were grown overnight at 30°C in a 96-well tissue culture plate containing 7.5% glycerol in LB-Kan-Spec medium. Plasmid amplification at the Los Alamos genome sequencing facility using rolling-circle amplification in the presence of a forward primer specific of *tet*-promoter and a reverse primer specific of GFP 11 (Supplementary Data 2) yield high-quality sequence. DNA sequences were analyzed using BioEdit® software. Sequence alignments were performed by aligning individual fragments onto the full-length parent gene to determine the exact boundaries from the forward (start of the fragment) and reverse sequence (end of the fragment). Based on the *in vitro* solubility assays, fragments were color-coded black, light green and bright green, where the black side of the spectrum identifies the least soluble protein fragments and the bright green side corresponds to the top 25% of the most soluble ones. Fraction of color-coded black and light green fragments varies from 25% and 50% for a full-length gene mapping to 37.5% in the case of IPCRs.

Small scale expression and solubility tests

An ensemble of 10 fragments spanning the PpsC polypeptide chain were selected and subcloned from the pTET-GFP, 11 plasmid into a N6–HIS or C6–HIS pET vector. The resulting clones were grown at 37°C in 1 ml cultures using 35 μg/ml kananycin. Cells were induced in exponential phase with 1 mM IPTG for 3 h. Cell culture pellets of 1 ml of each fragment were separately resuspended in 40 μl 150 mM NaCl, 100 mM Tris–HCl pH = 7.5, 10% (v/v) glycerol (TNG buffer) and sonicated. The lysate was fractionated by centrifugation to yield the soluble and pellet fractions. The pellet fraction was washed twice with 100 μl TNG, centrifuged and resuspended in the same starting volume. Samples corresponding to the soluble (S) and pellet (P) fractions were resolved on a 4–20% gradient Criterion SDS–PAGE gel (Bio-Rad, Hercules, CA, USA). Protein samples were stained using Gel Code Blue stain reagent (Pierce, Rockford, IL) and imaged using a GS-800 Calibrated Densitometer (Biorad, Hercules, CA, USA).

Metal affinity resin purification of selected fragments

Five hundred milliliter cultures of BL21(DE3) cells expressing selected protein fragments were grown to OD (600 nm) ∼0.5–0.7 in LB medium supplemented with 1 mM kanamycin, induced with 0.5 mM IPTG for 5 h at 32°C, pelleted by centrifugation, resuspended in 15 ml 100 mM Tris–Hcl pH = 8.1 containing 150 mM NaCl and sonicated. The soluble extract of 15 ml was mixed with an equal volume of 50% v/v slurry of metal affinity resin beads (Talon resin, Clontech, Palo Alto, CA, USA) in TNG buffer for 10 min and centrifuged briefly. The unbound fraction was removed by pipetting and the beads were washed twice with 10 volumes of TNG loading buffer. After an additional wash with TNG buffer supplemented with 10 mM imidazole, His-tagged proteins were eluted with 250 mM imidazole in TNG buffer. For each purification step, the proteins elution samples were resolved on a 4–20% gradient Criterion SDS–PAGE gel (Bio-Rad, Hercules, CA, USA) and stained using the same procedure described above. Using this procedure, 98% pure AT (∼4 mg), DH (∼20 mg) and ER (∼5 mg) proteins were obtained. The absence of aggregates in the samples and polydispersity levels of <10% were confirmed using a DynaPro™ Dynamic Light Scattering (DLS) Instrument from Wyatt Technology.

^{15}N–Protein labeling and sample preparation for NMR spectroscopy

Protein expression level in minimal media was enhanced by increasing cell density using a 4:1 cell concentrating method (24). For ^{15}N uniform labeling, cells from 21 of LB media were grown at 37°C until an OD_{600} of 0.5–0.7 was reached, then harvested and resuspended in 500 ml of M9 minimal media containing 0.5 g $^{15}NH_4Cl$. Following an additional 30 min of shaking, cells were induced with 0.5 mM IPTG for 7 h at 25°C. Proteins were purified as described above using Talon resin and dialyzed against 50 mM Na phosphate buffer pH = 7, 1 mM DTT, 1 mM EDTA overnight at 4°C to remove imidazole and salt from elution buffer. Protein solution was concentrated using an Amicon Ultra-15 centrifugal filter device (10 kDa cutoff; Millipore). Final NMR samples usually contain 0.5–1 mM protein in 50 mM Na phosphate buffer pH = 7, 1 mM DTT, 1 mM EDTA and 10% D2O. ^{1}H–^{15}N HSQC spectra were recorded at 298 K on

Varian Inova 720 MHz spectrometer using a conventional probe. Complex points of 1024 in the direct dimension (^{1}H) and 256 complex points in the indirect dimension (^{15}N) were collected. All spectra were processed using nmrPipe (25) and analyzed by the Sparky software (Goddard and Kneller, University of California, San Francisco). Chemical shifts were referenced to 4,4-dimethyl-4-silapentane-1-sulfonic acid (26).

RESULTS

Selection of in-frame, well-expressed and soluble fragments

Our strategy has four distinct steps, where selection pressure forces the false positives at any given step to be effectively eliminated and the number of false negatives to be reduced. First, a library of 5×106 clones expressing DNA fragments is created by fragmentation of a PCR amplified gene using DNase I or mechanical shearing (Figure 1). Fragments of the desired size are excised from preparative agarose gel, blunt-ended with 5′-3′-exonuclease, and then cloned between the two halves of bacterial DHFR at a permissive site between amino acids 86 and 87 in the presence of TMP (Supplementary Data 1). We found that 3–6 μg/ml TMP killed DH10B *E. coli* cells, but allowed clones expressing inserts without stop codons to subsist. Under this selective antibiotic pressure, approximately 1 in 18 cells survive, corresponding to fragments translated in the same reading frame as the parent open reading frame (ORF) and in-frame with the reporter destination vector (8). This leads to 3×105 clones recovered at this step (Figure 1), a number to compare with the few hundreds in-frame clones from published methods (5,6,11). Only fully automated platforms capable of picking and assaying tens of thousands of clones in parallel for expression and solubility can compete with our approach. In this respect, Expression of Soluble Proteins by Random Incremental Truncation (ESPRIT) (9) has led to remarkable results on several challenging targets (10,27–30). In the third step, in-frame fragments are subcloned into the split-GFP system. At this point, full-length libraries can be screened for solubility after *in vivo* sequential induction of the GFP 11-tagged protein fragments and the complementary GFP 1–10 detector (22). Clones displaying a wide

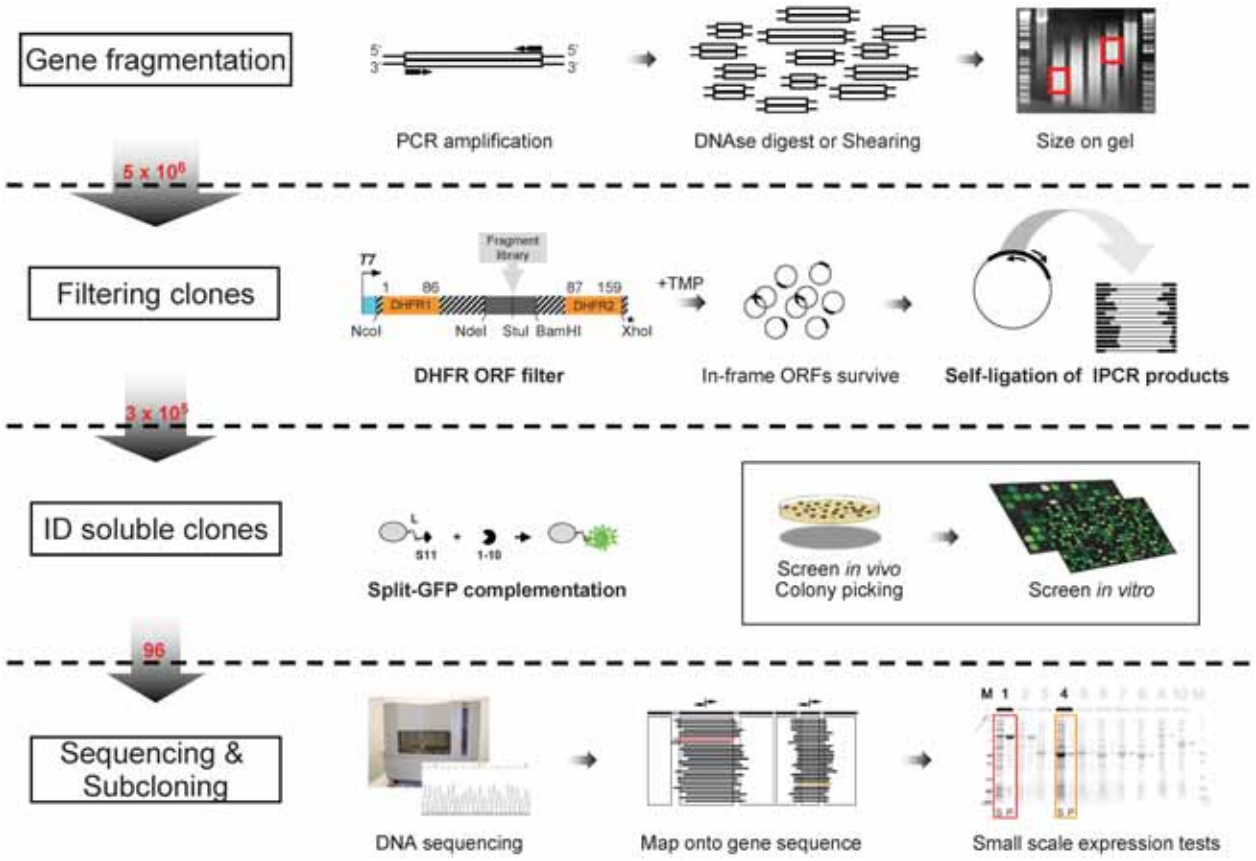


Figure 1. The GFP-enabled domain trapping strategy. The PCR-amplified gene is fragmented by chemical or mechanical means and DNA fragments of desired size are excised from agarose gel. Blunt-end fragments are cloned into the iDHFR ORF filter, where only the in-frame ones permitting the expression of the second half of DHFR will survive. Inserts from the recovered plasmids are cloned into the split-GFP vector and used for IPCR to create high density, focused sublibraries of fragments prior to the split-GFP assay. A range of fluorescent clones are picked and grown in 96-well liquid cultures for *in vitro* quantification of the soluble and insoluble protein fractions. Clones are sequenced and the fragments are aligned onto the full parent gene. Fragments can be directly tested for expression or subcloned without the S11 tag into a pET vector. Numbers to the left indicate the approximate library size at the different steps.

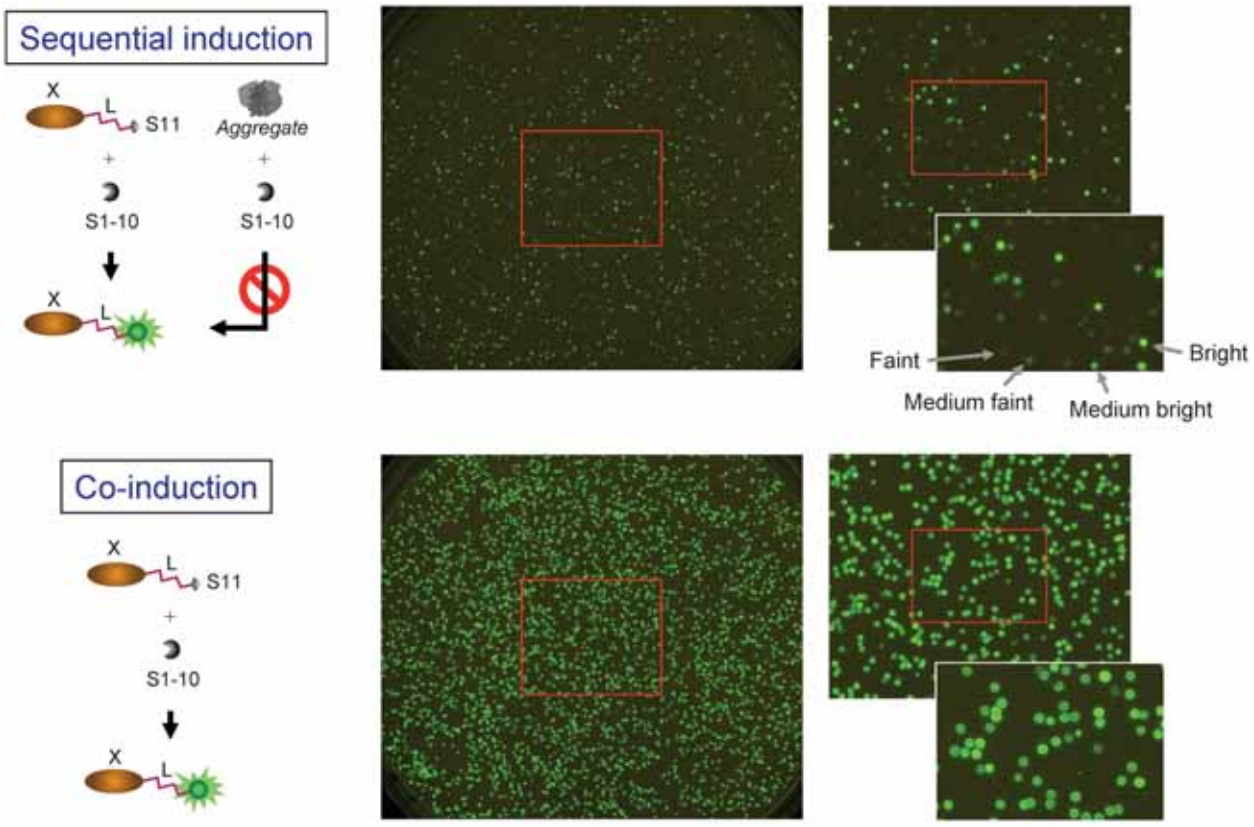


Figure 2. Screening clones using the split-GFP reassembly assay. Images showing cell colony fluorescence from agar plates after sequential induction (solubility reporter) and co-induction (expression reporter) of the GFP 11-tagged protein fragments and its complementary GFP 1–10 detector. Clones displaying a wide range of fluorescence are visible after sequential induction.

range of fluorescence are picked from the agar plates and grown in 96-well liquid culture plates (Figure 2). Within each library, we picked a total of 384 clones in four 96-well plates with fluorescence intensity levels from bright to faint (see Methods in Supplementary Data). As an alternative, IPCR (21) can first be used to create high density, focused sublibraries of fragments prior to the split-GFP assay. In this case, 96 clones were picked per sublibrary of fragments to ensure maximum coverage (Figure 1). For *in vitro* quantification, *E. coli* cells were first induced with anhydrotetracycline (AnTET) to overexpress the GFP 11-tagged proteins. Soluble lysates and insoluble fractions were assayed by adding the GFP 1–10 detector fragment, as previously described (22). In the final step, DNA sequencing was used to determine the boundaries of each fragment by reference to the parent gene. *In silico*, all fragment sequences within a library were aligned onto the parent gene and color-coded by solubility levels. We used the color scheme *black/light green/bright green*, where the black side of the spectrum identifies the least soluble fragments and the bright-green color identifies the most soluble ones. This provides a visual tool to correlate fragment boundaries with solubility levels and makes it easy to identify the most compact and soluble fragments for downstream functional and structural characterization.

p85α as a benchmark for testing our domain trapping strategy

The structural organization of the regulatory subunit p85α of the class I_A PI3K has been very well studied. Except for the two coiled-coil regions CC1 and CC2, the documented structures of the well-folded SH3 (31), BCR (32), N–SH2 (33) and C–SH2 (34) make p85α a good benchmark for demonstrating the feasibility and efficiency of domain trapping strategies.

Large quantities of PCR amplified target DNA (1–2 μg) were produced using either the Platinum® *Taq* DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA, USA) or the Phusion® High-Fidelity DNA polymerase (NEB, Ipswich, MA, USA). Since DNA ligation efficiency varies inversely with the size of the insert, and to avoid biased ligation of small fragments, two individual libraries of fragments were created: from 250 to 500 bp and from 350 to 750 bp. Indeed, the larger size fragment library would let us 'fish' soluble fragments encompassing the larger BCR domain (550 bp), whereas the smaller size fragment library would favor the selection of smaller soluble fragments from the SH3 domain (252 bp) and the two SH2 domains (345 and 324 bp for N–SH2 and C–SH2, respectively). DNase I reaction conditions were optimized to narrow the window of highly concentrated DNA fragments in the desired size range (see Methods in Supplementary Data).

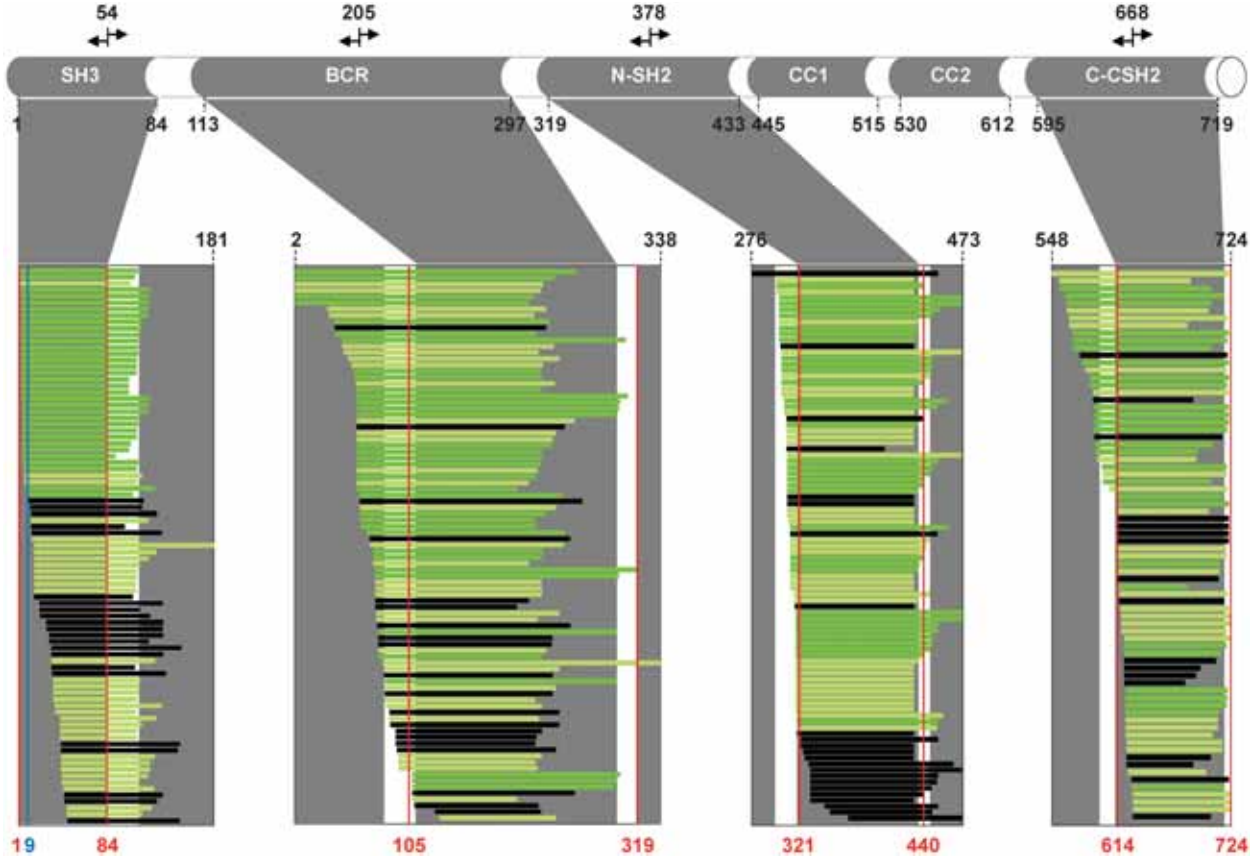


Figure 3. Mapping of the IPCR p85α targeted fragments. IPCRs using oppositely directed primers were used to generate sublibraries of fragments. For each sublibrary, an ensemble of 96 clones with a wide range of *in vivo* fluorescence intensities were picked and grown in 96-well liquid culture plates for *in vitro* split-GFP solubility screen. Only the correctly sequenced in-frame fragments are represented. Based on the *in vitro* solubility assays, fragments were color-coded black, light green and bright green, where the black side of the spectrum identifies the bottom 20% least soluble protein fragments and the bright green side corresponds to the top 20% most soluble ones. IPCRs within p85α were used to generate four large-size sublibraries (350–750 kb) centered onto the SH3, BCR, N–SH2 and C–SH2 domains. Within each library, solubility values from three or more identical fragments were averaged in order to keep the color-coded representation as clear as possible. Boundaries of structure solved domains are indicated in red and the junction at amino acid position 9 is indicated in blue.

The two independent pools of blunted *p85α* fragments were cloned into the iDHFR vector and transformed into *E. coli* DH10B electro-competent cells (Invitrogen, Carlsbad, CA, USA). Transformation of the recovered plasmids into *E. coli* BL21 Tuner ^{TM}(DE3) competent cells (Novagen, San Diego, CA, USA) containing the *lacY* permease mutation allowed the uniform induction of the cells by IPTG during the subsequent selection of in-frame clones. To ensure that the selection process is consistent with the theoretical 1 in 18 clones expected to survive after the ORF-filter step, each library was diluted and plated on medium in the presence and absence of TMP for fast and accurate colony counting.

Initially, pools of 'in-frame' DNA fragments were subcloned into the pTET-GFP 11 solubility vector using NdeI and BamHI sites (Supplementary Data 1), and transformed into BL21 (DE3) cells containing the pET GFP 1–10 plasmid to screen for soluble expression (20). After sequential induction of the GFP 11-tagged fragments, we could take advantage of the split-GFP complementation assay to screen thousands of clones *in vivo* by visual assessment of their intrinsic fluorescence, which is well correlated with the amount of soluble protein expressed (Figure 2). Both fragment libraries displayed substantial phenotypic variability as observed from the wide distribution of fluorescence intensities. Even though sequencing information was missing for some of the 384 manually picked clones (79 for the small size and 16 for the large size library), probably due to cross contamination with nearby faint or black clones, only a limited number of the sequenced fragments were not in the authentic reading frame (9 out of 305 for the small size and 2 out of 368 for the large size library), thus illustrating the efficiency of selection process. Despite the fact that most p85α domains are well represented, with the exception of the BCR domain, we noticed a bias in the distribution of fragments towards the second-half of the gene, from the N–SH2 domain to the C-terminal end of the gene (Supplementary Figure S1). In contrast, the distribution of randomly picked fragments is more homogeneous, as no information concerning fluorescence intensity levels was taken into consideration (data not shown).

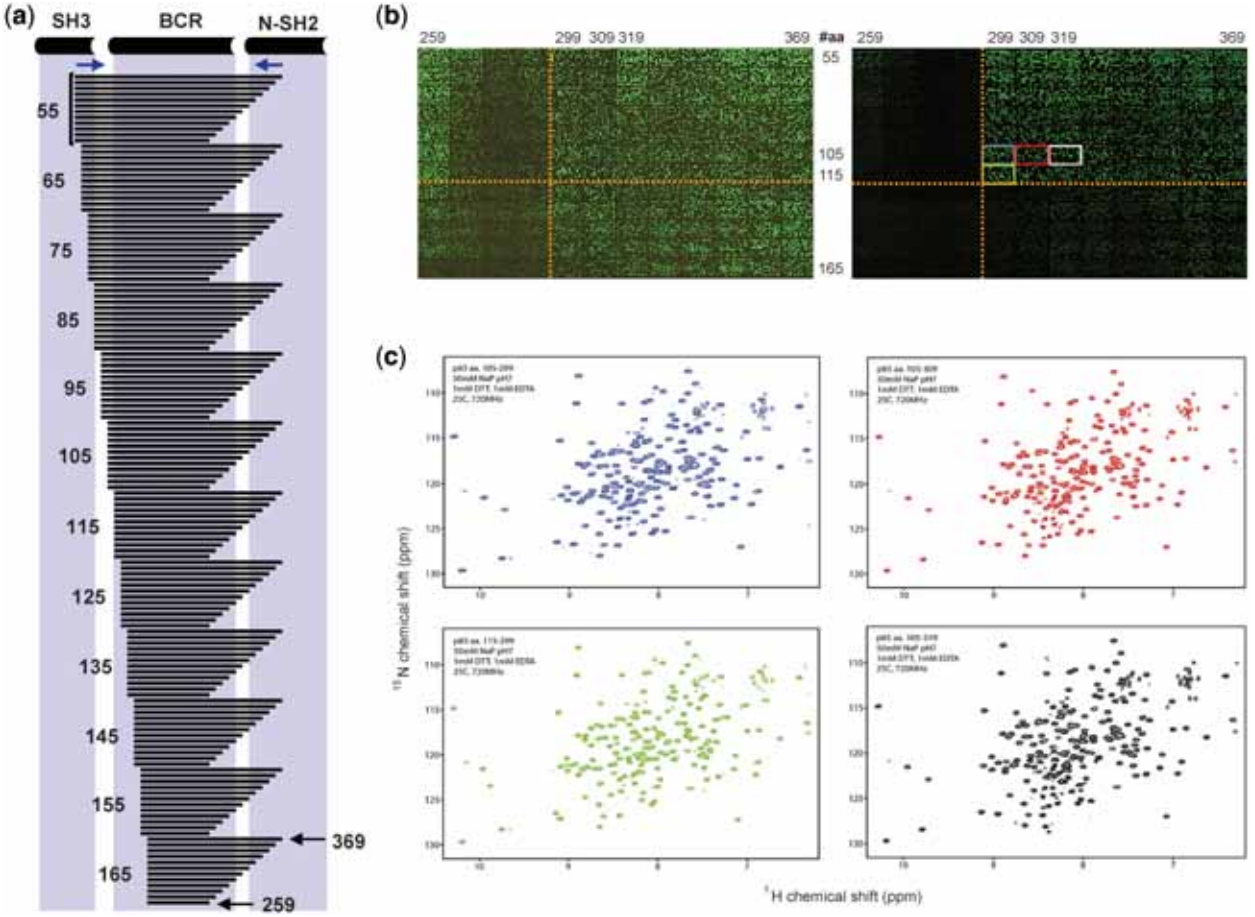


Figure 4. PCR-directed truncations of the p85α BCR domain. (**a**) Schematic representation of all 144 constructs aligned onto the p85α amino-acid sequence. N- and C-terminal positions are indicated. Fragments are organized in groups of 12 with identical N-terminal positions. (**b**) Expression (left) and solubility (right) levels of *E. coli* BL21(DE3) cells expressing the corresponding fragments in fusion with S11 following complementation with GFP 1–10. Orange dashed lines mark boundaries, where dramatic changes in expression and solubility levels were observed. (**c**) In addition to 105–319 for which the X-ray structure is known (PDB code: 1PBW), well-expressed and soluble fragments were selected for downstream NMR studies. HSQC spectra of fragments 105–299 (blue), 105–309 (red), 105–319 (black with corresponding white rectangle in the solubility screen) and 115–299 (green) are represented.

We reasoned that the bias may originate from our multistep selection process that favors the selection of extremely well-behaved fragments to the detriment of less soluble ones. To circumvent this representation artifact and starting with the same original library of in-frame fragments (following iDHFR selection), we used the IPCR (21) technique to selectively enrich for DNA sequences in regions of *p85α*. Phosphorylated primers were designed (Supplementary Data 2) to generate three small-size sublibraries (250–400 bp) centered on the SH3, N–SH2 and C–SH2 domains and one large-size sublibrary (400–750 bp) centered onto the BCR domain (Figure 3). We noticed that cutting into the N-terminal region of the SH3 and N–SH2 domains had a dramatic effect onto the solubility levels of the selected fragments. Also, the incorporation of amino acid residues from the CC2 domain and linker region drastically improved the solubility of the C–SH2 centered fragments, a somewhat surprising and difficult result to predict considering the unstructured nature of these regions.

Fragments can occasionally be used as a starting point to identify more compact and soluble constructs with a clear objective of maximizing the chances of 3D structure (35). As illustrated in Figure 3, IPCR sublibrary centered on the BCR domain contains three soluble fragments (110–294, 110–297 and 110–300) with C-terminal positions slightly shorter than the structurally characterized fragments 105–319 (PDB code: 1PBW). We note that the low solubility of another BCR fragment (117–297) from the full-length *p85α* library could be attributed to its shorter N-terminal end (Supplementary Figure S1b). In an effort to rationalize the effect of N- and C-terminal truncations in the BCR region, we cloned an ensemble of 144 constructs corresponding to a 10 amino acids walk from positions 55 and 369 (Figure 4 and Supplementary Data 2). Protein solubility levels were assessed both *in vivo* and *in vitro* using the split-GFP assay (20). Figure 4 illustrates the expression and solubility levels of all 144 constructs, respectively. As expected, fragments are soluble when the BCR core domain is

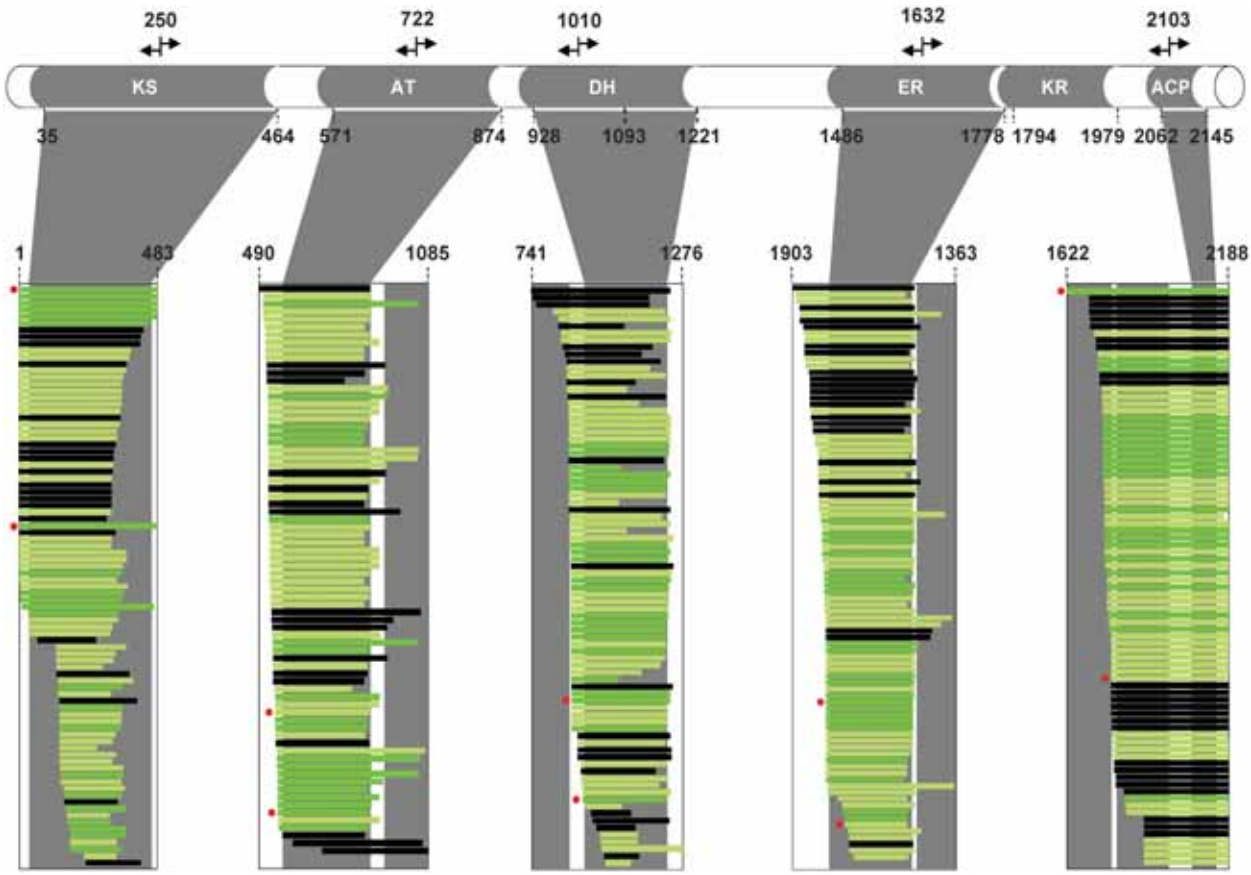


Figure 5. Mapping of the IPCR PpsC targeted fragments. Five large-size sublibraries of fragments (850–1650 kb) centered onto the KS, AT, DH, ER and ACP domains were generated. Picking and color coding of fragments follow the same rules as in Figure 3. Red dots indicate the fragments selected for downstream biochemical and structural characterization.

present. As truncations within the core proceed, protein solubility levels fall. Approximate boundaries of soluble fragments range between positions 115 and 125 in the N-terminal region, and positions 289 and 299 in the C-terminal region. A finer screening using one amino acid incremental truncation from the N-terminal (116–123) and C-terminal (290–297) regions was then performed (Supplementary Figure S2). Based on the fluorescence intensity ratios of solubility over expression, the fragments 116–294 is the shortest most soluble fragment of the BCR domain. Interestingly, cutting into the C-terminal end of helix α10, from positions 297 to 294, did not seem to have an effect on the solubility of the last three fragments within the group. The addition of a three residue linker (Gly–Ser–Asp) to the C-terminus of the protein as a result of molecular cloning site in the pTET-GFP 11 (20) vector may compensate for residues Ser–Thr–Glu of the real protein sequence, thus making truncated variants more stable. A total of six most compact soluble protein fragments, including the structurally characterized fragments 105–319 (PDB code: 1PBW), were selected and purified for 2D NMR experiments. ^{1}H–^{15}N HSQC spectra are well resolved and dispersed, thus indicating that all protein fragments are well folded (Figure 4c and Supplementary Figure S2). These spectra overlay each other, thus suggesting that the extra tails at both N- and C-terminal ends do not affect the domain core structure.

Application to the polyketide synthase PpsC from *M. tuberculosis*

Polyketides comprise natural compounds that are essential for the virulence of major human mycobacterial pathogens, namely *M. tuberculosis* (36,37) and other emerging infectious agents. Polyketide biosynthesis is accomplished by polyketide synthases (PKS), which are giant and multifunctional enzymes. PpsC belongs to the family of type I PKS with six domains present on a single polypeptide chain.

Information on the 3D structure of full-length type I PKS is not available and the current data collected worldwide on fatty acid synthases (FAS) and the 6-deoxyerythronolide B synthase (DEBS) from *Saccharopolyspora erythraea*, the most studied modular PKS, only allow rough modeling of their molecular architectures, thus providing only a low-resolution picture. High-resolution structural information is needed to describe at the atomic level each individual domain in order to understand the full-length PKS catalytic machinery. Dedicated software for the analysis of PKS, MAPSI

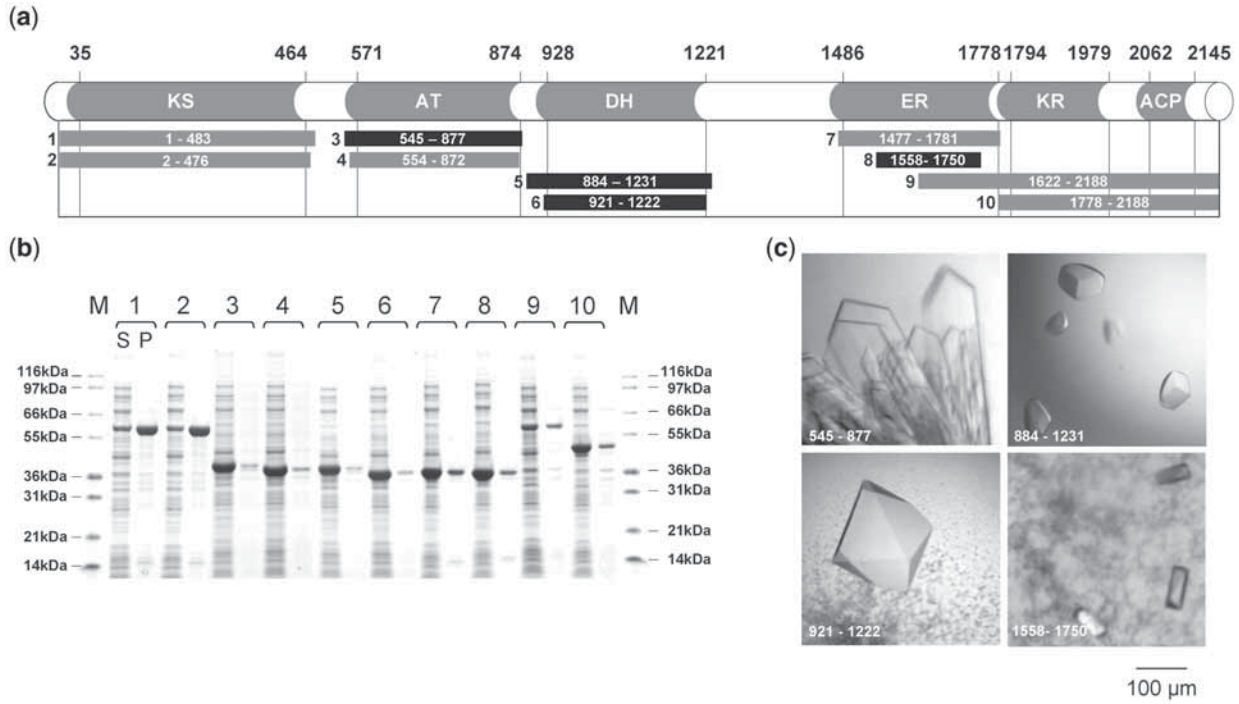


Figure 6. Expression, solubility and crystallization trials of selected PpsC fragments. (**a**) Ten fragments covering the structural domains of PpsC were subcloned from the pTET-GFP 11 plasmid into a N6–HIS pET vector for biochemical and biophysical characterization. (**b**) SDS–PAGE of soluble (S) and pellet (P) fractions of *E. coli* BL21 (DE3) cells expressing the different fragments. (**c**) Pictures of X-ray quality crystals of selected fragments represented as a black rectangle in (a) are shown. The 3D structures of the AT (545–877), DH (921–1222), and truncated ER (1558–1750) domains have been recently determined at 1.8, 2.8, and 2.5 Å resolutions, respectively.

(http://gate.smallsoft.co.kr:8008/pks/mapsitools/index.pl) and SEARCHPKS (http://www.nii.res.in/searchpks.html) were used to identify approximate boundaries for each individual domain. Only in the case of the DH domain, the identification of C-terminal boundaries differs substantially for the methods used in the two approaches (1093 with SEARCHPKS and 1221 with MAPSI).

We used a hydrodynamic point-sink shearing method (38) to create large-size (850–1650 kb) DNA libraries of *Ppsc* fragments. Compared to enzymatic digestion with DNase I, the resulting DNA fragment libraries were tightly distributed in size (data not shown). For each targeted domain, IPCR was used to generate focused sublibraries of fragments, as shown in Figure 5. Initially, phosphorylated forward and reverse IPCR primers were designed to the center of the KS, AT, DH, ER and ACP domains (Supplementary Data 1). Multiple priming sites originating from the high GC content of the *M. tuberculosis* genome lead to PCR amplification failures. To circumvent this problem, 32-mer primers were blasted against the full-length gene to ensure priming at a unique site. An ensemble of 10 fragments covering all six PpsC domains was selected (Figure 6a) and subcloned from a pTET-GFP 11 vector using NdeI and SpeI sites (Supplementary Data 1) into a N6–HIS pET vector. Selected fragments had to satisfy two main criteria: (i) N- and C-terminal boundaries should incorporate the predicted ones and (ii) fragments with different levels of solubility should be considered if possible. SDS–PAGE of the soluble and pellet fractions of *E. coli* BL21 (DE3) cells, expressing the different fragments, were in good agreement with the *in vitro* split-GFP solubility screen (Figure 6b). Crystallization trials have been performed on 6 out of 10 selected constructs centered on the AT, DH and ER domains. Diffraction quality crystals have been obtained for all three domains (Figure 6c). Analysis of the 3D structure of the fragment 921–1222 centered onto the DH domain clearly indicates that the typical double-hotdog fold extends from positions 933 to 1216 (A. Faille *et al.*, manuscript in preparation).

DISCUSSION

Our approach to screening libraries of over a million clones can be of general interest for identifying soluble constructs of 'recalcitrant' proteins, as it does not require the use of an automated robotic platform. It simply relies on well-proven technologies capable of effectively eliminating the unwanted constructs, thus reducing the population size of fragments to be analysed. Although *E. coli* cells expressing self-associated DHFR in the presence of TMP somewhat forces the selection of inserted in-frame fragments, we noticed the presence of short antisense peptides devoid of stop codons and in-frame with both ends of DHFR. Following the *in vivo* split-GFP filtration step, the proportion of false positives drops to <3% of the total number of picked clones. The most plausible explanation is that they

have been eliminated during gel purification prior to subcloning into the GFP S11 vector.

The IPCR step is crucial as it helps remove the bias originating from the overrepresentation of regions of the protein that never get amplified due to the lack of homology to the IPCR primers. It also permits single amino acid resolution boundary mapping as seen from both p85α and PpsC sublibraries of fragments. For example, in the case of the p85α SH3 domain, two distinct populations of fragments at the junction between amino acid positions 6 and 9 were visible (Figure 3). Detailed examination of the solution structure of the SH3 domain (31) revealed that Tyr6 is the last amino acid residue of the N-terminal tail that connects to a four-residues β-strand central to a triple-stranded antiparallel β-sheet. Cutting into this strand would not only have a destabilizing effect on interactions within the β-sheet, but also with a parallel β-sheet of two strands crossing at right angle. In the case of PpsC, single amino acid truncations also have a dramatic impact on the solubility of the expressed fragments, a result which clearly demonstrates the potential of our approach in an attempt to identify potential candidates for downstream functional and structural applications. To date, only the crystal structure of a fragment covering partially the ER domain has been reported (PDB code 1PQW). Thanks to our approach, we successfully crystallized fragments encompassing the AT, DH and ER domains and solved the X-ray structure of the active AT and DH domains (Alexandre Faille *et al*, manuscripts in preparation).

Our domain trapping strategy is also particularly well adapted to situations, where unstructured regions are essentials for the stability of isolated domains. As seen with p85α and PpsC, it offers a rapid and easy way to identify N- and C-terminal boundaries of soluble fragments often difficult to predict solely by theoretical means. In the near future, we anticipate our approach will facilitate decipher structure–function relationships in mechanistically diverse and complex enzymatic machineries, thus opening the way to atomic level description of active sites and domain–domain interactions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the staff of synchrotron beamlines ID14-1, ID14-2, ID23-2 and ID29 at the European Synchrotron Radiation Facility (Grenoble, France); Françoise Viala for her help in preparing the figures. We would like to thank Dr Fengli Zhang at National High Magnetic Field Laboratory (Florida, USA) for his help with NMR spectroscopy.

FUNDING

National Institutes of Health's Protein Structure Initiative (grant number 5U54GM074946-4); the French 'Agence Nationale de la Recherche' (ANR-09-BLAN-0298-01). Funding for open access charge: National Institutes of Health's Protein Structure Initiative (grant number 5U54GM074946-4).

Conflict of interest statement. The split-GFP and related intellectual properties are the subject of domestic and foreign patent applications by Los Alamos National Laboratories on behalf of the Department of Energy and LANS, L.L.C.

REFERENCES

1. Pedelacq,J.D., Piltch,E., Liong,E.C., Berendzen,J., Kim,C.Y., Rho,B.S., Park,M.S., Terwilliger,T.C. and Waldo,G.S. (2002) Engineering soluble proteins for structural genomics. *Nat. Biotechnol.*, **20**, 927–932.
2. Pedelacq,J.D., Cabantous,S., Tran,T., Terwilliger,T.C. and Waldo,G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.*, **24**, 79–88.
3. Sippl,M.J. (2009) Fold space unlimited. *Curr. Opin. Struct. Biol.*, **19**, 312–320.
4. Prodromou,C., Savva,R. and Driscoll,P.C. (2007) DNA fragmentation-based combinatorial approaches to soluble protein expression Part I. Generating DNA fragment libraries. *Drug Discov. Today*, **12**, 931–938.
5. Cornvik,T., Dahlroth,S.L., Magnusdottir,A., Herman,M.D., Knaust,R., Ekberg,M. and Nordlund,P. (2005) Colony filtration blot: a new screening method for soluble protein expression in *Escherichia coli*. *Nat. Methods*, **2**, 507–509.
6. Reich,S., Puckey,L.H., Cheetham,C.L., Harris,R., Ali,A.A., Bhattacharyya,U., Maclagan,K., Powell,K.A., Prodromou,C., Pearl,L.H. *et al.* (2006) Combinatorial domain hunting: an effective approach for the identification of soluble protein domains adaptable to high-throughput applications. *Protein Sci.*, **15**, 2356–2365.
7. Waldo,G.S., Standish,B.M., Berendzen,J. and Terwilliger,T.C. (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.*, **17**, 691–695.
8. Kawasaki,M. and Inagaki,F. (2001) Random PCR-based screening for soluble domains using green fluorescent protein. *Biochem. Biophys. Res. Commun.*, **280**, 842–844.
9. Yumerefendi,H., Tarendeau,F., Mas,P.J. and Hart,D.J. (2010) ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. *J. Struct. Biol.*, **172**, 66–74.
10. Nadal,M., Mas,P.J., Blanco,A.G., Arnan,C., Sola,M., Hart,D.J. and Coll,M. (2010) Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain. *Proc. Natl Acad. Sci. USA*, **107**, 16078–16083.
11. Dyson,M.R., Perera,R.L., Shadbolt,S.P., Biderman,L., Bromek,K., Murzina,N.V. and McCafferty,J. (2008) Identification of soluble protein fragments by gene fragmentation and genetic selection. *Nucleic Acids Res.*, **36**, e51.
12. Nakayama,M. and Ohara,O. (2003) A system using convertible vectors for screening soluble recombinant proteins produced in *Escherichia coli* from randomly fragmented cDNAs. *Biochem. Biophys. Res. Commun.*, **312**, 825–830.
13. Zacchi,P., Sblattero,D., Florian,F., Marzari,R. and Bradbury,A.R. (2003) Selecting open reading frames from DNA. *Genome Res.*, **13**, 980–990.
14. Cho,G., Keefe,A.D., Liu,R., Wilson,D.S. and Szostak,J.W. (2000) Constructing high complexity synthetic libraries of long ORFs using *in vitro* selection. *J. Mol. Biol.*, **297**, 309–319.
15. Seehaus,T., Breitling,F., Dubel,S., Klewinghaus,I. and Little,M. (1992) A vector for the removal of deletion mutants from antibody libraries. *Gene*, **114**, 235–237.
16. Daugelat,S. and Jacobs,W.R. Jr (1999) The Mycobacterium tuberculosis recA intein can be used in an ORFTRAP to select for open reading frames. *Protein Sci.*, **8**, 644–653.

17. Lutz,S., Fast,W. and Benkovic,S.J. (2002) A universal, vector-based system for nucleic acid reading-frame selection. *Protein Eng.*, **15**, 1025–1030.
18. Gerth,M.L., Patrick,W.M. and Lutz,S. (2004) A second-generation system for unbiased reading frame selection. *Protein Eng. Des. Sel.*, **17**, 595–602.
19. Smith,V.F. and Matthews,C.R. (2001) Testing the role of chain connectivity on the stability and structure of dihydrofolate reductase from *E. coli*: fragment complementation and circular permutation reveal stable, alternatively folded forms. *Protein Sci.*, **10**, 116–128.
20. Cabantous,S., Terwilliger,T.C. and Waldo,G.S. (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.*, **23**, 102–107.
21. Hoskins,R.A., Stapleton,M., George,R.A., Yu,C., Wan,K.H., Carlson,J.W. and Celniker,S.E. (2005) Rapid and efficient cDNA library screening by self-ligation of inverse PCR products (SLIP). *Nucleic Acids Res.*, **33**, e185.
22. Cabantous,S. and Waldo,G.S. (2006) *In vivo* and *in vitro* protein solubility assays using split GFP. *Nat. Methods*, **3**, 845–854.
23. Listwan,P., Terwilliger,T.C. and Waldo,G.S. (2009) Automated, high-throughput platform for protein solubility screening using a split-GFP system. *J. Struct. Funct. Genomics*, **10**, 47–55.
24. Marley,J., Lu,M. and Bracken,C. (2001) A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR*, **20**, 71–75.
25. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
26. Wishart,D.S., Bigam,C.G., Yao,J., Abildgaard,F., Dyson,H.J., Oldfield,E., Markley,J.L. and Sykes,B.D. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, **6**, 135–140.
27. Tarendeau,F., Boudet,J., Guilligay,D., Mas,P.J., Bougault,C.M., Boulo,S., Baudin,F., Ruigrok,R.W., Daigle,N., Ellenberg,J. *et al.* (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat. Struct. Mol. Biol.*, **14**, 229–233.
28. Tarendeau,F., Crepin,T., Guilligay,D., Ruigrok,R.W., Cusack,S. and Hart,D.J. (2008) Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog.*, **4**, e1000136.
29. Guilligay,D., Tarendeau,F., Resa-Infante,P., Coloma,R., Crepin,T., Sehr,P., Lewis,J., Ruigrok,R.W., Ortin,J., Hart,D.J. *et al.* (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.*, **15**, 500–506.
30. Angelini,A., Tosi,T., Mas,P., Acajjaoui,S., Zanotti,G., Terradot,L. and Hart,D.J. (2009) Expression of Helicobacter pylori CagA domains by library-based construct screening. *FEBS J.*, **276**, 816–824.
31. Booker,G.W., Gout,I., Downing,A.K., Driscoll,P.C., Boyd,J., Waterfield,M.D. and Campbell,I.D. (1993) Solution structure and ligand-binding site of the SH3 domain of the p85 alpha subunit of phosphatidylinositol 3-kinase. *Cell*, **73**, 813–822.
32. Musacchio,A., Cantley,L.C. and Harrison,S.C. (1996) Crystal structure of the breakpoint cluster region-homology domain from phosphoinositide 3-kinase p85 alpha subunit. *Proc. Natl Acad. Sci. USA*, **93**, 14373–14378.
33. Nolte,R.T., Eck,M.J., Schlessinger,J., Shoelson,S.E. and Harrison,S.C. (1996) Crystal structure of the PI 3-kinase p85 amino-terminal SH2 domain and its phosphopeptide complexes. *Nat. Struct. Biol.*, **3**, 364–374.
34. Siegal,G., Davis,B., Kristensen,S.M., Sankar,A., Linacre,J., Stein,R.C., Panayotou,G., Waterfield,M.D. and Driscoll,P.C. (1998) Solution structure of the C-terminal SH2 domain of the p85 alpha regulatory subunit of phosphoinositide 3-kinase. *J. Mol. Biol.*, **276**, 461–478.
35. Derewenda,Z.S. (2010) Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 604–615.
36. Camacho,L.R., Ensergueix,D., Perez,E., Gicquel,B. and Guilhot,C. (1999) Identification of a virulence gene cluster of Mycobacterium tuberculosis by signature-tagged transposon mutagenesis. *Mol. Microbiol.*, **34**, 257–267.
37. Cox,J.S., Chen,B., McNeil,M. and Jacobs,W.R. Jr (1999) Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice. *Nature*, **402**, 79–83.
38. Oefner,P.J., Hunicke-Smith,S.P., Chiang,L., Dietrich,F., Mulligan,J. and Davis,R.W. (1996) Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic Acids Res.*, **24**, 3879–3886.

Domain trapping has proven to be a robust and efficient method for identifying soluble fragments from multi-domain proteins. This methodology has been successfully applied to PpsC and has resulted in the production of libraries of fragments centered on each catalytic domain of this enzyme (**J.D. Pedelacq *et al*., Figure 5**).

The most soluble and well-expressed fragments can be efficiently produced in a classical expression system involving the fast-growing bacterium *E. coli* and easy-to-use pET vectors for downstream applications. As highlighted in this publication, crystallization trials have yielded at least one diffraction-quality crystals for fragments representing the AT, DH and ER domains of PpsC. Subsequent structural analysis of these fragments have then demonstrated that domain trapping is able to correctly identify fragments encompassing one functional domain (or sub-domain, in the case of the ER domain), without prior knowledge of domain boundaries.

A summary of all the results obtained from the fragments studied during my PhD, including those concerning the AT, DH and ER domains, are presented in the next chapter.

N°	Domain(s)	Tag	Purif.	Amount	Conditions	Crystals	Res.	Phasing	Structure
1	KS	N	2	1 mg	1440				
2	AT	N	1	25 mg	960	Large platelet	1.8 Å	Molrep (2HG4)	X
2	AT	No	1	20 mg	960				
3	AT-DH	N	1	6 mg	960				
4	DH	N	1	50 mg	864	Bipyramidal	2.9 Å	MAD (Se-Met, 2 λ)	X
4	DH	C	1	50 mg	960	Bipyramidal	3.5 Å	Molrep	X
4	DH	No	2	30 mg	960	Bipyramidal	3.0 Å	Molrep	X
5	DH	N	1	60 mg	960	Rhomboedric	7.0 Å		
5	DH	No	1	30 mg	864				
6	DH-ER-KR	N	2	7 mg	1440	Thin needles			
7	DH-ER-KR	N	2	8 mg	864				
8	ER	N	1	20 mg	960				
9	ER	N	1	20 mg	960				
10	ER	N	1	20 mg	960				
11	ER	N	2	5 mg	1920	Bipyramidal	10.0 Å		
12	ER	N	2	8 mg	768				
13	ER	N	2	10 mg	960				
14	ER NB sub	N	1	60 mg	1152	Rhomboedric	1.5 Å	Molrep (1PQW)	X
15	ER-KR-ACP	N	2	4 mg	864				
15	ER-KR-ACP	C	2	4 mg	960				
16	KR	N	2	12 mg	960				
17	KR-ACP	C	2	2 mg	672				
18	KR-ACP	C	2	2 mg	480				
19	ACP	N	2	25 mg	1152				

Table 6. Crystallization study of PpsC fragments.

Fragment numbering is in accordance with Figure 24. Tag corresponds to the position of the 6His affinity tag (N-, C-terminal or No tag). Purif. corresponds to the protocol applied for purification, 1 means step 1 only, whereas 2 means both step 1 and step 2 (see **section II.A.3.B**). Amount corresponds to the quantity of pure protein obtained per liter of culture medium. Conditions are the number of different crystallization conditions tested. Res. is the maximum diffraction resolution observed. ER NB sub is the nucleotide-binding subdomain of the ER domain.

3. PPSC: A FRAGMENT STUDY

a) Overview of the PpsC Fragment Study

Classical pET vectors, bearing a six histidine tag either in N- or C-terminus, have been used to facilitate purification and the resulting constructions were produced by *Escherichia coli* BL21 (DE3). The most soluble ones were selected for downstream applications. In most cases, relatively small culture volumes (250 ml to 2 l) were sufficient to efficiently produce these fragments. Purification consisted in a batch affinity step using a cobalt-bound resin followed by dialysis. In some cases, an additional size exclusion chromatography on an automated AKTA system was needed. A total of 19 different fragments, representing all PpsC domains, have been brought to crystallization trials using this methodology (**Table 6 and Figure 24**). Experimental procedures are detailed in **section II.A.3.B**.

Diffraction-quality crystals have been obtained for fragments representing the AT domain, the DH domain and the nucleotide-binding subdomain of the ER domain. However, despite extensive crystallization trials on six different fragments representing the full-length ER domain, only low-resolution diffracting crystals could be obtained. This is in accordance with the absence of any structure of this domain from a PKS when it is expressed alone. The only structure containing the full-length ER domain is indeed that of the ER-KR didomain from DEBS module 4 [71]. KR and ACP domains also seem to be concerned by this issue, although only four fragments have been brought to crystallization in our study. Fragments representing the KS domain could only be produced at low temperatures (<20 °C), in small amounts and crystallization trials never yielded hits. Likewise, no KS domain expressed alone has ever been structurally characterized and all structures available concern KS-AT didomains from FASs or PKSs. We were then interested in di- and tri-domain fragments to establish the structure of these domains. So far, the most promising results have been obtained for fragments encompassing the DH, ER and KR domains. Thin protein needles that need to grow bigger and are, so far, not suitable for

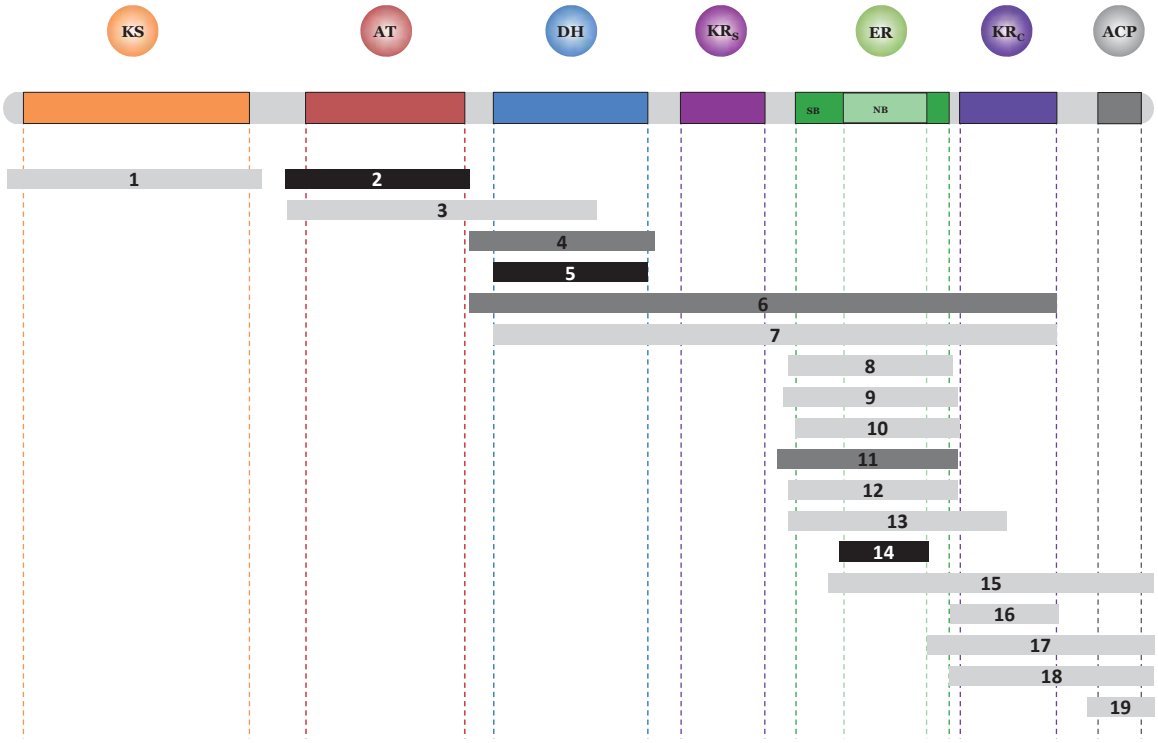


Figure 24. Fragment mapping onto the PpsC sequence.

Fragment numbering is in accordance with Table 6. Light grey boxes correspond to fragments for which no crystallization hit has been obtained. Dark grey boxes correspond to those which have been crystallized. Black boxes correspond to those which have been structurally characterized. SB stands for substrate-binding subdomain and NB for nucleotide-binding subdomain.

diffraction data collection, have been recently obtained (fragment 6, **see Figure 24**). To conclude with this study, we have been able to characterize the structures of the AT domain, of the DH domain and of the nucleotide-binding subdomain of the ER domain. The structure of the AT domain will be part of an article that will also include other PKS AT domains, which were structurally characterized in our research group. Therefore, it will not be discussed in this manuscript. The following chapter reports the functional and structural results obtained for the DH and ER domains.

b) Experimental Procedures

CLONING

All fragments were amplified by PCR using the Phusion High-Fidelity DNA polymerase (New England Biolabs), and a *ppsC* containing vector as matrix. Primers consisted in 22-mer complementary sequences bearing 5'-NdeI and 3'-BamHI restriction sites. Resulting amplicons were then digested and ligated into a modified pET28 plasmid bearing a STOP codon and a six-histidine tag at either side of the cloning site. Plasmids were transformed into chemically competent *E. coli* BL21 (DE3) cells.

PRODUCTION

Frozen cells expressing the different fragments were used to start an overnight 3 ml LB-kanamycin (112 μg/ml) culture at 37 °C prior to inoculation in baffled flasks containing 500 ml to 2 l of the same media. Cells were allowed to grow for approximately 2 h at 37 °C before temperature was dropped from 37 °C to 30 °C (or 20 °C in the case of KS-encompassing fragments). When OD_{600} reached 0.5–0.7, cells were induced with IPTG at a final concentration of 1 mM and grown for an additional 4 h (or 20 h in the case of KS-encompassing fragments) prior to harvesting by centrifugation at 4000*g* for 30 min and storage at –80 °C.

PURIFICATION

Step 1. The pelleted cells, grown in 1 l of LB broth, were suspended in 40 ml of buffer A (100 mM Tris, 150 mM NaCl pH 8.0) and lysed by sonication (6 cycles of 30 s pulse, 50% amplitude, power 5) prior to centrifugation at 40,000*g* for 30 min. The supernatant was poured on 3 ml of dried cobalt Talon Superflow resin (Clontech) previously washed with buffer A, and centrifuged at 300*g* for 5 min. The fragment-bound resin was then washed 3 times with buffer A by centrifugation. 30 ml of buffer A containing 10 mM imidazole was used to elute non-specific interactants, and after centrifugation, 5 ml of buffer A containing 250 mM imidazole was used to elute the fragment of interest.

Step 2. The eluted protein was concentrated down to 2 ml using a Vivaspin 20 centrifugal concentrator (Sartorius), prior to injection into a HiLoad 16/60 Superdex 75 ((or Superdex 200 for fragments 6 and 7,), (GE Healthcare)) pre-equilibrated with buffer B (25 mM Tris, 25 mM NaCl pH 8.0).

CRYSTALLIZATION TRIALS

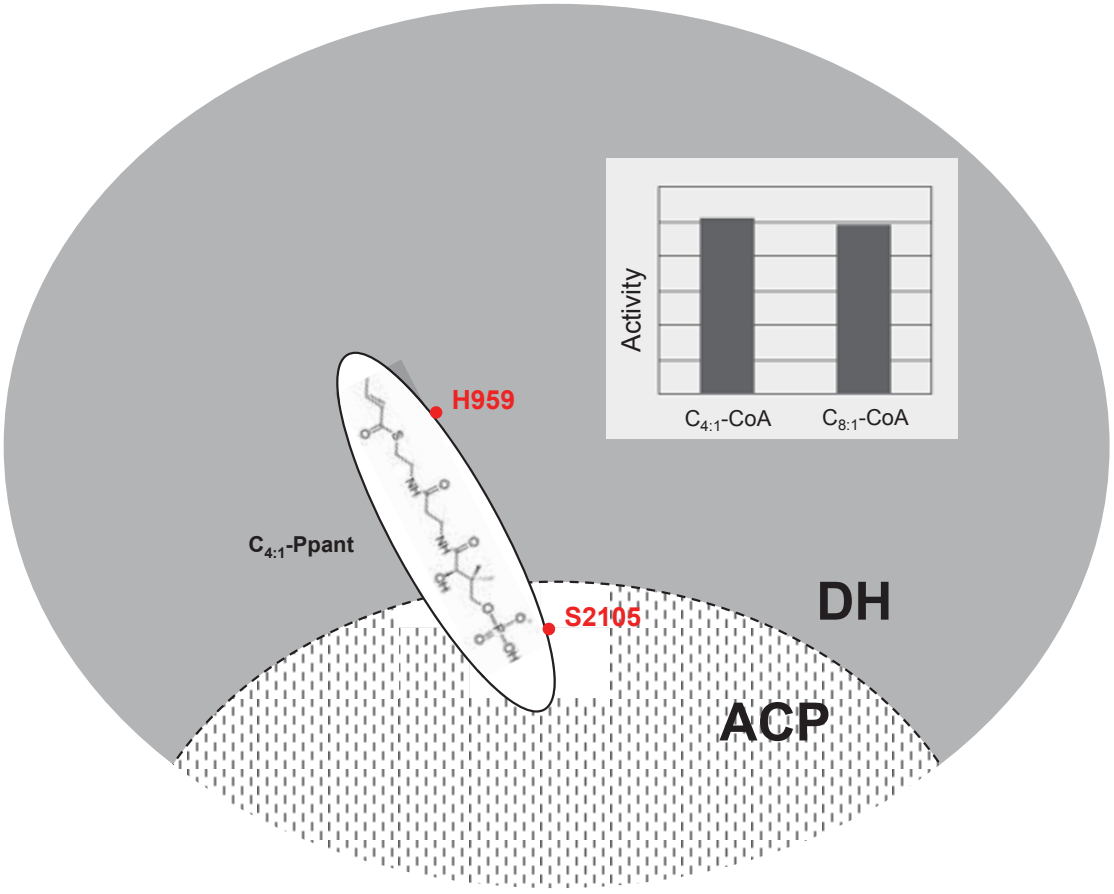
Crystallization trials were conducted with concentrated proteins using QIAGEN NeXtal commercial kits assisted by a Nanodrop ExtY crystallization robot. 400 nl drops consisting in 200 nl protein and 200 nl crystallization condition were set in Innovadyne Innovaplate SD-2 plates. Crystallization conditions that yielded crystals were refined in Linbro 24-well plates with 2 µl drops consisting in 1 µl protein and 1 µl crystallization condition.

4. THE DEHYDRATASE DOMAIN AND THE ASSEMBLY OF THE BETA-CARBON PROCESSING DOMAINS

To date, the DH domain substrate specificity and catalytic mechanism have been poorly characterized. Indeed, only suicide inhibitors that are not representative of a physiological substrate have been structurally identified with stand-alone dehydratases from type-II fatty acid synthases (FAS-II). Until now, determinants of substrate specificity were unknown with no experimental evidence demonstrating the activity of a given substrate, although a few conserved residues have been proposed to assist this transient catalytic mechanism. In addition to the X-ray structure of the native DH domain of PpsC, we have solved, for the first time, the structure of an inactive mutant version (DH-H959F) in complex with the substrate analog crotonyl-CoA ($C_{4:1}$-CoA), demonstrated as suitable for (de)hydration by *in vitro* enzymatic tests. We also report the structure of the nucleotide-biding sub-domain of the ER domain of PpsC. This structure has helped us model the assembly of the DH and ER domains of PpsC using the mammalian FAS (mFAS) as a template. Finally, the crotonyl-CoA bears a full-length phosphopantetheinyl moiety whose β-phosphate is normally covalently attached to the serine acceptor of the ACP domain. We have imposed a of 3 Å constraint between the two atoms to generate a docking model of the ACP domain of PpsC onto the DH domain.

All these results are presented in the following article that we are currently writing, which is then a temporary version:

Graphical Table of Contents



New Insights into Substrate Modification by Dehydratases from Type I Fatty Acid and Polyketide Synthases

Alexandre Faille[1,2], Nawel Slama[1,2], Annaïk Quémard[1,2], Valérie Guillet[1,2], Lionel Mourey[1,2,*] & Jean-Denis Pedelacq[1,2,*]

[1]Institut de Pharmacologie et de Biologie Structurale (IPBS), Centre National de la Recherche Scientifique (CNRS), 205 route de Narbonne, BP 64182, F-31077 Toulouse, France; [2]Université de Toulouse, Université Paul Sabatier, IPBS, F-31077 Toulouse, France

*To whom correspondence should be addressed. Tel: +33 5 61 17 54 11; Fax: +33 5 61 17 59 94; Email: Jean-Denis.Pedelacq@ipbs.fr

Correspondence may also be addressed to Lionel Mourey. Tel.: +33 5 61 17 54 36; Fax: +33 5 61 17 59 94; Email: Lionel.Mourey@ipbs.fr

Abstract

Dehydration reactions play a crucial role for *de novo* biosynthesis of fatty acids and a wide range of pharmacologically active polyketide natural products with strong emphasis in human medicine. Until now, the dehydration mechanism remained poorly documented due to the lack of structural studies implying metabolizable substrates. Here, we describe the first crystal structure of a dehydratase domain, from the *Mycobacterium tuberculosis* type I polyketide synthase PpsC, in complex with the substrate analog *trans*-2-butenoyl-CoA for which activity was confirmed *in vitro*. Residues essential for binding and activity have been identified and a dehydration mechanism has been confirmed. We have also determined the X-ray structure of the dimeric nucleotide-binding subdomain of the enoylreductase domain that we used to propose an assembly model for the dehydratase, enoylreductase and acyl carrier protein domains of PpsC, similar to mammalian fatty acid synthase.

Polyketides represent the most important class of naturally synthesized metabolites in human health with proven antibacterial, anticholesterolemic, immunosuppressant, antifungal, anti-cancer, antiparasitic, and insecticide activities[1]. They are biosynthesized by complex machineries, named polyketide synthases (PKS), that human pathogens such as *M. tuberculosis* use in combination with fatty acid synthases (FAS) to produce lipid virulence factors. Indeed, these polyketide-derived lipids form a permeability barrier within the mycobacterial cell envelope, which constitutes the first line of defense against toxic molecules from the host and confers intrinsic resistance against treatments[2,3]. Among these virulence factors, phthiocerol dimycocerosates (PDIM) participate directly into the early steps of microbial infection through a mechanism of host plasma membrane rearrangement that encourages phagocytosis and inhibits phagosome acidification[4]. More recently, Cambier and colleagues have highlighted a central role for PDIM in masking pathogen-associated molecular patterns, thus preventing the production of microbicidal reactive nitrogen species through the Toll-like receptor recruitment of macrophages[5].

The *M. tuberculosis* genes involved in the biosynthesis of PDIM and closely related phenolic glycolipids (PGL) are clustered in a 73-kb region of the chromosome[6]. Biosynthesis of these components implies the concerted action of an ensemble of six type I PKS: five mono-modular phenol phtiocerol synthases (PpsA to E) responsible for the elongation of long-chain fatty acid backbones that form the phtiocerol (or phenolphtiocerol in the case of PGL biosynthesis) moiety and the mycocerosic acid synthase (Mas) for the production of multi-methyl branched mycocerosates[7]. PDIM result from the esterification of two mycocerosic acids onto the hydroxyl groups of the phthiocerol, a reaction mediated by the acyltransferase PapA5[7].

Both type I PKS and FAS belong to the superfamily of megasynthases characterized by the presence of multiple catalytic domains onto the same polypeptide chain. Three domains are essential for their activity: an acyl transferase (AT) domain, often referred to as the "gatekeeper", which catalyzes the transfer of a starter/elongation unit to its cognate acyl carrier protein (ACP) domain, and a ketosynthase (KS) domain for the condensation of the growing chain with the elongation unit to give an ACP bound β-keto ester. Optional modification reactions, also referred to as β-carbon processing, involve the enzymatic ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains. First, the β-keto ester undergoes a reduction by the KR domain to yield a hydroxyl subsequently dehydrated by the DH domain, thus resulting in carbon-carbon double bond formation. The growing chain can also be reduced by the ER domain to yield a saturated carbon chain longer than the original one by two methylene units. Overall, the catalytic mechanism underlying β-carbon modifications remains poorly described as no structure of individual DH, ER and KR domains in complex with their physiological substrates is yet available. Only suicide inhibitors in complex with stand-alone FabA[8] and FabZ[9] dehydratases from *Pseudomonas aeruginosa* have been published so far. In this paper, we report for the first time the three-dimensional structure of a DH domain from the type I mono-modular PKS PpsC in complex with a substrate analog for which activity was confirmed *in vitro*. We also solved the X-ray structure of the PpsC ER domain that we used to propose a DH-ER-ACP assembly model in accordance with the mammalian FAS structure[10].

Two libraries of fragments centered onto the DH and ER domains of *M. tuberculosis* PpsC were created to select for well-expressed and soluble constructs, using the domain trapping method[11] (**Figure 1a**). DH fragments A (884-1231) and B (921-

1222) could be easily purified and crystallized (**Figure 1b**). Small crystals of fragment A never diffracted beyond 8-10 Å despite the implementation of extensive optimization screens. On the other hand, tetragonal bipyramidal crystals of fragment B could grow overnight and diffracted to a maximum resolution of 2.7 Å (**Supplementary Table 1**). The ER domain was initially represented by soluble fragments C (1465-1786), D (1470-1784), E (1477-1784), and F (1477-1781), with N- and C-termini close to the predicted boundaries for this domain (**Figure 1b**). Extensive crystallization attempts failed to produce good quality crystals for these fragments, although nice bipyramids could be obtained for fragment C. The search for the shorter fragment G (1558-1758) was driven by the fact that only the nucleotide-binding subdomain of the ER domain interact with DH in the mFAS X-ray structure[12]. Diffraction quality crystals of fragment G were obtained and X-ray data were collected to a maximum resolution of 1.5 Å (**Supplementary Table 1**).

The DH domain belongs to the hotdog-fold enzyme superfamily whose first identified member was the *Escherichia coli* type II dehydratase FabA[9]. Unlike FabA, which functions as a homodimer with hotdog folds from each monomer associating to create an extended β-sheet, the PpsC dehydratase double hotdog occurs in a single polypeptide chain. Landmarks include a 13-stranded β-sheet (topology β7-β8-β9-β11-β12-β13-β10-β3-β6-β5-β4-β2-β1), hotdog helices αHD1 and αHD2, a cap motif comprising helices η1, η2, and η3, a three-stranded β-sheet (β2a, β2b, and β7a), and helix αC1 (**Figures 2a & Supplementary Figure 1a**). Residues in loop regions (1049-1063 and 1144-1158) that connect the two hotdog motifs are missing from the refined X-ray structure. The search for structural homologues in the Protein Data Bank (http://www.rcsb.org) highlighted some strong similarities with dehydratase domains

from the curacin A biosynthetic pathway, CurF being the closest structural homologue[13], and from mFAS[12], with 2.1-2.5 Å root mean square deviation between superimposed Cα atoms. The structural alignment also revealed that ~85% of the DH Cα atoms could be superimposed onto the CurF homologue and ~70% in the case of the mFAS X-ray structure, with most structural differences occurring in the cap motif (**Supplementary Figures 1a,b**).

In an attempt to elucidate the dehydration mechanism of the DH domain, we conducted a series of experiments using *trans*-2-butenoyl-CoA ($C_{4:1}$-CoA) and *trans*-2-octenoyl-CoA ($C_{8:1}$-CoA), two molecules resembling the natural substrate with an intact 4'-phosphopantetheine (P-pant) moiety (**Figure 2b**). We chose to monitor the hydration of these substrate analogues since isolated dehydratases preferentially catalyze hydration rather than dehydration *in vitro*[14]. Disappearance of the C=C double bond at 263 nm showed a specific activity of 0.5 $\mu mol.min^{-1}.mg^{-1}$ in the presence of either substrate (**Figure 2c**), thus indicating that the reaction rate is not limited by the length of the aliphatic carbon chain covalently attached to the P-pant moiety. Neither co-crystallization nor soaking of wild-type DH crystals with $C_{4:1}$-CoA could give rise to the structure of a stable complex. Instead, we focused our efforts on engineering an inactive mutant that could sequester the substrate within the active site cavity. We relied onto two independent studies that had shown that mutating the catalytic His residue into a phenylalanine was sufficient to abolish the activity of the DH domains of DEBS[15] and the picromycin/methymycin synthase[16]. Using this strategy, crystals of the PpsC DH domain bearing the analogous mutation H959F were soaked with $C_{4:1}$-CoA. The presence of $C_{4:1}$-CoA in the active site cavity was confirmed after examination of the 2Fo-Fc and Fo-Fc difference Fourier maps (**Figure 2a**). Hydrogen bond interactions occur between the P-

pant group and residues G969, R1000, M1002, and P1163 (**Figure 2a**). The highly conserved G969 interacts with the substrate ketone via its main-chain nitrogen and guides the substrate towards the active site tunnel (**Figures 2a,d**). Interestingly, this ketone superimposed well with the bound phosphate found in the DH apo structure (**Figure 2d**). The cavity entrance at the junction between the two hotdogs is delineated by two stretches of residues located between strands β3 and β4 (positions 999 to 1002) and within strand β10 (positions 1162 to 1164) (**Figure 2d, top**). In the apo structure, V966 and Q1084 make hydrogen bonds with the catalytic H959 that maintain its side chain in a position suitable for hydration of the carbon-carbon double bond (**Figure 3a**). The strictly conserved Y1076 is part of a polar interaction network that probably increase the pKa value of the carboxylic acid group of D1129, thus allowing it to catalyze dehydration at a physiological pH. In the holo form, the D1129 side-chain would facilitate the proton transfer to the 3'-hydroxyl of the substrate, while the histidine side-chain nitrogen ε2 could abstract a proton from C2 of the substrate, thus forming a carbon-carbon double bond between C2 and C3 (**Figure 3b**). This dehydration reaction also generates a water molecule that makes interactions with G969 and the side chain of D1129 (**Figure 3c**). This water molecule had previously been hypothesized as the dehydration side product for the DEBS module 4 homologue[17].

To date, FAS remains the only megasynthase for which X-ray structures of the full-length enzyme have been described[10,18,19]. mFAS is the closest homologue to type I PKS, in terms of module organization[10] and sequence and structural similarities. However differences in protomer-protomer association exist between isolated DH domains from PpsC, CurF[13], DEBS module 4[17] type I PKS and mFAS[10]. This change is orchestrated by the intertwined "crab claw"-like structure at the dimer interface, that slightly opens up in

mFAS to adopt a V-shaped structure (**Figures 4a,b**). Dimerization of the PpsC DH domain occurs both in the crystal structure and in solution, for which a molecular weight of 75.3 ± 1.2 kDa has been measured by size exclusion chromatography multi-angle laser light scattering (SEC-MALLS) experiments (**Supplementary Figure 2**). Although we could not confirm a mFAS-like V-shaped structure of our DH dimer in solution and in the crystal structure, evidences that the β-carbon processing DH and ER domains could interact similarly as in mFAS (**Supplementary Figure 3a**) was reinforced by the observed variability in the DH dimerization mode in addition to the dimer interface of the PpsC ER X-ray structure that may play a key role in stabilizing the complex. Indeed, the linear dimerization mediated by an extended 12-stranded β-sheet in the PpsC ER structure is also shared by mFAS, whereas the presence of an extra α-helical segment in the multimodular spinosyn PKS structure prevents dimerization (**Figure 4c**)[20]. We then generated a model of PpsC DH-ER interactions using mFAS as a template, where a PpsC DH protomer was first superimposed onto each of the two protomers constituting the mFAS DH dimer, prior to the positioning of the ER dimer (**Supplementary Figure 3b**). We felt confident that the resulting model could provide a fair picture of the PpsC DH-ER assembly. Indeed, steric clashes were only found between the side-chain of residue K810 from the ER domain and G263 from the loop that connects β11 to β12 in the DH domain (**Supplementary Figure 3c**). This loop region may play a major role in stabilizing the DH-ER interface as it protrudes from the 12-stranded β-sheet into a cleft with potential interactions with main-chain and side-chain atoms of residues from each protomer of the ER domain. In an attempt to address how the ACP domain could fit into our model of the DH-ER assembly, we first generated a DH-ACP complex by imposing a distance constraint of 3 Å between the β-phosphate of the CoA derivative in the active

site cavity of the DH domain and the serine acceptor of the ACP domain. Although the three-dimensional structure of the PpsC ACP domain is still unknown, a structural alignment with homologues from FAS and other type I PKS indicates the conservation of a four-helix bundle motif covering the entire domain. We first generated a PpsC ACP model based on the module 2 of DEBS, with respect to the criteria of highest sequence identity (~37 %) and conservation of key residues in the active site. Modeling of the PpsC DH-ACP interface led to only one solution permitting favorable interactions of the ACP domain with the ER and DH domains (**Supplementary Figures 3c,d**). Recently, the X-ray structure of a crosslinked complex between the dehydratase FabA and the acyl carrier protein AcpP involved in the fatty acid biosynthesis pathway of *E. coli* has been determined[21]. Although the covalently bound probe used for crosslinking follows the same route in FabA as in PpsC DH, the orientation of the ACP domain differs substantially. Examination of the FabA-AcpP interface highlighted significant differences with the mFAS structure and our DH-ACP complex model that would lead to severe steric clashes with the ER domain. As suggested by the authors, the interactions observed in the FabA-AcpP complex are not relevant to megasynthases for which an atomic level description of the interface still remains to be elucidated[21].

In conclusion, we have determined the first structure of a β-carbon processing domain from a type I PKS, i.e. the dehydratase domain from PpsC, in complex with the non-covalently bound substrate analog *trans*-2-butenoyl-CoA ($C_{4:1}$-CoA) whose activity has been clearly demonstrated *in vitro*. To stabilize the complex, the universally conserved catalytic histidine residue was mutated into a phenylalanine, an approach that can easily be transposed to any dehydratase whether it is part of a megasynthase or a stand-alone enzyme. A detailed structural comparison of the PpsC DH domain in its apo

form and in complex with *trans*-2-butenoyl-CoA allowed us to unambiguously identify residues implicated in the dehydration mechanism. This should enable the engineering of modified DH domains of PKS and FAS to process non-natural substrates with improved or novel pharmacological properties.

Methods

PpsC fragment library, ORF selection, and cloning. The open reading frame for the *M. tuberculosis ppsC* gene was amplified by PCR using forward 5'-GATATACATATGACCGCAGCGACACCAGATCG-3' and reverse 5'-AATTCACTAGTTGACTCGCCTCGCGTCGCAGC-3' primers. The underlined bases represent the NdeI and SpeI restriction sites, respectively. A large size DNA library (850-1650 bp) was created from cleaned PCR product using a HydroShear device from Genomics Solutions (Ann Arbor, MI). Extremities of the fragments were polished using Vent polymerase (New England Biolabs, Beverly, MA) at 72°C for 20 min. A slab of the gel containing the DNA fragments region 850-1650 bp was excised and recovered with a QIAquick® gel extraction kit (Qiagen). Library screening for open reading frames (ORFs) into a permissive site of the *E. coli* dihydrofolate reductase (DHFR) was conducted as previously described[11]. Recovered DHFR insertion library was diluted for plasmid preparation. Gel extracted and cleaned inserts from NdeI/SpeI restriction digests were ligated into the pTET ColE1 GFP11 vector and transformed into chemically competent

BL21 (DE3), pET GFP 1–10 cells. *In vivo* solubility screenings were performed as previously described [22].

Enzymatic assays. Hydratase activity was monitored at 263 nm using a thermostated Uvikon 923 spectrophotometer (Kontron Instruments) in the presence of *trans*-2-enoyl-CoA (ΔA of 0.67 for a variation of concentration of 100 µM). Kinetic assays were performed in a quartz cuvette for 1.5 min at 25°C, in 100 mM sodium phosphate buffer pH 7.0 in the presence of *trans*-2-octenoyl-CoA ($C_{8:1}$-CoA, 25 µM) or *trans*-2-butenoyl-CoA ($C_{4:1}$-CoA, 10 or 25 µM). After equilibration of the baseline, reactions were started by adding purified DH domain (fragment A) at 125 nM or 250 nM. $C_{8:1}$-CoA was synthesized as previously described (Quémard et al., 1995) and $C_{4:1}$-CoA was purchased from Sigma-Aldrich.

Additional Methods. For details on protein production and purification, crystallization and structure determination, modeling of the assembly of the DH-ER-ACP domains, SEC-MALLS experiments, please see Supporting Information.

References

(1) Weissman, K. J., and Leadlay, P. F. (2005) Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol. 3*, 925–36.

(2) Nguyen, L., Chinnapapagari, S., Charles, J., and Thompson, C. J. (2005) FbpA-Dependent Biosynthesis of Trehalose Dimycolate Is Required for the Intrinsic Multidrug Resistance , Cell Wall Structure , and Colonial Morphology of Mycobacterium smegmatis FbpA-Dependent Biosynthesis of Trehalose Dimycolate Is Required for the Intri. *J. Bacteriol. 187*, 6603–6611.

(3) Jarlier, V., and Nikaido, H. (1990) Permeability barrier to hydrophilic solutes in Mycobacterium chelonei. *J. Bacteriol. 172*, 1418–23.

(4) Astarie-Dequeker, C., Le Guyader, L., Malaga, W., Seaphanh, F.-K., Chalut, C., Lopez, A., and Guilhot, C. (2009) Phthiocerol dimycocerosates of M. tuberculosis participate in macrophage invasion by inducing changes in the organization of plasma membrane lipids. *PLoS Pathog. 5*, e1000289.

(5) Cambier, C. J., Takaki, K. K., Larson, R. P., Hernandez, R. E., Tobin, D. M., Urdahl, K. B., Cosma, C. L., and Ramakrishnan, L. (2014) Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature 505*, 218–22.

(6) Siméone, R., Constant, P., Guilhot, C., Daffé, M., and Chalut, C. (2007) Identification of the missing trans-acting enoyl reductase required for phthiocerol dimycocerosate and phenolglycolipid biosynthesis in Mycobacterium tuberculosis. *J. Bacteriol. 189*, 4597–602.

(7) Trivedi, O. a, Arora, P., Vats, A., Ansari, M. Z., Tickoo, R., Sridharan, V., Mohanty, D., and Gokhale, R. S. (2005) Dissecting the mechanism and assembly of a complex virulence mycobacterial lipid. *Mol. Cell 17*, 631–43.

(8) Moynié, L., Leckie, S. M., McMahon, S. a., Duthie, F. G., Koehnke, A., Taylor, J. W., Alphey, M. S., Brenk, R., Smith, A. D., and Naismith, J. H. (2013) Structural Insights into the Mechanism and Inhibition of the β-Hydroxydecanoyl-Acyl Carrier Protein Dehydratase from Pseudomonas aeruginosa. *J. Mol. Biol. 425*, 365–377.

(9) Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M., and Smith, J. L. (1996) Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure 4*, 253–64.

(10) Maier, T., Leibundgut, M., and Ban, N. (2008) The crystal structure of a mammalian fatty acid synthase. *Science 321*, 1315–22.

(11) Pedelacq, J.-D., Nguyen, H. B., Cabantous, S., Mark, B. L., Listwan, P., Bell, C., Friedland, N., Lockard, M., Faille, A., Mourey, L., Terwilliger, T. C., and Waldo, G. S. (2011) Experimental mapping of soluble protein domains using a hierarchical approach. *Nucleic Acids Res. 39*, e125.

(12) Maier, T., Leibundgut, M., and Ban, N. (2008) Supporting Info - The crystal structure of a mammalian fatty acid synthase. *Science 321*, 1315–22.

(13) Akey, D., Razelun, J., and Tehranisa, J. (2010) Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure 18*, 94–105.

(14) Rock, C., and Cronan, J. (1996) Escherichia coli as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochim. Biophys. Acta 1302*, 1–16.

(15) Bevitt, D. J., Staunton, J., and Leadlay, P. F. (1993) Mutagenesis of the dehydratase active site in the erythromycin-producing polyketide synthase. *Biochem. Soc. Trans. 21*, 30S.

(16) Wu, J., Zaleski, T. J., Valenzano, C., Khosla, C., and Cane, D. E. (2005) Polyketide double bond biosynthesis. Mechanistic analysis of the dehydratase-containing module 2 of the picromycin/methymycin polyketide synthase. *J. Am. Chem. Soc. 127*, 17393–404.

(17) Keatinge-Clay, A. (2008) Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol. 384*, 941–953.

(18) Jenni, S., Leibundgut, M., Boehringer, D., Frick, C., Mikolásek, B., and Ban, N. (2007) Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science 316*, 254–61.

(19) Leibundgut, M., Jenni, S., Frick, C., and Ban, N. (2007) Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science 316*, 288–90.

(20) Zheng, J., Gay, D. C., Demeler, B., White, M. a, and Keatinge-Clay, A. T. (2012) Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol. 8*, 615–21.

(21) Nguyen, C., Haushalter, R. W., Lee, D. J., Markwick, P. R. L., Bruegger, J., Caldara-Festin, G., Finzel, K., Jackson, D. R., Ishikawa, F., O'Dowd, B., McCammon, J. A., Opella, S. J., Tsai, S.-C., and Burkart, M. D. (2014) Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature 505*, 427–31.

(22) Cabantous, S., and Waldo, G. (2006) In vivo and in vitro protein solubility assays using split GFP. *Nat. Methods 3*, 845–854.

Acknowledgment

This work was supported by the Agence Nationale de la Recherche (XPKS-MYCO, grant 09-BLAN-0298-01). The equipment used for crystallization experiments and X-ray crystallography are part of the Integrated Screening Platform of Toulouse (PICT, IBiSA). We thank the staff of synchrotron beamlines ID14 and ID29 at the European Synchrotron Radiation Facility (Grenoble, France).

Figure Legends

Figure 1. The domain trapping strategy. (a) Our strategy integrates a DHFR filter to eliminate the incorrect reading frames and a split-GFP solubility screen to identify the soluble constructs. A 850-1500 bp library of fragments was inserted at position 86/87 of the DHFR. Inverse PCR focused sub-libraries of fragments helped identifying soluble fragments centered onto a specific region of the protein of interest. **(b)** The top 25% most soluble clones centered onto the DH and ER domains of the *M. tuberculosis* polyketide synthase PpsC are represented. Fragments that permitted the X-ray structure determination of the DH and ER domains are colored in gray.

Figure 2. **Structure and activity of the DH domain of the *M. tuberculosis* PpsC. (a)** Ribbon representation of the X-ray structure of the DH domain in complex with *trans*-2-butenoyl-CoA. The active site H959 has been mutated into a Phe. The presence of $C_{4:1}$-CoA in the active site cavity was confirmed after examination of the difference Fourier 2Fo-Fc (blue) and Fo-Fc (green) maps contoured at 1σ and 3σ levels, respectively. **(b)** Enzymatic assays were performed using *trans*-2-butenoyl-CoA ($C_{4:1}$-CoA) and *trans*-2-octenoyl-CoA ($C_{8:1}$-CoA), two molecules resembling the natural substrate. **(c)** Disappearance of the C=C double bond at 263 nm indicates a specific activity of 0.5 $\mu mol.min^{-1}.mg^{-1}$ in the presence of either $C_{4:1}$-CoA or $C_{8:1}$-CoA. **(d)** Top, the cavity entrance is delineated by two stretches of residues located between strands β3 and β4 (positions 999 to 1002) and within strand β10 (positions 1162 to 1164) with small differences in the shape and size between the apo (left) and holo (right) X-ray structures. A phosphate molecule is present in the active site of the apo enzyme whereas a water molecule (black ball) is at hydrogen bond distance to the thioester group of $C_{4:1}$-CoA. Bottom, orthogonal view around the vertical axis.

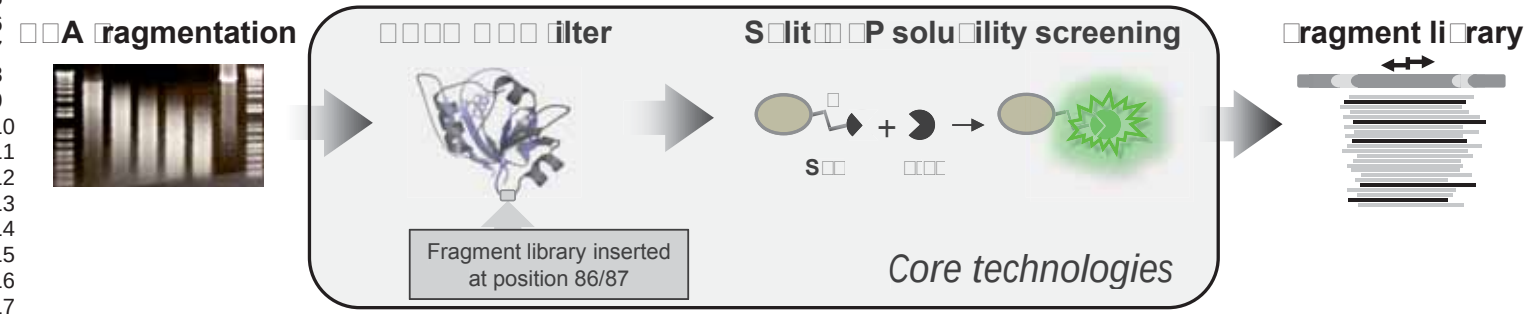
Figure 3. Catalytic mechanism of the dehydration reaction. Ball and stick representation of residues involved in the dehydration reaction as seen in the structure of (**a**) wild-type apo DH and (**b**) the DH-H959F mutant in complex with *trans*-2-butenoyl-CoA. (**c**) Proposed dehydration reaction mechanism.

Figure 4. The PpsC DH and ER dimers. (a) Ribbon diagram representation of the DH dimer present in the PpsC crystal structure and **(b)** its comparison with known homologues from DEBS module 4 (PDB code 3EL6), CurF (PDB code 3KG6) and mFAS (PDB code 2VZ8). Reference molecule in black is in the same orientation. Interacting residues at the interface are represented as spheres. **(c)** X-ray structure of the nucleotide-binding subdomain of the ER domains from PpsC (left), mFAS (middle) and

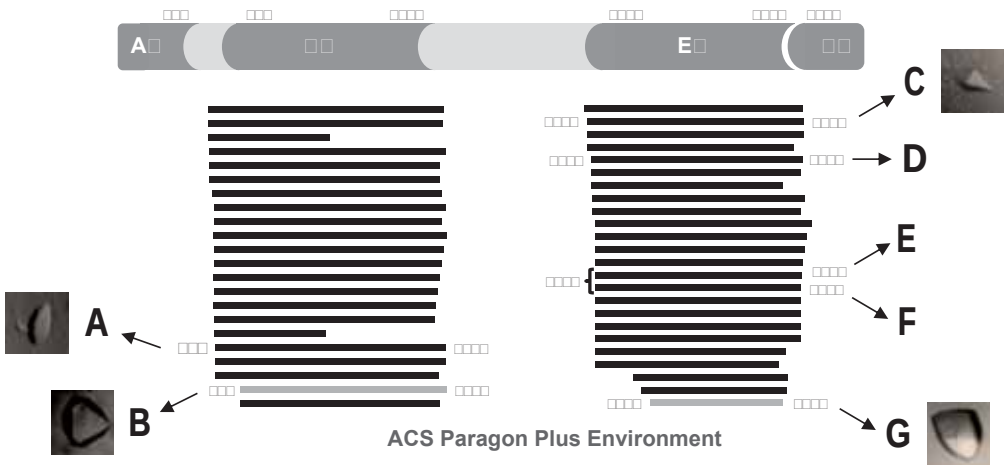
module 2 of the Spinosyn PKS (PDB code 3SLK, right). The extra α-helical segment in the spinosyn PKS structure is represented in red.

□igure □□□he □omain tra□□ing strategy□

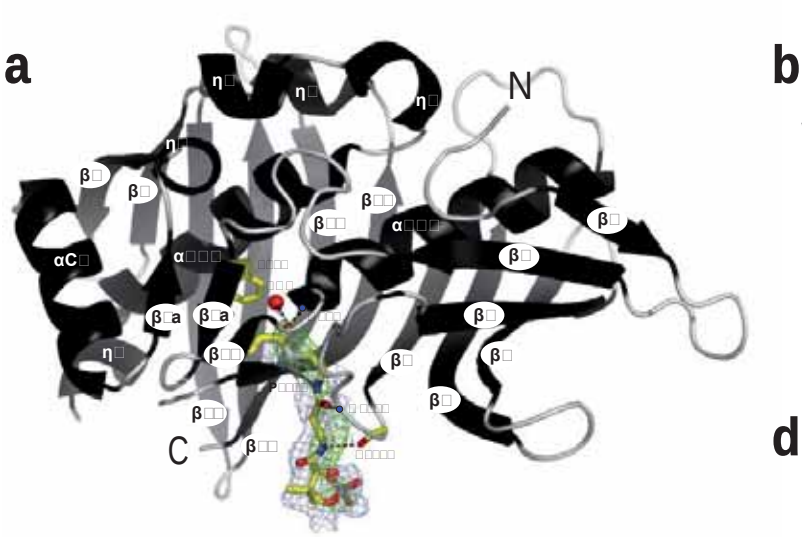
a

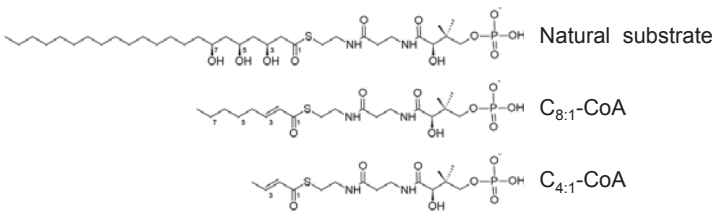


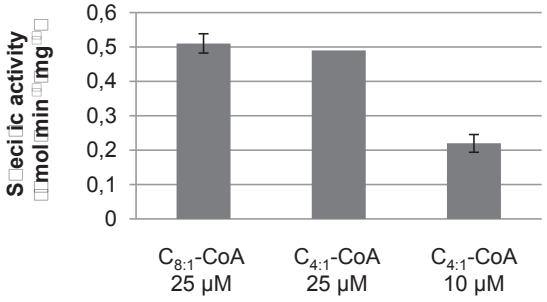
b

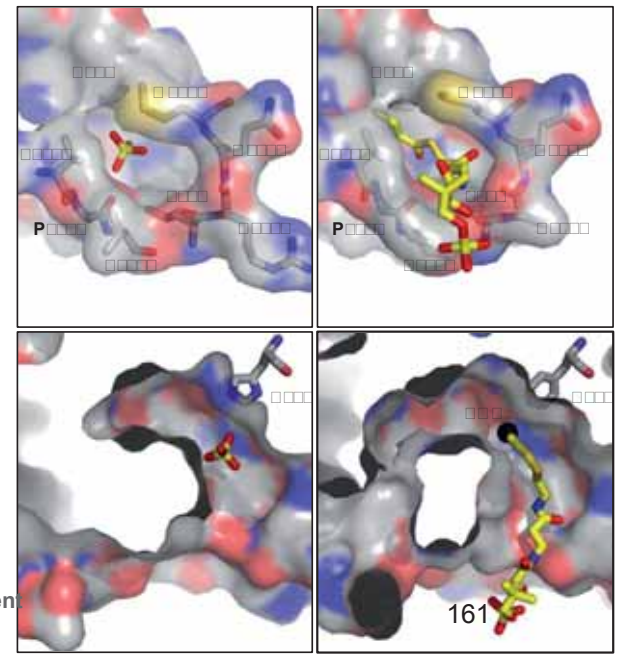


□igure □. Structure an□ activity o□the □□ □omain o□the *M. tuberculosis* P□sC□

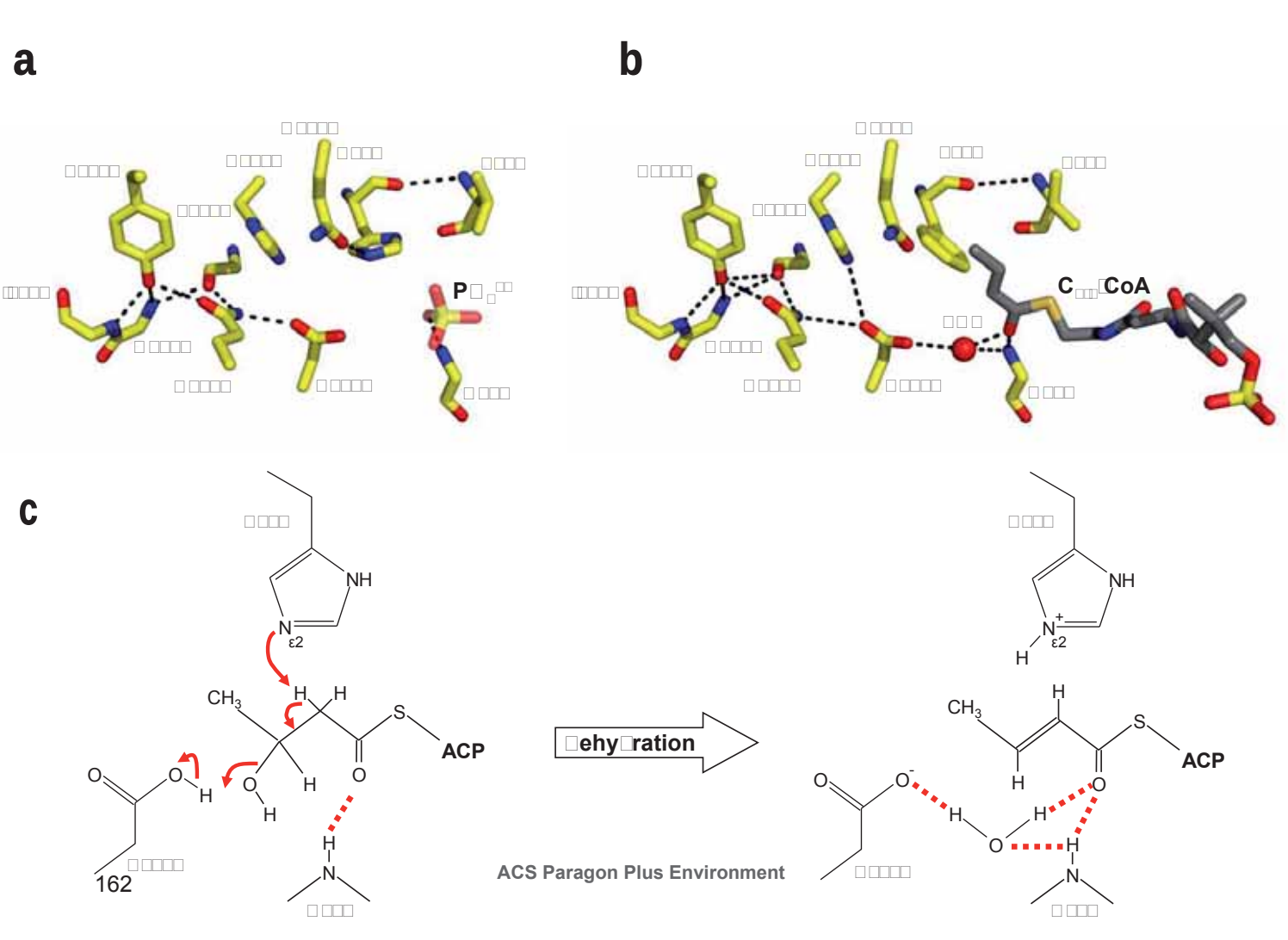




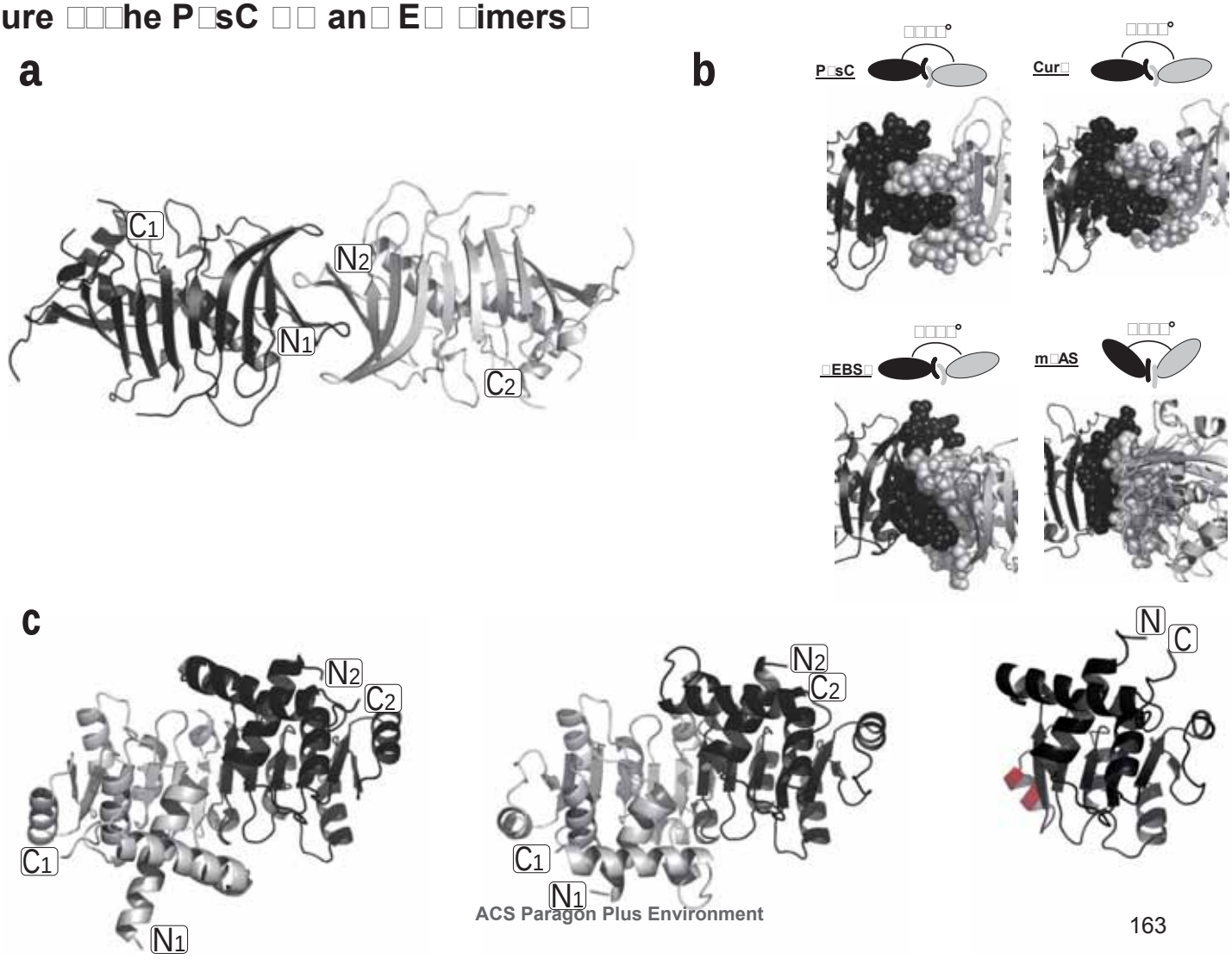




□igure □□Catalytic mechanism o□the □ehy□ration reaction□



□igure □□□he P□sC □□ an□ E□ □imers□



Supporting Information

New Insights into Substrate Modification by Dehydratases from Type I Fatty Acid and Polyketide Synthases

Alexandre Faille[1,2], Nawel Slama[1,2], Annaïk Quémard[1,2], Valérie Guillet[1,2], Lionel Mourey[1,2,*] & Jean-Denis Pedelacq[1,2,*]

[1]Institut de Pharmacologie et de Biologie Structurale (IPBS), Centre National de la Recherche Scientifique (CNRS), 205 route de Narbonne, BP 64182, F-31077 Toulouse, France; [2]Université de Toulouse, Université Paul Sabatier, IPBS, F-31077 Toulouse, France

*To whom correspondence should be addressed. Tel: +33 5 61 17 54 11; Fax: +33 5 61 17 59 94; Email: Jean-Denis.Pedelacq@ipbs.fr

Correspondence may also be addressed to Lionel Mourey. Tel.: +33 5 61 17 54 36; Fax: +33 5 61 17 59 94; Email: Lionel.Mourey@ipbs.fr

Experimental procedures

Protein production and purification

Frozen cells from constructs A, B, C and G (**Figure 1b**) were used to grow 3 ml LB-kanamycin (35 µg/ml) cultures overnight at 32˚C prior to inoculation in baffled flasks containing 500 ml of the same media. Cells were allowed to grow for approximately 2 h at 37˚C before temperature was dropped from 37˚C to 30˚C. When OD_{600} reached 0.5-0.7, cells were induced with IPTG at a final concentration of 0.5 mM and grown for an additional 4 h prior to harvesting by centrifugation at 4000×*g* for 20 min. Selenomethionine-labeled fragment B was produced following the protocol described by Guerrero and colleagues [1]. Cell pellets from both wild-type and selenomethionylated protein fragments were resuspended in 25 ml buffer A (100 mM Tris pH 8.1, 150 mM NaCl) and lysed by sonication on ice. 3 ml of washed cobalt Talon Superflow resin (GE Healthcare 28-9575-02) was then added to the supernatant and the resulting mixture was shaken for a few seconds allowing the 6His tagged proteins to bind the cobalt resin. Two cycles of washing were then realized before the elution of non-specific contaminants with buffer A containing 10 mM imidazole and finally the elution of the bound protein with buffer A containing 250 mM imidazole. 6 ml eluted solution was dialyzed twice against 1 l of buffer B (Tris pH 8.1 25 mM, 25 mM NaCl). Protein purity was assessed on 10% SDS-PAGE along with PageRuler Prestained Protein Ladder (Thermo Scientific #26616) followed by Coomassie blue detection.

Crystallization and structure determination

Highly purified proteins were used for crystallization using commercially available kits from Qiagen (Venlo, Netherlands) in Greiner 3 round plates (Stonehouse, UK) and the hanging drop vapor diffusion technique. Optimization of crystallization conditions was realized

manually in Linbro plates (Hampton Research, Aliso Viejo, USA) containing 500 µl reservoir solution. For crystallization of the DH domain, 1 µl of protein (30 mg/ml) was mixed with 1 µl of 1.8 M Na/K phosphate. For crystallization of the ER domain, 1 µl of protein (16 mg/ml) was mixed with 1 µl of 10% MPD, 0.1 M MES pH 6.0. Crystals of the DH fragment B belong to space group $P4_32_12$ with cell parameters a=b=89.5 Å, c=159.3 Å. Crystals of the ER fragment G belong to space group $P2_12_12_1$ with cell parameters a=59.9 Å, b=78.7 Å, c=88.8 Å. Crystals of fragment B were soaked in paratone for 5 seconds prior to flash freezing in liquid nitrogen whereas crystals of fragment G were directly flash frozen.

Native data sets were collected for the DH and ER domains at the ESRF beamlines ID29 and ID14 to a maximum resolution of 2.7 Å and 1.5 Å, respectively (**Table 1**). Peak and inflection datasets for selenomethionylated crystals of the DH domain were collected at the selenium absorption edge (ID29, ESRF). The DH structure was solved by the multi-wavelength anomalous dispersion method using anomalous scattering from the six selenium-substituted methionine residues. Datasets were indexed using MOSFLM[2] and scaled with SCALA[3]. Phase calculation and density modification were realized with the SHELX software suite[4]. The ER domain structure was solved by molecular replacement using PHASER[5] and PDB code 1PQW as the search model. Structures were then built and refined by iterative cycles of manual model building in COOT[6] and refinement using REFMAC5[7] included in the CCP4 software suite[8].

Size exclusion chromatography multi-angle laser light scattering (SEC-MALLS) experiments

Prior to SEC-MALLS experiments, the affinity-purified DH fragment B was loaded onto a Superdex 75 analytical column (GE Healthcare) equilibrated with 300 mM Na/K phosphate

pH 6.9. Then, 80 µl protein solution at 0.28 mg/ml were loaded onto a Shodex (300A) column (Wyatt Technology, France) using an Agilent 1260 Infinity LC chromatographic system (Agilent Technology) coupled to a multi-angle laser light scattering (MALLS) detection system. The column was equilibrated with a 0.1 mm filtered 300 mM Na/K phosphate pH 6.9 buffer and the separation was performed at a flow rate of 0.35 ml.min^{-1} at 15°C. Data were collected using a DAWN HELEOS 8+ (8-angle) and Optilab T-rEX refractive index detector (Wyatt Technology Corp., Toulouse France). Results were analyzed using the ASTRA 6.0.2.9 software (Wyatt Technology Corp.).

Construction and structure determination of the H959F DH domain mutant

The H959F DH mutant was amplified by inverse PCR using complementary primers bearing two point mutations shifting the catalytic histidine codon CAC to a phenylalanine codon TTC. Forward (5'-GCTGTGGCTCGCCGATTTCGTCATCGACGATC-3) and reverse (5'-GATCGTCGATGACGAAATCGGCGAGCCACAGC-3') primers were from Sigma-Aldrich. Amplified DNA was digested using DpnI and transformed into *E. coli* BL21 (DE3). Transformants were picked and the mutation was verified by sequencing the plasmid (GATC Biotech, Konstanz, Germany). The resulting mutant DH fragment was produced, purified, and crystallized in the exact same conditions than the wild-type DH fragment. *Trans*-2-butenoyl-CoA ($C_{4:1}$-CoA) (Sigma-Aldrich) was solubilized in water and added to the drop containing DH mutant crystals to a final concentration of 1 mM for 1 h. Crystals were flash frozen in liquid nitrogen. A dataset was collected on ID29 at the ESRF and processed using the XDS program package[9]. The structure was solved by molecular replacement using the wild-type DH structure as template in PHASER[5]. *Trans*-2-butenoyl-CoA was then added and structure was refined by iterative cycles of manual model building in COOT[6] and refinement using PHENIX refine[10].

Modeling of the assembly of the DH-ER-ACP domains

First, the sixteen N-terminal residues (921-936) of the DH domain were removed since this region coincides with an extended linker that connects the AT domain in the mFAS structure (PDB code: 2VZ8). One PpsC DH protomer was first superimposed onto each of the two protomers constituting the mFAS DH dimer, prior to the positioning of the ER dimer. The DH-ER interface was then optimized using the local high-resolution minimization option in RosettaDock from the Rosettacommons software suite v3.4[11]. The model of the ACP domain of PpsC (2062-2145) was generated using MODELLER with the structure of the ACP domain from DEBS module 2 (PDB code: 2JU1) as input, and taking into account sequence alignment between the two proteins. The resulting pdb file was then uploaded and set as the ligand molecule for molecular docking using PatchDock[12] while the structure of the DH-H959F mutant in complex with *trans*-2-butenoyl-CoA, whose ADP moiety had been removed, was used as the receptor molecule. Default values were used to run docking simulation, to the exception of a 3 Å distance constraint that was set between the side-chain oxygen atom of ACP S2105 and the β-phosphate phosphorus atom of *trans*-2-butenoyl-phosphopantetheinyl. The twenty best results were then streamed to FireDock[13] for local refinement. Only one result still satisfying the imposed constraint was obtained (global binding energy of -22.72 $kcal.mol^{-1}$). Finally, the model of the DH-ER-ACP tripartite assembly was generated by superimposing the DH-ACP docking model with the DH-ER FAS-like model.

Supplementary Table 1. Data collection, phasing, and refinement statistics.

	DH			DH/$C_{4:1}$-CoA	ER
	peak	*inflection*	*Native*		
Data collection					
Wavelength	0.979138	0.979300	0.979500	0.97625	0.97626
Space group	**P**4_32_12	**P**4_32_12	**P**4_32_12	**P**4_32_12	**P**$2_12_12_1$
Cell dimensions					
a, b, c (Å²)	89.4, 89.4, 160.3	89.4, 89.4, 160.3	89.5, 89.5, 159.3	83.9, 83.9, 166.8	59.9, 78.7, 88.8
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)[a]	46.0 - 3.5 (3.71 - 3.50)	46.0 - 3.5 (3.71 - 3.50)	36.4 - 2.7 (2.8 - 2.7)	41.9 - 2.9 (2.98 - 2.90)	42.0 - 1.5 (1.54 - 1.50)
R_{meas} (%)[b]	7.8 (28.5)	10.3 (92.0)	7.3 (116.8)	23.7 (235.0)	3.7 (96.7)
CC* (%)[c]	99.9 (99.8)	99.9 (98.0)	99.9 (91.9)	99.8 (68.8)	99.9 (78.5)
Anomal Corr.	72* (11*)	32* (1)			
I/σ(I)	23.99 (7.78)	21.21 (2.98)	15.84 (1.38)	6.51 (0.43)	18.5 (1.1)
Completeness (%)	99.8 (99.3)	99.6 (97.7)	99.8 (99. 7)	95.0 (96.9)	90.4 (57.8)
Reflections, total	179465 (29567)	183277 (29044)	116065 (15692)	58177 (4464)	168564 (8132)
Reflections, unique	15640 (2534)	15618 (2493)	18423 (1796)	13146 (968)	61343 (3864)
Redundancy	11.5 (11.7)	11.7 (11.7)	6.3 (5.7)	4.4 (4.6)	2.7 (2.1)
Refinement					
Resolution (Å)			36.4 - 2.7	41.9 - 2.9	42.0 - 1.5
R_{work}/R_{free} (%/%)[d]			0.208/0.2 51	0.272/0.318	0.182/0.2 18
No. of atoms			1989	1967	3033
protein			1979	1905	2717
water			5	28	301
ligand			5	9	15
Mean B value (Å²)			90.1	36.0	28.8
protein			90.0	35.9	27.6
water			138.9	47.0	46.7
ligand			81.5	29.4	38.5
Rmsd bond lengths (Å)			0.011	0.021	0.025
Rmsd bond angles (°)			1.65	2.35	2.30

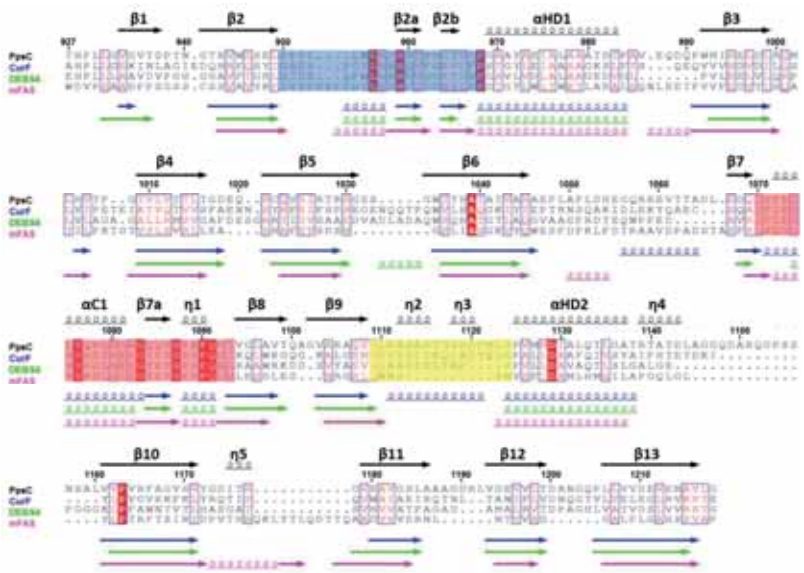
[a] Highest resolution shell is shown in parenthesis

[b.] $Rmeas = (\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{i=1}^{n}|I_i(hkl) - \overline{I}(hkl)|)/\sum_{hkl}\sum_{i=1}^{n}I_i(hkl)$

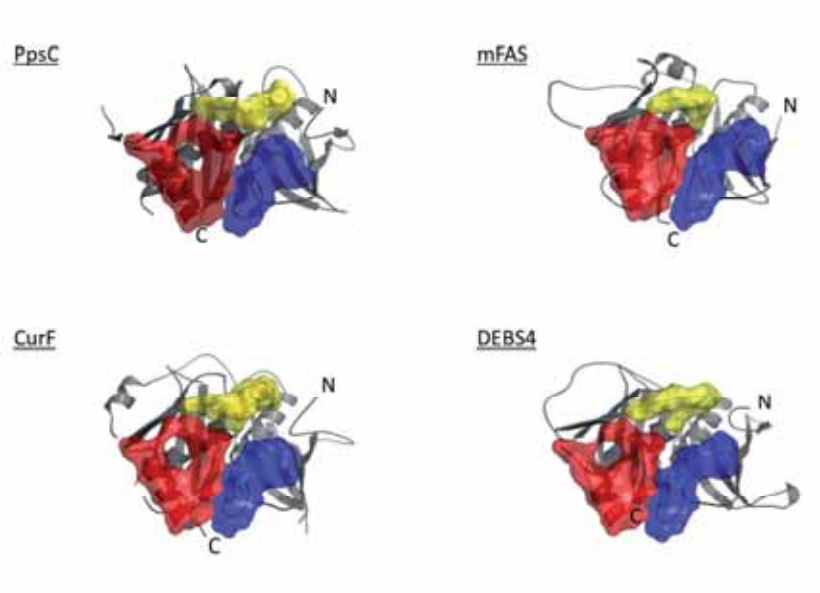
[c.] $CC^* = \sqrt{2CC_{1/2}/(1+CC_{1/2})}$

[d.] $Rwork = \sum_{hkl}(|F_{obs}(hkl)| - |F_{calc}(hkl)|)/\sum_{hkl}|F_{obs}(hkl)|$ and Rfree is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.

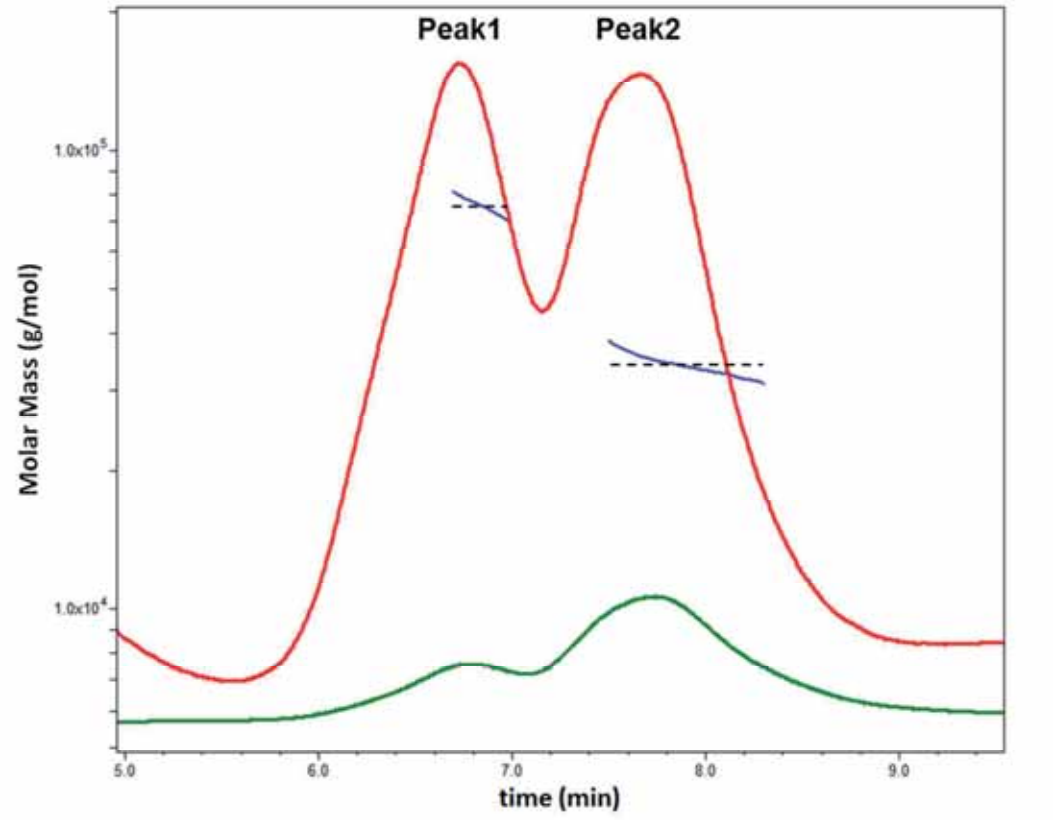
a



b

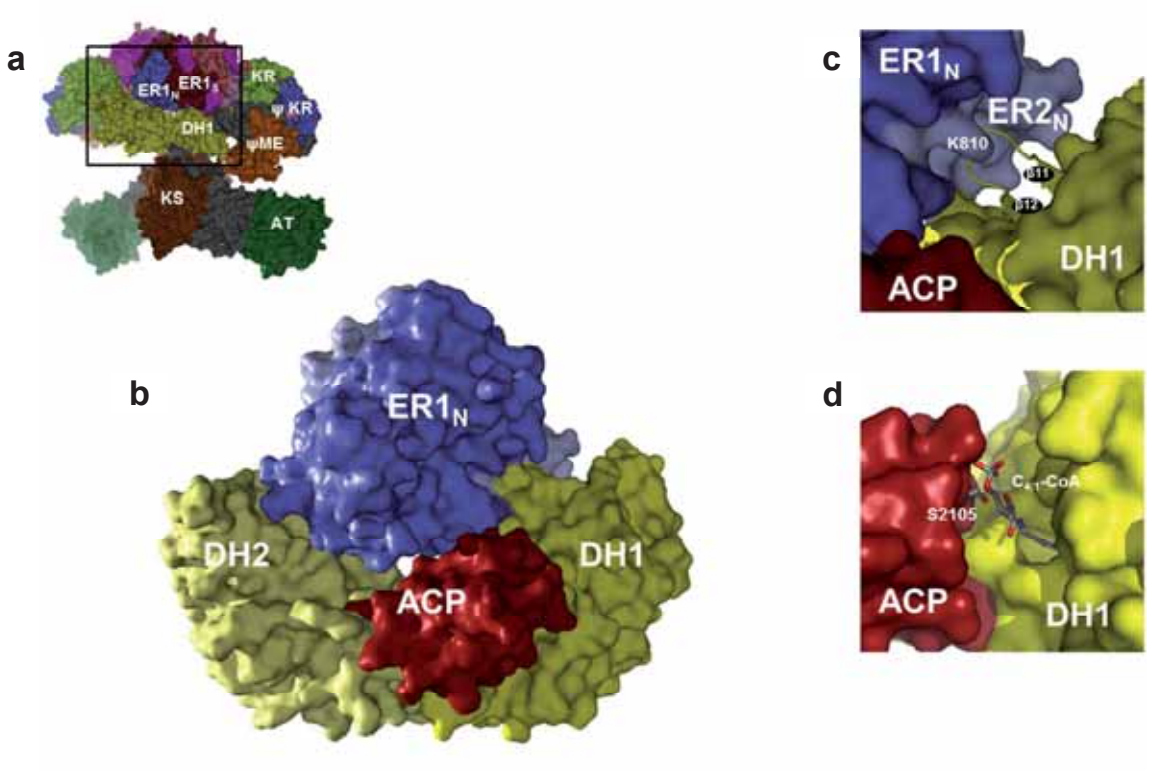


Supplementary Figure 1. Structure-based sequence alignment of DH domains. **(a)** Sequence alignment showing identical residues in blue rectangles with red background and conserved residues in blue rectangles with white background. Secondary structure elements are depicted in black (PpsC), blue (CurF), green (DEBS module 4), and pink (mFAS). Regions with high variability in sequence and/or structure are colored in blue, red and yellow. **(b)** Ribbon representation of PpsC (top left) and its comparison with homologues from CurF (bottom left), DEBS module 4 (bottom right), and mFAS (top right). Regions displayed in blue, red and yellow in the sequence alignment are also represented.



	Total	Peak1	Peak2
Masses			
Injected Mass (µg)	23.00	23.00	23.00
Calculated Mass (µg)	26.12	11.75	2.35
Mass Recovery (%)	113.6	51.1	10.2
Mass Fraction (%)	100.0	45.0	9.0
Molar mass moments (g/mol)			
Mw	$5.059 \cdot 10^4$ (±3.580%)	$3.430 \cdot 10^4$ (±2.568%)	$7.531 \cdot 10^4$ (±1.588%)
Refractive index peak parameters			
Retention Time (min)	7.741	7.741	6.776

Supplementary Figure 2. SEC-MALLS profile of the PpsC DH domain. Displayed are the light scattering signal (red trace) and the differential refractive index signal (green trace) against elution time. The blue trace represents the molar mass ($g.mol^{-1}$) calculated across the elution peak according to SLS measurement whereas the average molar mass (7.53×10^4 $g.mol^{-1}$ and 3.43×10^4 $g.mol^{-1}$) is indicated by the dashed black line for the dimer and monomer. respectively.



Supplementary Figure 3. Assembly model of the PpsC DH, ER and ACP domains. (a) Surface representation of the X-ray structure of mammalian FAS color-coded according to the different domains. The nucleotide-binding (ER_N) and substrate-binding (ERs) subdomains of the ER dimer are also represented with different colors. The navigating ACP domain is missing from the 3D structure. **(b)** mFAS-like assembly model of the DH, ER and ACP domains of PpsC and close-up view of the **(c)** DH/ER and **(d)** DH/ACP interfaces. The ACP domain from the module 2 of DEBS was used for generating the *in silico* docking model of the DH-ACP interactions.

References

(1) Guerrero, S. a., Hecht, H.-J., Hofmann, B., Biebl, H., and Singh, M. (2001) Production of selenomethionine-labelled proteins using simplified culture conditions and generally applicable host/vector systems. *Appl. Microbiol. Biotechnol. 56*, 718–723.

(2) Leslie, A. G. W., and Powell, H. R. (2007) Processing diffraction data with mosflm, in *Evolving Methods for Macromolecular Crystallography*, pp 41–51.

(3) Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D. Biol. Crystallogr. 67*, 282–92.

(4) Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D. Biol. Crystallogr. 66*, 479–85.

(5) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658–674.

(6) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr. 66*, 486–501.

(7) Murshudov, G. N., Vagin, a a, and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr. 53*, 240–55.

(8) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. a, Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr. 67*, 235–42.

(9) Kabsch, W. (2010) XDS. *Acta Crystallogr. D. Biol. Crystallogr. D66*, 125–132.

(10) Adams, P. D., Afonine, P. V, Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr. 66*, 213–21.

(11) Leaver-fay, A., Tyka, M., Lewis, S. M., Lange, F., Thompson, J., Jacak, R., Kaufmann, K. W., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y. A., Fleishman, S. J., Corn, E., Kim, D. E., Lyskov, S., Berrondo, M., Havranek, J. J., Mentzer, S., Popovic, Z., Meiler, J., Kortemme, T., Karanicolas, J., Das, R., Gray, J. J., Kuhlman, B., Baker, D., and Bradley, P. (2011) R OSETTA 3 : An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods Enzymol. 487*, 545–574.

(12) Schneidman-duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005) PatchDock and SymmDock : servers for rigid and symmetric docking. *Nucleic Acids Res. 33*, 363–367.

(13) Mashiach, E., Schneidman-duhovny, D., Andrusier, N., Nussinov, R., and Wolfson, H. J. (2008) FireDock : a web server for fast interaction refinement in molecular docking y. *Nucleic Acids Res. 36*, 229–232.

In this study, we have deciphered the catalytic mechanism of the DH domain of PpsC and we have identified key residues for the substrate specificity of dehydratases. Using two different substrate analogues, we have demonstrated that the chain length is not a determinant in terms of substrate specificity. The approach we used to obtain a complex between the DH domain and its substrate analogue can easily be transposed to any dehydratase, both stand-alone enzymes and domain from a megasynthase. We have also established that the DH dimer is flexible and could adopt a "V-shaped" conformation similar to the one seen in the mFAS X-ray structure. Indeed, considering that the ER dimer superimposes well with its counterpart in mFAS where it plays a major role in stabilizing the bent DH dimer, we have modeled the assembly of the DH and ER domains based on the mFAS structure. This modeling led to the identification of undocumented flexible loops from the DH domain which lies into a cavity in the ER domain and could help form a stable complex between these two domains. Finally, the presence of the bound $C_{4:1}$-CoA in the active site cavity of the DH domain has greatly facilitated the docking of the ACP domain onto the DH domain. Taken together these results should open the route to a more detailed and comprehensive study of the dehydration mechanism, and of the assembly of the β-carbon processing domains in type-I mono-modular PKSs.

We have recently crystallized the DH domain in a different crystalline form which, based on preliminary diffraction tests, will allow us to improve resolution. Soaking these crystals into a crotonyl-CoA solution will certainly bring a clearer view of the interactions than what we are able to observe with the complex structure refined at 3.2 Å resolution we had so far.

B. PROJECT II: STUDY OF AN ESSENTIAL PPTASE FROM MYCOBACTERIUM TUBERCULOSIS

1. PROJECT

PptT has been clearly identified as a new target for the development of antituberculosis drugs by our collaborators at the IPBS [103], [104]. PptT is essential for the replication and survival of *Mtb* during acute and especially chronic phases of infection in mice, a very encouraging feature for the search for new drugs. Indeed, anti-TB drugs usually have only poor bactericidal activity during the chronic phase of infection [184], which extend the treatment course and thus indirectly promote antibiotic resistance. Moreover, a conserved PptT-encoding gene has been identified in all sequenced mycobacterial species, including clinical isolates. Also, PptT may be druggable, since submicromolar inhibitors have already been found for its closest homologue, the *B. subtilis* Sfp [185]. Besides, the low amino acid identity (12%) with the human AcpS (hAcpS) will certainly allow the design of specific inhibitors for PptT, and the availability of the hAcpS structure should enable careful selection of hit compounds.

With this in mind, our collaborators have developed a miniaturized assay suitable for the search of inhibitors of PptT by high-throughput screening [104]. In

order to help the rational drug design, we have decided to determine the tridimensional structure of PptT using X-ray crystallography, in parallel to its functional characterization. We hope the structure will help understand the binding mechanism of inhibitors and will contribute in a near future to the design of new anti-TB drugs. Finally, PptT also represents an invaluable biotechnological tool for the activation of mycobacterial PKSs and may help their structural and functional study as well.

When PptT is produced in *E. coli* using classical pET vectors, only very small amounts can be purified and the protein precipitates irreversibly overtime. To overcome this problem, our collaborators, on the basis of published results [153], have managed to produce and purify PptT N-terminally fused to the maltose binding protein (MBP). However, this construction may not be adapted to a structural study since flexibility between the two proteins would probably hinder crystallization. Furthermore, removing the MBP tag induces precipitation of PptT. In an effort to produce PptT expressed alone, we have first explored for a better production strategy that led us to develop a new method for the diagnosis of co-factor dependent protein expression. Using this strategy, sufficient amounts of stable PptT could be obtained. PptT has then been subjected to crystallization trials and characterized using biophysical methods. Finally, its 4'-phosphopantetheinyl transferase activity has been tested with PpsC ACP domain as its carrier protein substrate.

2. DESIGN OF A SPECIFIC PROTOCOL TO RECOVER PPTT

Based on length, activity, and sequence motifs, PptT was thought to belong to the group II PPTase class of enzymes, along with hAcpS and the well characterized Sfp from *B. subtilis*. According to the biochemical and structural data accumulated for Sfp [88], [97] and hAcpS [89], [92], this group of PPTase would use Mg^{2+} and CoA as cofactors to realize its catalytic activity.

Before considering producing PptT in other organisms than *E. coli*, we have sought to determine whether PptT was produced in a soluble form *in-vivo* using the split-GFP technology. We have also generated N- and C-terminally truncated versions of the open reading frame in an effort to obtain a soluble and stable fragment of the enzyme. Interestingly, the most soluble construction in *E. coli* was the full-length enzyme, although once purified, aggregation occurred overtime *in vitro*. Our knowledge on the catalytic mechanism of group II PPTases have then led us to introduce the naturally-occurring co-factors, since they might be required to stabilize PptT, even though this feature has never been encountered in any functional homologue so far [88], [89].

Introducing $MgCl_2$ and CoA all along the purification process, as well as using a weaker promoter, have indeed been sufficient to recover large amounts of PptT in a soluble and non-aggregated form. The impact of co-factors on PptT stability has further been estimated using differential scanning fluorimetry (DSF). We have also performed enzymatic tests to verify the *in vitro* PptT activity and determine its metal-ion specificity, some homologues being able to catalyze the transfer reaction with other divalent cations than Mg^{2+}, like Ca^{2+} [90], or Mn^{2+} [91]. To do so, we have monitored the transfer of P-pant onto a fragment containing PpsC ACP domain isolated by *domain trapping*, by both PptT and Sfp.

All these results, along with the methodology developed to diagnosis protein requirement for co-factors, are presented in the following paper:

A. Faille*, K. Rottier*, T. Prudhomme, C. Leblanc, C. Chalut, S. Cabantous, C. Guilhot, L. Mourey, and J.-D. Pedelacq, "Detection of soluble co-factor dependent protein expression *in vivo*: application to the 4'-phosphopantetheinyl transferase PptT from *Mycobacterium tuberculosis*." *Journal of Structural Biology,* , vol. 183, no. 3, pp 320-8, Sept. 2013.

*These authors contributed equally to this work.

Journal of Structural Biology 183 (2013) 320–328

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Detection of soluble co-factor dependent protein expression *in vivo*: Application to the 4′-phosphopantetheinyl transferase PptT from *Mycobacterium tuberculosis*

Karine Rottier [a,b,1], Alexandre Faille [a,b,1], Thomas Prudhomme [a,b], Cécile Leblanc [a,b], Christian Chalut [a,b], Stéphanie Cabantous [c,d,e], Christophe Guilhot [a,b], Lionel Mourey [a,b], Jean-Denis Pedelacq [a,b,*]

[a] *CNRS, IPBS (Institut de Pharmacologie et de Biologie Structurale), 205 Route de Narbonne, BP 64182, F-31077 Toulouse, France*
[b] *Université de Toulouse, UPS, IPBS, F-31077 Toulouse, France*
[c] *INSERM UMR 1037, Cancer Research Center of Toulouse, 20–24 Rue du Pont St. Pierre, 31052 Toulouse Cedex, France*
[d] *Université de Toulouse, 31052 Toulouse Cedex, France*
[e] *Institut Claudius Regaud, 31052 Toulouse Cedex, France*

ARTICLE INFO

Article history:
Received 13 May 2013
Received in revised form 24 July 2013
Accepted 25 July 2013
Available online 31 July 2013

Keywords:
Split-GFP
Domain trapping
Co-factor
4′-Phosphopantetheinyl transferase
Polyketide
Tuberculosis

ABSTRACT

The need for early-on diagnostic tools to assess the folding and solubility of expressed protein constructs *in vivo* is of great interest when dealing with recalcitrant proteins. In this paper, we took advantage of the picomolar sensitivity of the bipartite GFP1–10/GFP11 system to investigate the solubility of the *Mycobacterium tuberculosis* 4′-phosphopantetheinyl transferase PptT, an enzyme essential for the viability of the tubercle bacillus. *In vivo* and *in vitro* complementation assays clearly showed the improved solubility of the full-length PptT compared to its N- and C-terminally truncated counterparts. However, initial attempts to purify the full-length enzyme overexpressed in *Escherichia coli* cells were hampered by aggregation issues overtime that caused the protein to precipitate within hours. The fact that the naturally occurring Coenzyme A and Mg^{2+}, essentials for PptT to carry out its function, could play a role in stabilizing the enzyme was confirmed using DSF experiments. *In vitro* activity assays were performed using the ACP substrate from the type I polyketide synthase PpsC from *M. tuberculosis*, a 2188 amino-acid enzyme that plays a major role in the virulence and pathogenicity of this microbial pathogen. We selected the most soluble and compact ACP fragment (2042–2188), identified by genetic selection of in-frame fragments from random library experiments, to monitor the transfer of the P-pant moiety from Coenzyme A onto a conserved serine residue of this ACP domain.

1. Introduction

Ways to reduce the toxicity and improve the solubility of recombinant proteins in heterologous hosts have been the subject of prolific research over the past two decades. Major progresses have been made with the advent of structural genomics initiatives and the development of tools to circumvent this major obstacle. Available strains and expression vectors with different promoters now make possible parallel approaches to increase the chance of isolating protein targets in a form suitable for functional and structural studies. Oftentimes, recombinant proteins partition into "inclusion bodies" as a result of misfolding, aggregation, and intracellular accumulation. With the advent of DNA polymerase-based error-prone PCR protocols (Cadwell and Joyce, 1994) as well as *Escherichia coli* mutator strains (Greener et al., 1997), introducing mutations that enhance solubility and stability of a target protein while retaining its activity has now become possible. Wider sequence variations introduced by DNA shuffling techniques (Stemmer, 1994; Zhao et al., 1998) have expanded the functional diversity and facilitated the identification of variants with superior crystallization propensity (Keenan et al., 2005). Libraries of se-

Abbreviations: ACP, acyl carrier protein; AcpS, acyl carrier protein synthase; AnTet, anhydrotetracycline; BSA, bovine serum albumin; CoA, Coenzyme A; DHFR, dihydrofolate reductase; DSF, differential scanning fluorimetry; DTT, dithiothreitol; GFP, green fluorescent protein; IMAC, immobilized metal ion affinity chromatography; IPTG, isopropyl β-D-1-thiogalactopyranoside; Kan, kanamycin; LB, Luria–Bertani; MBP, maltose binding protein; PCR, polymerase chain reaction; PDIM, phtiocerol dimycocerosate; PKS, polyketide synthase; P-pant, 4′-phosphopantetheine; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; SEC, size-exclusion chromatography; Sfp, surfactin synthetase-activating enzyme; Spec, spectinomycin; SUMO, small ubiquitin-related modifier.

* Corresponding author at: Université de Toulouse, UPS, IPBS, F-31077 Toulouse, France.
E-mail address: Jean-Denis.Pedelacq@ipbs.fr (J.-D. Pedelacq).

[1] These authors contributed equally to this work.

1047-8477/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2013.07.010

quences fused to reporters with specific marker phenotype (fluorescence in the case of GFP reporters) may be screened to detect soluble variants. Successful results were obtained with the GFP (Pedelacq et al., 2002; Waldo, 2003), although the bulky C-terminally fused reporter may affect the solubility of the passenger protein. Insertion into the GFP scaffolding or fusion to GFP11 prior to complementation with the large fragment GFP1–10 (Cabantous et al., 2005) can overcome this limitation.

When a full-length protein fails to maintain in a soluble form, N- and C-terminal truncations generated by PCR amplification can help identifying variants with increased solubility level (Pedelacq et al., 2011). Another alternative is to select soluble fragments encompassing one or several domains. This can be achieved when domain boundaries are easily predictable from known homologues. The situation becomes more difficult with protein targets that have no sequence or structural homologues. One way to tackle this problem is to generate libraries of truncated DNA fragments coupled with a genetic screen to discriminate between the thousands of constructs. The expression of soluble proteins by random incremental truncation (ESPRIT) (Yumerefendi et al., 2010) and colony filtration (CoFi) blot (Cornvik and Dahlroth, 2006) methods have their limits, since they only permit to truncate unidirectionally one end of the gene. The bidirectional truncation version of ESPRIT (An et al., 2011) and the combinatorial domain hunting (CDH) (Reich et al., 2006) method have no filtering strategy over fragment orientation and frame selection. Upstream protein fusions to murine dihydrofolate reductase (mDHFR) (Dyson et al., 2008) or the β-lactamase (Fisher et al., 2006) can help eliminating incorrect reading frames. However, false positives were identified from translation initiation at internal ribosome binding sites (IRBS). A number of bipartite selection systems have also been developed in an attempt to overcome this issue. The DNA sequence of the target gene is inserted between the two halves of the reporter, which are both required to give an observable phenotype (Cabantous et al., 2008; Daugelat and Jacobs, 1999; Gerth et al., 2004). Our insertional DHFR system combined with the split-GFP assay has been successfully applied to identify soluble fragments from each domain of human p85α and PpsC (Pedelacq et al., 2011).

In this paper, we adopted a PCR approach and a library screening strategy to identify soluble constructs of an enzyme and its protein substrate, respectively. We extend the use of the split-GFP complementation assay (Cabantous et al., 2005) as a diagnostic tool to investigate the solubility of co-factor dependent proteins expressed *in vivo*. We applied this system to the *Mycobacterium tuberculosis* PptT, an enzyme required for growth and persistence of the bacteria *in vivo* (Leblanc et al., 2012). PptT is responsible for the covalent transfer of the P-pant group of Coenzyme A (CoA) to a conserved serine residue onto the acyl carrier protein (ACP) domain of various type I polyketide synthases (PKS) (Quadri et al., 1998a). This post-translational modification converts PKS into functional *holo* enzymes that play a key role in the biosynthesis of various lipids of the mycobacterial cell envelope (Chalut et al., 2006). Split-GFP complementation assays indicated that the full-length enzyme is the most soluble among all the variants tested. A soluble fragment containing the predicted boundaries for the PpsC ACP substrate identified using our domain trapping method (Pedelacq et al., 2011) was used for *in vitro* activity assays. We showed that the natural endogenous CoA and divalent metal ion Mg^{2+}, which play a crucial role for PptT to carry out its function, are also a prerequisite for its stability overtime. This easy-to-use split-GFP system can be generalized to detect soluble co-factor dependent protein expression *in vivo* whenever the presence of co-factors is required for folding and stability.

2. Materials and methods

2.1. PptT small scale expression and solubility tests

The *pptT* gene from *M. tuberculosis* was amplified from genomic DNA using top and bottom primers listed in Supplementary Table 1. The *Nde*I/*Bam*HI digested *pptT* was then inserted into commercially available pET28 vectors (Novagen, Madison, WI, USA) with N- or C-terminal 6His tag. pET vectors allowing the production of PptT with MBP (Leblanc et al., 2012) or SUMO (Invitrogen, Carlsbad, CA, USA) fused to its N-terminus were also tested. Ligated plasmids were transformed into chemically competent *E. coli* BL21(DE3) cells (Invitrogen, Carlsbad, CA, USA). Transformed cells were plated onto Luria–Bertani (LB) agar plates containing 35 μg/ml kanamycin, allowing them to grow overnight at 32 °C. The resulting clones were grown at 37 °C in 1 ml cultures using 35 μg/ml kananycin. Cells were induced in exponential phase with 1 mM IPTG for 3 h. Cell culture pellets of 1 ml of each fragment were separately resuspended in 40 μl TNG buffer (150 mM NaCl, 100 mM Tris–HCl pH = 7.5, 10% (v/v) glycerol) and sonicated. The lysate was fractionated by centrifugation to yield the soluble and pellet fractions. The pellet fraction was washed twice with 100 μl TNG buffer, centrifuged and resuspended in the same starting volume. Samples corresponding to the soluble (S) and pellet (P) fractions were resolved on a 4–20% gradient Criterion SDS–PAGE gel (Bio-Rad, Hercules, CA, USA). Protein samples were stained using Gel Code Blue stain reagent (Pierce, Rockford, IL, USA) and imaged using a GS-800 Calibrated Densitometer (Biorad, Hercules, CA, USA).

2.2. Cloning and in vivo split-GFP solubility screen

Full-length and truncated variants of the *M. tuberculosis PptT* gene were amplified by PCR using a series of forward and reverse primers listed in Supplementary Table 1. Cleaned inserts from *Nde*I/*Spe*I restriction digests were ligated into the pTET ColE1 GFP11 vector and transformed into chemically competent BL21(DE3) cells containing the pET GFP1–10 plasmid (Cabantous and Waldo, 2006; Cabantous et al., 2005). Frozen cells were used to grow 1 mlLuria–Bertani (LB) medium in a 96-well deep-well plate containing 35 μg/ml kanamycin and 112 μg/ml spectinomycin until OD_{600} reached ~1.0 at 37 °C. 50 μl of two successive 400-fold dilutions of the expressed constructs were plated onto a compartmentalized nitrocellulose membrane and grew overnight at 32 °C. The membrane was transferred to an LB/Agar plate containing the same antibiotics supplemented with 250 ng/ml anhydrotetracycline (AnTET) for 2 h at 32 °C, and then moved back onto the original plate for 1 h at the same temperature. Finally, the membrane was transferred onto a plate with same antibiotics and 1 mM IPTG to induce the expression of the GFP1–10 fragment at 32 °C. Colonies were illuminated using an Illumatool Bright light System LT-9900 (http://www.lightools.com/) at excitation and emission wavelengths of 470 nm and 515 nm, respectively. Pictures were taken after 30 min complementation at 32 °C.

2.3. In vitro split-GFP complementation assay

5 ml LB cultures with 112 μg/ml spectinomycin were grown at 37 °C until OD_{600} reached 0.5–0.7. After induction at 32 °C with AnTET at a final concentration of 8 μg/ml and growth for an additional 4 h, cells were harvested by centrifugation at 20,000*g* for 20 min. 20 μl of PptT-GFP11 soluble fractions were mixed with 180 μl of 0.35 mg/ml refolded GFP1–10 in a 96 well microplate (Nunc-Immuno plate, Nunc, Rochester, NY, USA), as previously described (Cabantous and Waldo, 2006). Fluorescence kinetics (λ_{exc} = 488 nm/λ_{em} = 530 nm) were monitored with a

FL600 Microplate Fluorescence Reader (Bio-Tek, Winooski, VT, USA), at 3 min intervals, for 15 h. The background fluorescence of a blank sample (20 μl *E. coli* lysate expressing an unrelated protein without the GFP11 tag, and 180 μl of 0.35 mg/ml GFP1–10 in TNG buffer) was subtracted from final fluorescence values.

2.4. Full-length PptT production and purification

Frozen cells expressing the full-length PptT from a pTet vector missing the GFP11 tag were used to start an overnight 3 ml LB-spectinomycin (112 μg/ml) culture at 32 °C prior to inoculation in baffled flasks containing 500 ml of the same media. Cells were allowed to grow for approximately 2 h at 37 °C before temperature was dropped from 37 °C to 20 °C. When OD_{600} reached 0.5–0.7, cells were induced with AnTET at a final concentration of 8 μg/ml and grown for an additional 18 h prior to harvesting by centrifugation at 4000*g* for 30 min and storage at −80 °C. The pelleted cells were suspended in 40 ml of buffer A (100 mM Tris pH 8.0, 300 mM NaCl, 10 mM $MgCl_2$, 50 μM CoA) and lysed by sonication (6 cycles of 30 s pulse, 50% amplitude, power 5) prior to centrifugation at 20,000*g* for 1 h. The supernatant was filtered (0.2 μm) and purified on an ÄKTA purifier system (GE Healthcare) following a two-step procedure that combined an immobilized metal ion affinity chromatography (IMAC) and a size-exclusion chromatography (SEC). The lysate was loaded onto a 1 ml HisTrap HP (GE Healthcare) affinity column. The N-terminally His tagged PptT was eluted from the column with a step of buffer A supplemented with 120 mM imidazole. The eluted protein fractions were pooled and concentrated down to 2 ml using a Vivaspin 20 centrifugal concentrator (Sartorius), prior to injection into a HiLoad 16/60 Superdex 200 (GE Healthcare) pre-equilibrated with buffer B (50 mM Tris pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 50 μM CoA).

2.5. ACP fragment cloning and purification

The *ppsC* gene from *M. tuberculosis* was cloned into the *Nde*I/*Spe*I sites of a pET26b plasmid (Novagen, Madison, WI) and PCR amplified using gene specific forward 5′-GATATA**CATATG**ACCGCAGCGAC ACCAGATCG-3′ and reverse 5′-AATTC**ACTAGT**TGACTCGCCTCGCGT CGCAGC-3′ primers. DNA fragmentation, cloning of the DNA library into the DHFR ORF filter and the split-GFP solubility reporter (Cabantous et al., 2005) were conducted as previously described (Pedelacq et al., 2011). Briefly, a 400 to 850 bp DNA library was created by mechanical shearing of the PCR amplified *ppsC* gene. Blunt fragments were ligated in a StuI-digested insertion DHFR pET vector to select for in-frame fragment in the presence of 6 μg/ml trimethoprim. Inverse PCR was performed with fragments ligated into the pTET ColE1 GFP11 vector using phosphorylated forward (5′-CGACTCGCTGATGGGCCTGGAATTGCGCAATC-3′) and reverse (5′-AGTCCCAGGGTTTCCAGCGGTCGGTGGTGATC-3′) primers. After transformation into chemically competent BL21(DE3) pET GFP1–10 cells, 96 individually picked clones were grown overnight at 30 °C in a 96-well tissue culture plate containing 7.5% glycerol in LB medium with kanamycin and spectinomycin prior to sequencing. DNA sequences were analyzed using the BioEdit software (http://www.mbio.ncsu.edu/bioedit/bioedit.html) and aligned onto the full-length parent gene to determine the exact boundaries. Based on the *in vitro* solubility assays (Cabantous et al., 2005), fragments were color-coded black, dark gray, and light gray, where the black side of the spectrum identifies the least soluble ACP fragments and the light gray side corresponds to the top 25% most soluble ones. A DNA fragment encompassing the ACP domain and linker regions of PpsC was amplified by conventional PCR and was cloned into the *Nde*I/*Bam*HI restriction sites of a pET28a plasmid (Novagen, Madison, WI, USA) using GATATA**CATATG**CATG ACTCGGCGGCCCGCAAAA and AATTC**GGATCC**TGACTCGCCTCGCGTC GCAGCTT as forward and reverse primers, respectively. The resulting plasmid was transformed into chemically competent BL21(DE3) *E. coli* cells. Cells expressing the ACP fragment were grown to $OD_{600} \sim 0.6$, induced with 0.25 mM IPTG for 4 h at 30 °C, pelleted by centrifugation at 4000*g* for 30 min and stored at −80 °C. The frozen pelleted cells were resuspended in 40 ml of buffer A and lysed by sonication (6 cycles of 30 s pulse, 50% amplitude, power 5) prior to centrifugation at 45,000*g* for 30 min. 4 ml of Talon Superflow beads (Talon resin, Clontech, Palo Alto, CA, USA) conditioned in buffer A were added to the supernatant for 5 min prior to centrifugation at 1000*g* for 5 min. The unbound fraction was removed and the beads were washed twice with 50 ml of buffer A then 30 ml of buffer A supplemented with 10 mM imidazole to eliminate non-specific binding to the beads. The N-terminally His-tagged protein was eluted with 5 ml of buffer A containing 250 mM imidazole. The protein elution sample was resolved on a 4–20% gradient SDS–PAGE gel and stained using a Coomassie Blue R-250 staining solution. The protein solution was stored at −20 °C after dialysis overnight against 25 mM Tris pH 8.0, 25 mM NaCl with 10% glycerol.

2.6. DSF experiments

Fluorescence of the SYPRO Orange Protein Gel Stain (Invitrogen, USA) wasmeasured using the CFX96 real-time PCR detection System (Bio-Rad, Hercules, CA, USA) from 20 °C to 90 °C in 0.3 °C increments of 3 s. PptT was concentrated to 400 μM in the presence of buffer A (50 mM Tris pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 50 μM CoA) prior to a 10-fold dilution with either buffer A or buffer B (50 mM Tris pH 8.0, 50 mM NaCl). 20 μl mixtures containing 2 μl PptT 40 μM and 1 μl Sypro Orange 200X were placed in 96-well clear-bottom Hard-Shell 96-Well Skirted PCR Plates, Low-Profile (BioRad). Addition of 17 μl of buffer A or B completes the reaction mixture. Each condition was realized in triplicate. Melting temperatures were then calculated using the BioRad CFX Manager software.

2.7. Activity assays

PptT at 50 nM was incubated at 30 °C with 10 μM of ACP domain of PpsC. 20 μM CoA, 25 mM DTT and divalent cations ($CaCl_2$, $MnCl_2$, or $MgCl_2$) at 10 mM final concentration were also present in the reaction mixture. Reaction was then stopped at different times with the addition of 100 mM EDTA. The same experiment was conducted with Sfp from *Bacillus subtilis* (NEB, Ipswich, MA) at the same concentration, instead of PptT. Samples were then loaded on a 10% polyacrylamide gel supplemented with 2.5 M urea. Appropriate positive and negative controls for the transfer of the P-pant moiety were carried out for one hour at 30 °C in the presence or absence of 200 nM Sfp, respectively.

3. Results

3.1. Expression of PptT in bacterial pET vectors

We first cloned *pptT* into pET28 derived vectors prior to transformation into BL21(DE3) cells. Only in a situation where the histidine tag is present at the N-terminal extremity of the protein that overexpression of PptT can be achieved. The construct is poorly soluble and expresses mainly as inclusion bodies after IPTG induction (Fig. 1). To maximize the amount of soluble protein, one strategy was to screen conditions where the inclusion bodies could be refolded in a stable and active conformation. This was achieved in the presence of 100 mM MES at pH 5.5. Unfortunately, analytical

gel filtration experiments indicated an aggregated form of the enzyme detected in the exclusion volume of the column and the activity of the enzyme could not be confirmed.

As an alternative, we cloned *pptT* into pET vectors with fusion partners maltose binding protein (MBP) (Fox and Waugh, 2003) and small ubiquitin-related modifier (SUMO) (Mossessova and Lima, 2000), previously described as solubility enhancers of passenger proteins (see Section 2). On the one hand, SUMO had no effect on the solubility of the fusion. On the other hand, the MBP–PptT fusion was mostly soluble and could be purified easily following a two-step procedure using metal-affinity resin followed by size exclusion chromatography. Unfortunately, the enzyme irreversibly precipitates after proteolytic cleavage of the N-terminal MBP tag, which indicates that the MBP fusion maintains PptT in a soluble form.

3.2. Split-GFP solubility screen of PptT variants

Single amino-acid truncation at the N- or C-terminal extremity of a protein of interest often has a dramatic effect onto the solubility level of the expressed variant (Listwan et al., 2010; Pedelacq et al., 2011). With this in mind, we explored the expression and solubility levels of an ensemble of 48 constructs *in vivo* (Fig. 2a and b), corresponding to the full-length PptT along with N- and C-terminally truncated variants, using the split-GFP based assay (Cabantous et al., 2005). In this system, two separately inducible expression systems are present in the same *E. coli* cell, one tetracycline-inducible vector bearing the small GFP11 tag N-terminally fused to the protein of interest while the large complementary fragment GFP1–10 is located onto a separate IPTG-inducible expression vector. After sequential induction, only soluble GFP11 fusions can spontaneously bind to GFP1–10, thus restoring the fluorescence of the full-length GFP. In this system, fluorescent complementation is directly proportional to the amount of soluble protein present in the living cells or purified from cell crude extracts, with detection levels within the picomole range. *In vivo*, the highest fluorescence levels could be attained for the full-length enzyme and N-terminally truncated variants whereas truncation of the last C-terminal residue L227 was sufficient to negatively impact the fluorescence intensity of the colonies (Fig. 2b). To confirm this result, an *in vitro* split-GFP complementation kinetic assay was performed by adding a large excess of GFP1–10 to the N- and C-terminally truncated variants in the soluble fraction. After 15 h of complementation at room temperature, fluorescence intensity measured for the Δ227 deletion mutant was reduced by ~75% compared to the full-length enzyme (Fig. 2c). Fluorescence intensity levels measured for the other six C-terminally truncated variants, from position 226 to position 220, remained low and comparable.

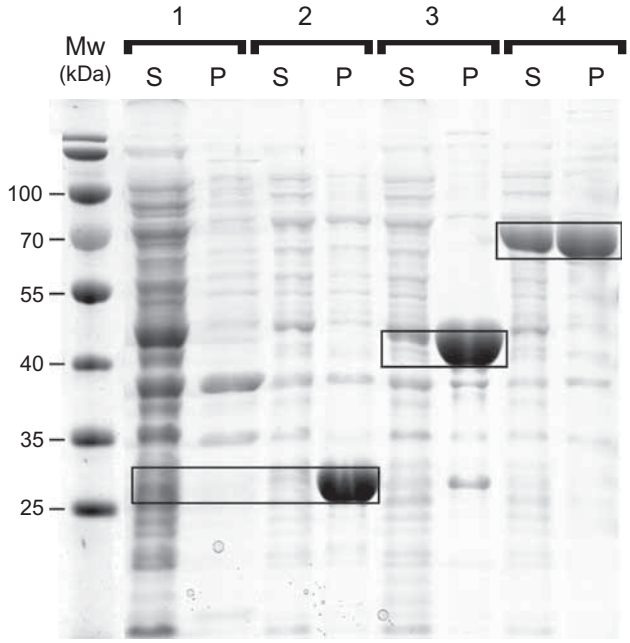


Fig.1. SDS–PAGE of soluble (S) and pellet (P) fractions of PptT constructs of *E. coli* BL21(DE3) cell cultures carrying a pET28-PptT plasmid allowing expression of PptT (1) with C-terminal and (2) N-terminal 6His tag, (3) with N-terminally fused SUMO and (4) MBP proteins. Molecular weight markers (Mw) are also indicated.

3.3. Both CoA and $MgCl_2$ are essential for the stability of PptT

In vivo expression and solubility levels of the recombinant His–PptT–GFP11 fusion were assessed by comparing the fluorescence intensities of co-induced and sequentially induced *E. coli* cell cultures, as previously described (Cabantous and Waldo, 2006). Approximately one third of the enzyme could be expressed in a soluble form, a result in accordance with the band intensities measured on a SDS–PAGE gel (data not shown). A two-step purification procedure using immobilized metal ion affinity chromatography (IMAC) and size-exclusion chromatography (SEC) indicated that the protein was mostly aggregated and could precipitate within hours.

Knowing the essentiality of CoA and Mg^{2+} for PptT to carry out its function *in vivo* (Quadri et al., 1998a), we were wondering if these compounds could also help maintaining the enzyme in a stable conformation *in vitro*. When CoA and $MgCl_2$ are present in a large molar excess over PptT in the cell lysate and all along the two-step purification strategy, high yields of soluble recombinant protein can be obtained. We then conducted a series of experiments using the Differential Scanning Fluorimetry (DSF) method (Niesen et al., 2007) to confirm the importance of the two protagonists onto the stability of PptT. We used Sypro Orange to detect the buried hydrophobic regions of the enzyme that become progressively exposed to solvent during a linear temperature ramp from 20 °C to 90 °C. The low initial fluorescence background and the high signal-to-noise fluorescence ratios have facilitated an accurate measurement of the transition midpoint (Tm) values. High Tm values around 50 °C were measured in the presence of CoA and $MgCl_2$ at a final concentration of 50 μM and 10 mM, respectively, in accordance with the purification conditions. Reducing the concentration of CoA to trace amount (0.5 μM final) while maintaining a 25-fold molar excess of $MgCl_2$ (100 μM final) over PptT profoundly impacted the stability of the full-length enzyme, as no Tm values could be measured.

3.4. In vitro enzymatic activity of PptT

PptT catalyzes the covalent transfer of the 4′-phosphopantetheine (P-pant) group from CoA onto the ACP domain of various type I PKS and nonribosomal peptide synthetases (NRPS) (Quadri et al., 1998a). Issues related to heterologous host expression, folding and solubility of these large-size multi-domain enzymes forced us to identify a more compact and stand-alone soluble fragment encompassing the functional ACP domain. We used our recently published domain trapping strategy (Pedelacq et al., 2011) to generate a 400–850 bp fragment library centered onto the ACP domain of PpsC, a 2188 amino-acid enzyme involved in the biosynthesis of phtiocerol dimycocerosates (PDIM) (Camacho et al., 2001). These lipids play a major role in the virulence of *M. tuberculosis* (Astarie-Dequeker et al., 2009). Our approach combines a two-body dihydrofolate reductase (DHFR) scaffolding for selecting in frame DNA sequences from a random library of the fragmented *ppsC*

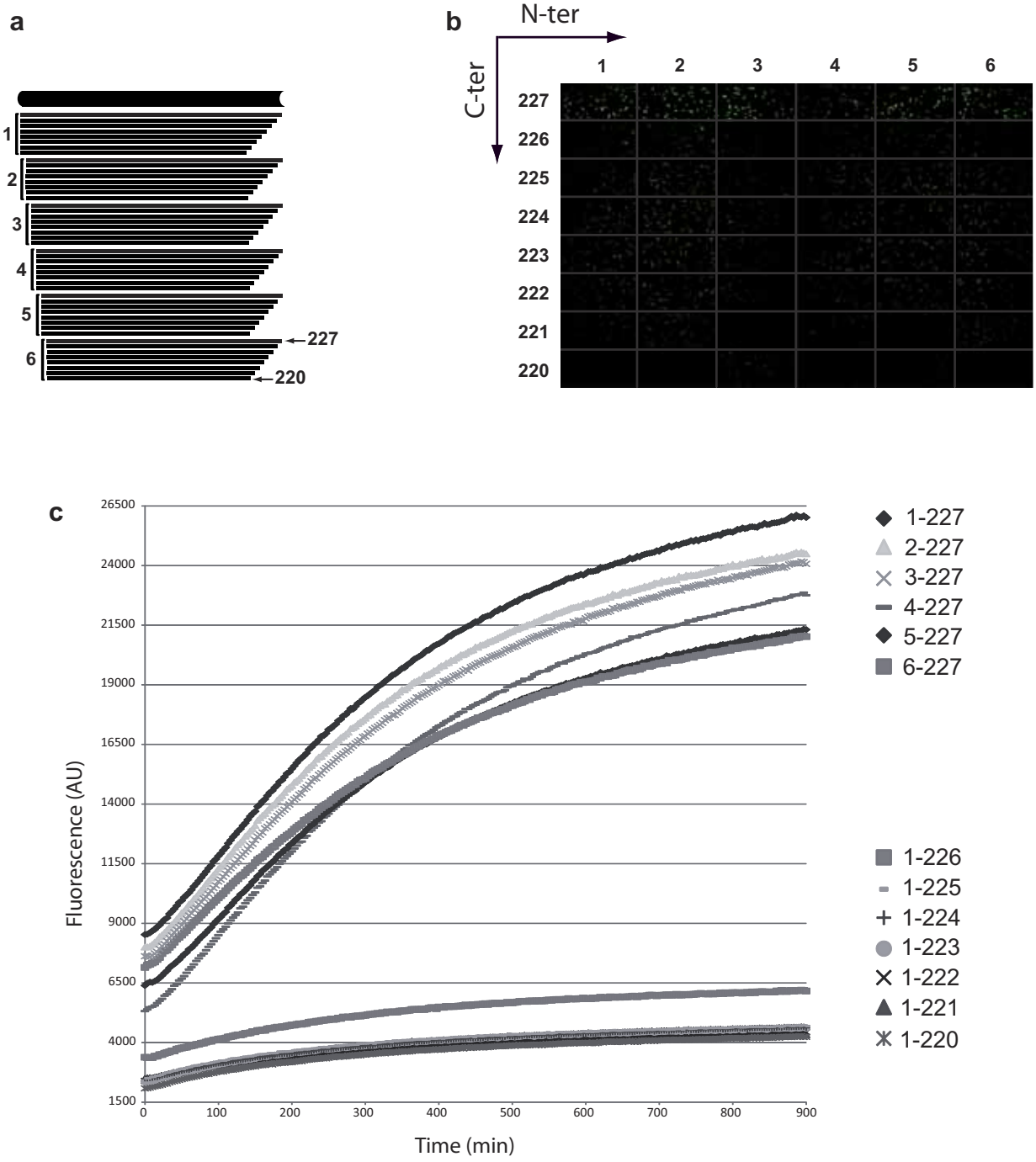


Fig.2. PCR-directed truncations of *pptT* and split-GFP complementation reaction. (a) Schematic representation of the 48 constructs aligned onto the PptT amino-acid sequence. Fragments are organized in groups of 6 with identical N-terminal positions and decreasing C-terminal positions. (b) Solubility screen of *E. coli* BL21(DE3) cells expressing the corresponding fragments in fusion with GFP11 following sequential complementation with GFP1–10. Most soluble fragments correspond to the full-length and N-terminally truncated constructs. (c) Superimposition of progress curves for complementation of PptT variants as a function of the fluorescence intensity versus time in minutes.

gene, with the split-GFP technology (Cabantous et al., 2005) to identify soluble candidates centered onto the ACP domain (Fig. 3). The main advantage of our approach lies in the efficiency of the DHFR step to filter-out ~95% of the constructs present in the original DNA library and the use of an inverse PCR step to select for fragments encompassing a specific region or domain of a given protein (Pedelacq et al., 2011). The implementation of the split-GFP assay to distinguish colonies on the basis of their fluorescence intensities completes this highly reliable and efficient strategy. We picked a total of 96 clones with fluorescence intensity levels varying from bright to faint depending on the solubility of the expressed fragment. Activity assays were carried out in the presence of the most soluble and compact ACP fragment 2042–2188 using either PptT or its closest homologue Sfp from *B. subtilis* (Quadri et al., 1998b). We then visualized differences in the migration of the ACP domain on a 10% native polyacrylamide gel supplemented with 2.5 M urea (Fig. 4). Knowing that the reaction catalyzed by Sfp is magnesium- (Reuter et al., 1999) or manganese-dependent

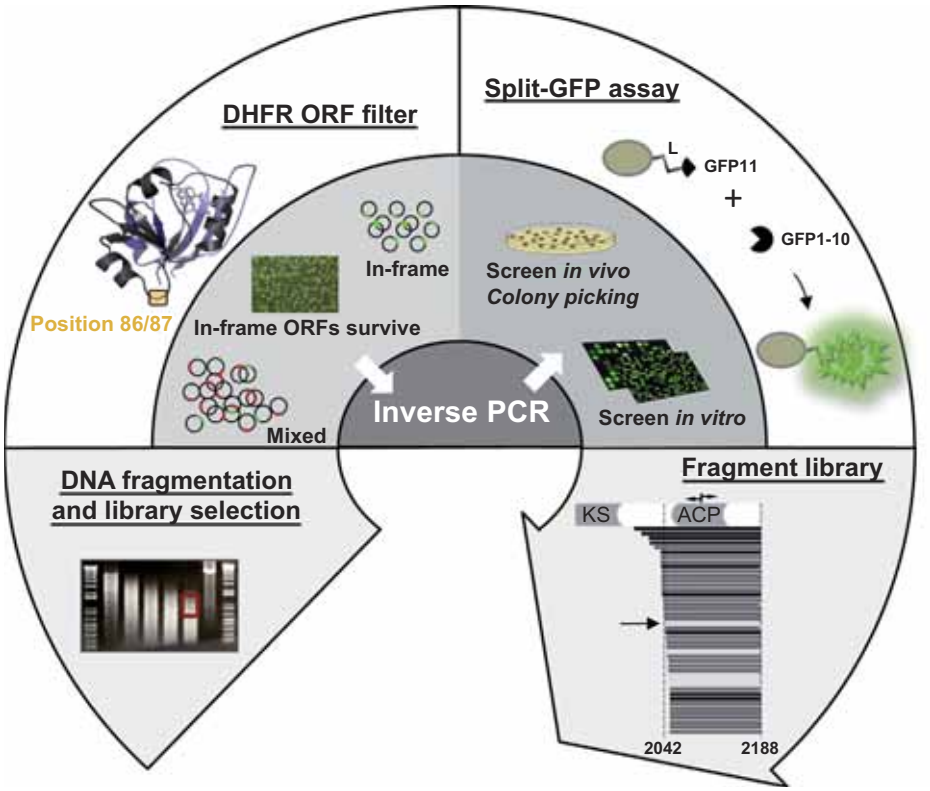


Fig.3. The domain trapping strategy. Our strategy integrates a DHFR filter to eliminate the incorrect reading frames and a split-GFP solubility screen to identify the soluble constructs. 450–850 bp generated library of fragments was inserted at position 86/87 of the DHFR. Inverse PCR focused sub-libraries of fragments helped identifying a soluble fragment (indicated by an arrow) centered onto the ACP domain of the *M. tuberculosis* polyketide synthase PpsC. Clones are picked and grown in 96-well liquid cultures for *in vitro* quantification of the soluble and insoluble protein fractions. We used the color scheme *gray – light gray – bright gray*, to distinguish between the least and most soluble ACP fragments.

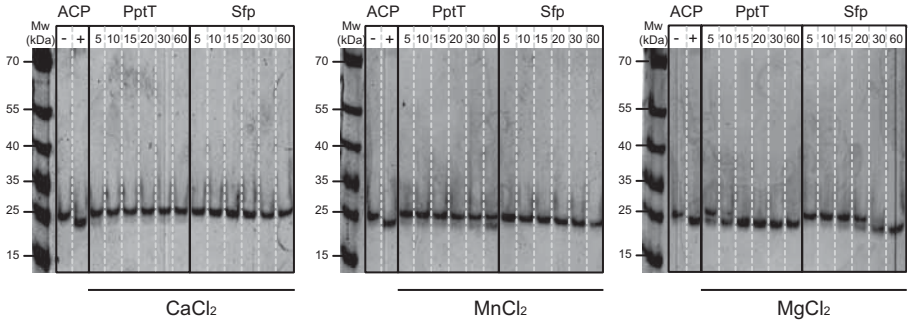


Fig.4. Activation time-course of the ACP domain by PptT and the Sfp from *B. subtilis*. The transfer of the P-pant group from CoA to the ACP domain of PpsC is visualized on a 10% native polyacrylamide gel supplemented with 2.5 M urea in the presence of $CaCl_2$ (left), $MnCl_2$ (middle) and $MgCl_2$ (right) at different time intervals (5, 10, 15, 20, 30 and 60 min). Negative (−) and positive (+) controls correspond to the inactivated and activated forms of the ACP domain, respectively.

(Mofid et al., 2004) whereas the *B. subtilis* AcpS X-ray structure revealed the presence of calcium in its active site (Parris et al., 2000), we then explored the role of all three divalent metal ions onto the activity of PptT. No activation of the ACP domain was detected in the presence of 10 mM $CaCl_2$ with either PptT or Sfp. When replacing $CaCl_2$ with $MnCl_2$, activation of the ACP was noticeable after 30 min incubation, only in the presence of PptT. At 10 mM $MgCl_2$, approximately half of the ACP domain was activated by the transfer of the P-pant moiety after 5 and 20 min incubation with PptT and Sfp, respectively. In conclusion, Mg^{2+} resulted in the highest activity among the divalent ions tested followed by Mn^{2+}, while Ca^{2+} has no effect on the activity of the enzyme.

4. Discussion

The main goal of this study was to examine the use of the GFP1–10/GFP11 system (Cabantous et al., 2005) as a diagnostic tool to assess the solubility of co-factor dependent protein targets *in vivo*. We have demonstrated the feasibility of our approach on the *M. tuberculosis* PptT whose essentiality for growth and persistence of the bacteria *in vivo* makes it a potential candidate for the development of anti-*tuberculosis* drugs (Leblanc et al., 2012). Therefore, identifying a functional and soluble construct of this enzyme is a prerequisite for high-throughput screening (HTS) of inhibitors and structure-based rational design of novel hit compounds. Until now, the soluble MBP–PptT fusion was used for the implementation

of an enzymatic assay suitable for HTS assays (Leblanc et al., 2012). *E. coli* expression of PptT alone using pET vectors was hampered by solubility issues that caused the protein to partition exclusively into inclusion bodies (Fig. 1). Here, we took advantage of the weak AnTet-inducible promoter of the GFP11 vector (Cabantous et al., 2005) to investigate the solubility levels of the full-length PptT from *M. tuberculosis* in addition to N- and C-terminally truncated variants. The picomolar detection level of the split-GFP system helped differentiating between the different constructs according to fluorescence intensity of the reconstituted GFP moiety. Interestingly, the full-length enzyme appeared to be the most soluble protein construct after sequential induction *in vivo* (Fig. 2). Although the *in vitro* experiments followed the same trend, aggregation of the protein over time ultimately led to its precipitation within hours.

We then evaluated the contribution of Mg^{2+} and CoA onto the stability and function of the full-length PptT *in vitro*. A series of experiments using the DSF method was performed. High quality and reproducible fluorescent melting curves with Tm values of 50 °C could be measured only when CoA and $MgCl_2$ were maintained in a large molar excess over PptT, at concentrations equivalent to the purification conditions. Essentiality of these two co-factors for PptT to carry out its function was also demonstrated using a soluble fragment encompassing the ACP domain of the polyketide synthase PpsC (Fig. 3). We opted for a domain trapping strategy (Pedelacq et al., 2011) to effectively identify soluble fragments encompassing the predicted boundaries for this ACP domain (2062–2145). All ~96 selected fragments incorporated the C-terminus of PpsC, probably as a result of less efficient mechanical breakage at the 3′ and 5′ extremities of the DNA molecule. The most compact and soluble ACP fragment (2042–2188) was selected to maximize the chances of monitoring the transfer of the P-pant moiety onto its conserved Ser2105 using either PptT or its closest homologue Sfp from *B. subtilis* (Reuter et al., 1999). Activation of the ACP fragment was confirmed in the presence of Mg^{2+} and, to a lesser extent, in the presence of Mn^{2+}. Surprisingly, activity of the *B. subtilis* Sfp was not detected in the presence of Mn^{2+}, a result not in accordance with the work published by Mofid et al. (2004), although the reaction mixture was incubated at 37 °C instead of 30 °C in our assay.

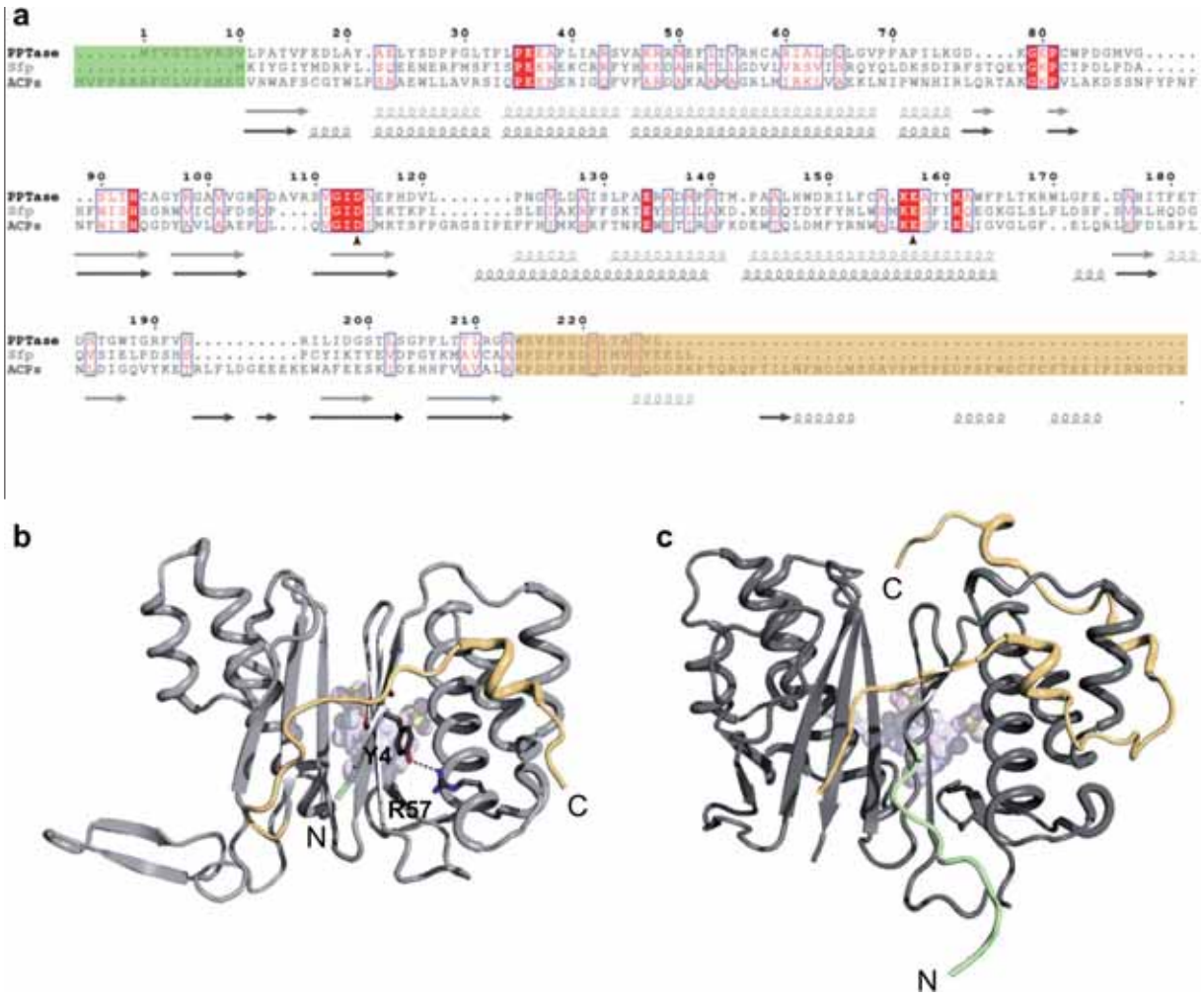


Fig.5. Sequence alignment of PptT with known homologues. (a) Structure-based sequence alignment of the *B. subtilis* Sfp (Uniprot code P39135) and the human AcpS (Uniprot code Q9NRN7) extended to PptT with respect to the conservation of sequence motifs with key residues important for the activity and interactions with CoA. Secondary structure elements deduced from the X-ray structure of Sfp (PDB code 1QR0) and the human AcpS (PDB code 2C43) are shown at the bottom of the aligned sequences in light and dark gray, respectively. Sequence similarities are highlighted in red; sequence identities are shown as white letters on a red background. Strictly conserved catalytic residues D114 and E157 are indicated with brown triangles. The figure was produced with ESPript (Gouet and Courcelle, 2002). Ribbon representation of the X-ray structure of the *B. subtilis* Sfp (b) and the human AcpS (c). CoA is shown in a ball-and-stick representation. N- and C-terminal ends are colored in green and orange in the X-ray structures and the sequence alignment. The side-chains of residues Tyr4 and Arg57 are indicated.

In an effort to rationalize the effect of N- and C-terminal truncations onto the solubiity of the expressed PptT constructs, we generated a structure-based sequence alignment between the homologues Sfp from *B. subtilis* (PDB code 1QR0, (Reuter et al., 1999)) and the human AcpS (PDB code 2c43, (Bunkoczi et al., 2007)), that we further extended to PptT with respect to the conservation of sequence motifs with key residues important for the activity and stability of CoA (Fig. 5a). In the Sfp X-ray structure, N-terminal residues 2 to 7 form the outermost strand of a three-stranded antiparallel β-sheet, at the opposite side of CoA (Fig. 5b). This N-terminal β-strand also interacts with the 26-residue long C-terminal region through hydrogen bond interactions that involve the main-chain nitrogen and oxygen atoms of Ile3 and Gly5. Tyr4 also contributes to the stabilization of the adjacent helix through interactions with Arg57. Disrupting all these interactions would have a dramatic impact onto the stability of the enzyme. In the human AcpS structure, the N-terminal domain is mostly disordered and unstructured with the first residue identified at position 8 in the X-ray structure and strand β1 starting at position 17 (Fig. 5c). We then anticipated that the N-terminal region of PptT adopts an unstructured conformation similar to the one observed for the human AcpS, in accordance with the sequence alignment presented in Fig. 5a. Surprisingly, deletion of C-terminal residues has a dramatic effect on the solubility of PptT, as illustratedby the fluorescence drop to 75% for the truncated variant missing the C-terminus residue L227. In the Sfp and AcpS structures, this highly variable and poorly structured region wraps around the α/β core of the protein (Fig. 5b and c). Stabilization occurs through numerous hydrophobic and van der Waals contacts and a constellation of hydrogen bond interactions that involve the main-chain and side-chain atoms of residues all along the C-terminal region. At this stage, considering the low level of amino-acid sequence similarity between the three enzymes, questions about the role of L227 onto the stability of PptT are difficult to answer as no NMR or X-ray structure is yet available. We are confident that our efforts to overcome this major solubility bottleneck will help determining the three-dimensional structure of PptT, which in turn will serve as a basis for the structure-based rational design of a selective PptT inhibitor.

Acknowledgments

The authors wish to acknowledge Geoffrey S. Waldo and Thomas C. Terwilliger for helpful discussions. This work was supported by European Structural Funds (FEDER) and the Région Midi-Pyrénées (MYCA, Grants 34249 and 09005193; TUBCAN, Grant 11052698) and the Agence Nationale de la Recherche (XPKS-MYCO, Grant 09-BLAN-0298-01).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2013.07.010.

References

An, Y., Yumerefendi, H., Mas, P.J., Chesneau, A., Hart, D.J., 2011. ORF-selector ESPRIT: a second generation library screen for soluble protein expression employing precise open reading frame selection. Journal of Structural Biology 175, 189–197.

Astarie-Dequeker, C., Le Guyader, L., Malaga, W., Seaphanh, F.-K., Chalut, C., Lopez, A., Guilhot, C., 2009. Phthiocerol dimycocerosates of *M. tuberculosis* participate in macrophage invasion by inducing changes in the organization of plasma membrane lipids. PLoS Pathogens 5, e1000289.

Bunkoczi, G., Pasta, S., Joshi, A., Wu, X., Kavanagh, K.L., Smith, S., Oppermann, U., 2007. Mechanism and substrate recognition of human holo ACP synthase. Chemistry & Biology 14, 1243–1253.

Cabantous, S., Rogers, Y., Terwilliger, T.C., Waldo, G.S., 2008. New molecular reporters for rapid protein folding assays. PLoS One 3, e2387.

Cabantous, S., Terwilliger, T., Waldo, G., 2005. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nature Biotechnology 23, 102–107.

Cabantous, S., Waldo, G., 2006. In vivo and *in vitro* protein solubility assays using split GFP. Nature Methods 3, 845–854.

Cadwell, R.C., Joyce, G.F., 1994. Mutagenic PCR. Genome Research, 3–8.

Camacho, L.R., Constant, P., Raynaud, C., Laneelle, M.A., Triccas, J.A., Gicquel, B., Daffe, M., Guilhot, C., 2001. Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. Evidence that this lipid is involved in the cell wall permeability barrier. The Journal of Biological Chemistry 276, 19845–19854.

Chalut, C., Botella, L., de Sousa-D'Auria, C., Houssin, C., Guilhot, C., 2006. The nonredundant roles of two 4′-phosphopantetheinyl transferases in vital processes of *Mycobacteria*. Proceedings of the National Academy of Sciences of the United States of America 103, 8511–8516.

Cornvik, T., Dahlroth, S., 2006. An efficient and generic strategy for producing soluble human proteins and domains in E. coli by screening construct libraries. Proteins: Structure, Function & Bioinformatics 273, 266–273.

Daugelat, S., Jacobs, W.R., 1999. The *Mycobacterium tuberculosis* recA intein can be used in an ORFTRAP to select for open reading frames. Protein Science 8, 644–653.

Dyson, M.R., Perera, R.L., Shadbolt, S.P., Biderman, L., Bromek, K., Murzina, N.V., McCafferty, J., 2008. Identification of soluble protein fragments by gene fragmentation and genetic selection. Nucleic Acids Research 36, e51.

Fisher, A., Kim, W., Delisa, M., 2006. Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. Protein Science 15, 449–458.

Fox, J., Waugh, D., 2003. Maltose-binding protein as a solubility enhancer. Methods in Molecular Biology 205, 99–117.

Gerth, M.L., Patrick, W.M., Lutz, S., 2004. A second-generation system for unbiased reading frame selection. Protein Engineering, Design & Selection: PEDS 17, 595–602.

Gouet, P., Courcelle, E., 2002. ENDscript: a workflow with web interface to display sequence and structure information. Bioinformatics 18, 767–768.

Greener, A., Callahan, M., Jerpseth, B., 1997. An efficient random mutagenesis technique using an *E. coli* mutator strain. Molecular Biotechnology 7, 189–195.

Keenan, R.J., Siehl, D.L., Gorton, R., Castle, L.A., 2005. DNA shuffling as a tool for protein crystallization. Proceedings of the National Academy of Sciences of the United States of America 102, 8887–8892.

Leblanc, C., Prudhomme, T., Tabouret, G., Ray, A., Burbaud, S., Cabantous, S., Mourey, L., Guilhot, C., Chalut, C., 2012. 4′-Phosphopantetheinyl transferase PptT, a new drug target required for *Mycobacterium tuberculosis* growth and persistence in vivo. PLoS Pathogens 8, e1003097.

Listwan, P., Pédelacq, J.-D., Lockard, M., Bell, C., Terwilliger, T.C., Waldo, G.S., 2010. The optimization of *in vitro* high-throughput chemical lysis of *Escherichia coli*. Application to ACP domain of the polyketide synthase ppsC from *Mycobacterium tuberculosis*. Journal of Structural and Functional Genomics 11, 41–49.

Mofid, M.R., Finking, R., Essen, L.O., Marahiel, M.A., 2004. Structure-based mutational analysis of the 4′-phosphopantetheinyl transferases Sfp from *Bacillus subtilis*: carrier protein recognition and reaction mechanism. Biochemistry 43, 4128–4136.

Mossessova, E., Lima, C.D., 2000. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Molecular Cell 5, 865–876.

Niesen, F.H., Berglund, H., Vedadi, M., 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nature Protocols 2, 2212–2221.

Parris, K.D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C.C., Seehra, J., Somers, W.S., 2000. Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. Structure 8, 883–895.

Pedelacq, J.-D., Nguyen, H.B., Cabantous, S., Mark, B.L., Listwan, P., Bell, C., Friedland, N., Lockard, M., Faille, A., Mourey, L., Terwilliger, T.C., Waldo, G.S., 2011. Experimental mapping of soluble protein domains using a hierarchical approach. Nucleic Acids Research 39, e125.

Pedelacq, J.-D., Piltch, E., Liong, E.C., Berendzen, J., Kim, C.-Y., Rho, B.-S., Park, M.S., Terwilliger, T.C., Waldo, G.S., 2002. Engineering soluble proteins for structural genomics. Nature Biotechnology 20, 927–932.

Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H., Walsh, C.T., 1998a. Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chemistry & Biology 5, 631–645.

Quadri, L.E., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., Walsh, C.T., 1998b. Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. Biochemistry 37, 1585–1595.

Reich, S., Puckey, L., Cheetham, C., 2006. Combinatorial domain hunting: an effective approach for the identification of soluble protein domains adaptable to high-throughput applications. Protein Science 15, 2356–2365.

Reuter, K., Mofid, M.R., Marahiel, M.A., Ficner, R., 1999. Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4′-phosphopantetheinyl transferase superfamily. The EMBO Journal 18, 6823–6831.

Stemmer, W., 1994. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. Proceedings of the National Academy of Sciences of the United States of America 91, 10747–10751.

Waldo, G.S., 2003. Genetic screens and directed evolution for protein solubility. Current Opinion in Chemical Biology 7, 33–38.

Yumerefendi, H., Tarendeau, F., Mas, P.J., Hart, D.J., 2010. ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. Journal of Structural Biology 172, 66–74.

Zhao, H., Giver, L., Shao, Z., Affholter, J., Arnold, F., 1998. Molecular evolution by staggered extension process (StEP) *in vitro* recombination. Nature Biotechnology 16, 258–261.

In this paper, we have used the split-GFP technology to design a general protocol to assess the solubility of co-factor dependent protein expression *in vivo*. We have applied this protocol to PptT and PptT truncated versions, and we have identified the full-length construction as the most soluble one as estimated by the fluorescence of the GFP. We have then demonstrated the role of $MgCl_2$ and coenzyme A to help stabilize PptT during the extraction and purification process. This role has been further confirmed by the increased stability of the enzyme in presence of these co-factors as monitored by DSF. We have also verified that PptT was able to transfer a P-pant moiety from CoA onto the ACP domain of PpsC. The ACP domain of PpsC could be activated in only a few minutes, which is also interesting for our structural and functional study of this enzyme.

We were then able to produce and purify large amounts of the full-length PptT in a stable and active form, suitable for crystallization. In an effort to assist the search for specific inhibitors of this essential enzyme in the virulence of *Mtb*, we have determined its structure at high resolution using X-ray crystallography. Crystallization, structural determination and characterization of PptT are detailed in the next chapter.

	Native	Sulfur-SAD
Data Collection		
Space group	*C* 2 2 21	*C* 2 2 21
Unit cell	99.81 121.26 48.83 90° 90° 90°	99.81 121.26 48.83 90° 90° 90°
No. unique reflections	58,592	32,126
Resolution	41.25 - 1.40 (1.44 - 1.40)	76.68 - 2.48 (2.58 - 2.48)
Completeness (%)	99.8 (97.7)	99.8 (98.7)
Redundancy	9.97 (5.23)	24.23 (22.44)
I/σ	14.8 (2.7)	27.5 (15.9)
Rmeas (%)	7.5 (56.5)	8.9 (18.7)
Anomalous correlation		43* (12*)
Refinement		
Rwork/Rfree	0.13/0.17 (0.27/0.35)	
Rmsd bond length	0.016 Å	
Rmsd bond angle	1.936°	
No. atoms		
Protein	1853	
Heteroatoms	91	
Water	297	
Mean B value	19.92 Å^2	

Table 7. Crystallographic data and refinement statistics for the PptT structure.

Values in parentheses refer to the highest resolution shell. * means significant correlation value.

3. STRUCTURAL CHARACTERIZATION OF PPTT

a) Crystallization, Data collection, and Phasing

Using the pTet vector construction of PptT missing the GFP11 strand, and thanks to Mg^{2+} and CoA, we have been able to produce and purify large amounts of PptT in a stable and active form. Screening a thousand crystallization conditions with the help of commercial kits and a crystallization robot, has been sufficient to identify a condition that resulted in the formation of thin platelets of PptT. These diffracted beyond 2 Å resolution using an in house X-ray source (Agilent PX-Scanner). In order to make PptT crystals easier to handle and more reproducible, we adapted a microseeding protocol where crushed platelets served as microseeds for the setup of new crystallization experiments [186], [187]. Using this protocol, a very similar crystallization condition was obtained which resulted in the formation of reproducible, individual thick platelets. These have been used to collect an X-ray diffraction dataset to 1.4 Å resolution at the PROXIMA 1 beamline at the SOLEIL synchrotron in Gif-sur-Yvette, France (**Table 7**).

We have first used the structures of *B. subtilis* Sfp (PDB code: 1QR0) or human AcpS (PDB codes: 2BYD or 2C43) as templates for the molecular replacement method. Unfortunately, a solution could be found only when using the conserved "core" structure as template, but no additional electron density was visible and the protein structure could only be partially built. We then took advantage of the anomalous signal provided by the 8 sulfur atoms (from 4 cysteines and 4 methionines) naturally present in PptT in addition to the 3 phosphorus and the sulfur atoms from CoA to collect a highly redundant SAD (single anomalous wavelength diffraction) dataset at a wavelength of 2Å at the beamline ID-29-1 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France (**Table 7**). The position of 12 atoms could be identified and the structure of PptT has been solved at 2.1 Å resolution. We have then used this structure as a molecular replacement template to

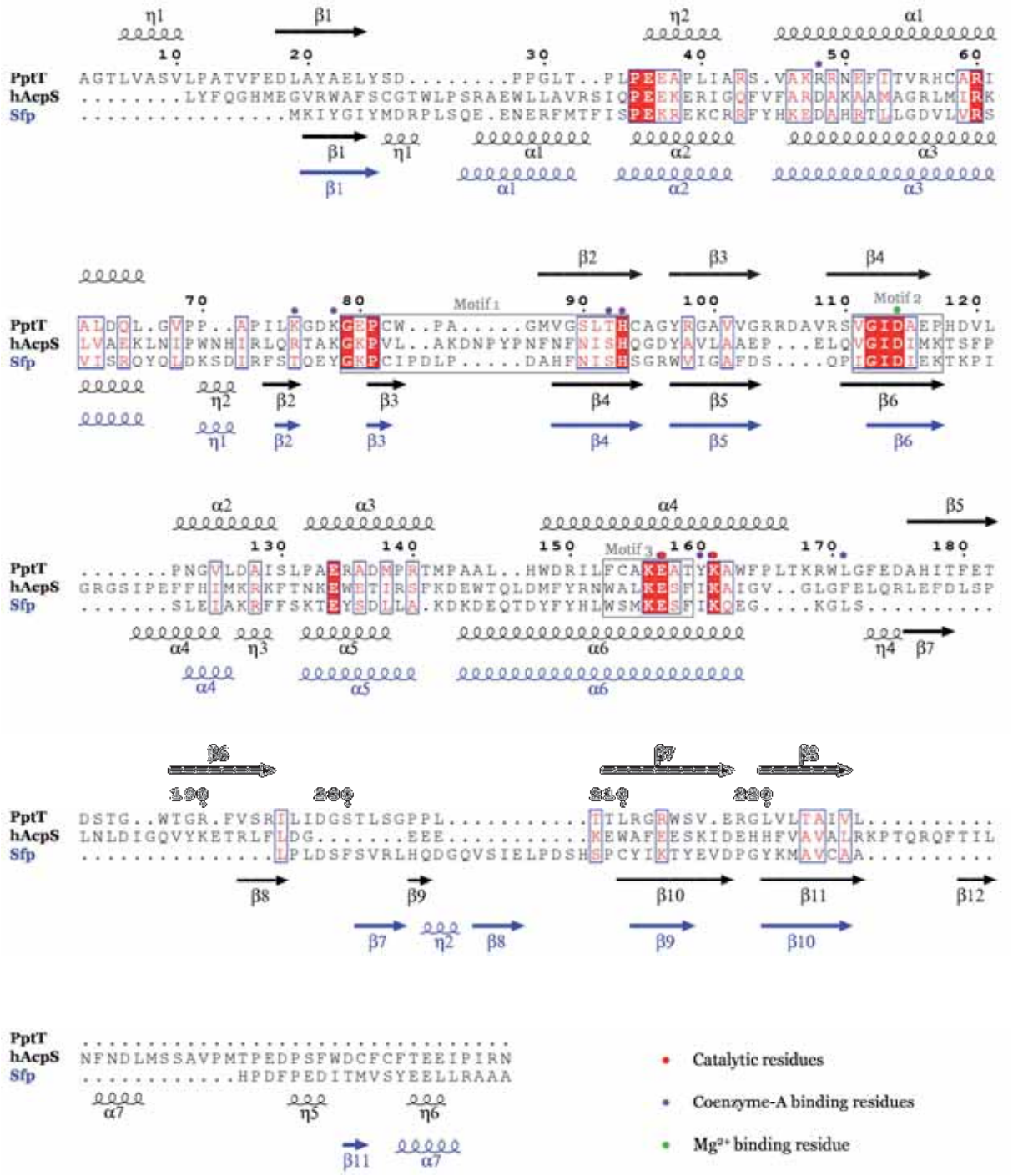


Figure 25. Structure-based sequence alignment of PptT, hAcpS and Sfp.

Identical residues are embedded in blue rectangles with red background. Similar residues are embedded in blue rectangles with white background. Secondary structure elements are represented in black for PptT (top) and hAcpS (bottom) and in blue for Sfp. η is for 3/10 helix, β is for β-strand and α is for α-helix. Sequence motifs of the group II PPTases family are embedded in gray rectangles as in Figure 15B.

solve the structure from the native dataset collected at SOLEIL. The final model has been refined at 1.4 Å resolution to an R-factor of 0.14 and an R-free of 0.17 (**Table 7**).

b) Overall Structure of PptT and Comparison with its Homologues

PptT folds into a pseudo-dimeric structure, similar to the well-characterized Sfp [97] and thUS belongs to the group II PPTases family of enzymes. This family can be divided into two sub-families (**Figure 15B**). Interestingly, PptT possess sequence patterns similar to both sub-families (**Figure 25**), although it lacks the motif 1A which is peculiar to the F/KES sub-family. Since members of these two sub-families are responsible for the same catalytic activity, and no structure from the F/KES sub-family is available yet, PptT cannot be unambiguously classified in either one of the subfamilies. A search using DALI (http://ekhidna.biocenter.helsinki.fi/dali_server/) has yielded only 2 homologues whose structure align with more than 100 of the 224 residues observed in the PptT structure. These correspond to the human AcpS (rmsd 3.0 Å on 190 Cα atoms) and the *B. subtilis* Sfp (rmsd 3.3 Å on 168 Cα atoms), which both belong to the W/KEA sub-family of group II PPTases. Other homologues align with only a hundred residues though with an average rmsd of about 2.5 Å and correspond to group I PPTases.

Like Sfp, PptT folds into two nearly identical domains connected by a short linker region (**Figure 26)**. The N-terminal domain extends from M1 to V108 whereas the C-terminus domain extends from R109 to L227. The main difference between these two domains lies in the first 18 residues of the N-terminus domain that are dedicated to the closure of the hydrophobic core formed by the β-sheets of the two sub-domains. They are located in a position similar to the β-sheet of the third monomer in the trimeric group I PPTase family (**Figures 27 and 13A**). The first residue fulfilling this role is L6 which is stabilized by the side chain of E23 and is involved in the short 3/10 helix η1 stacked against the hydrophobic core.

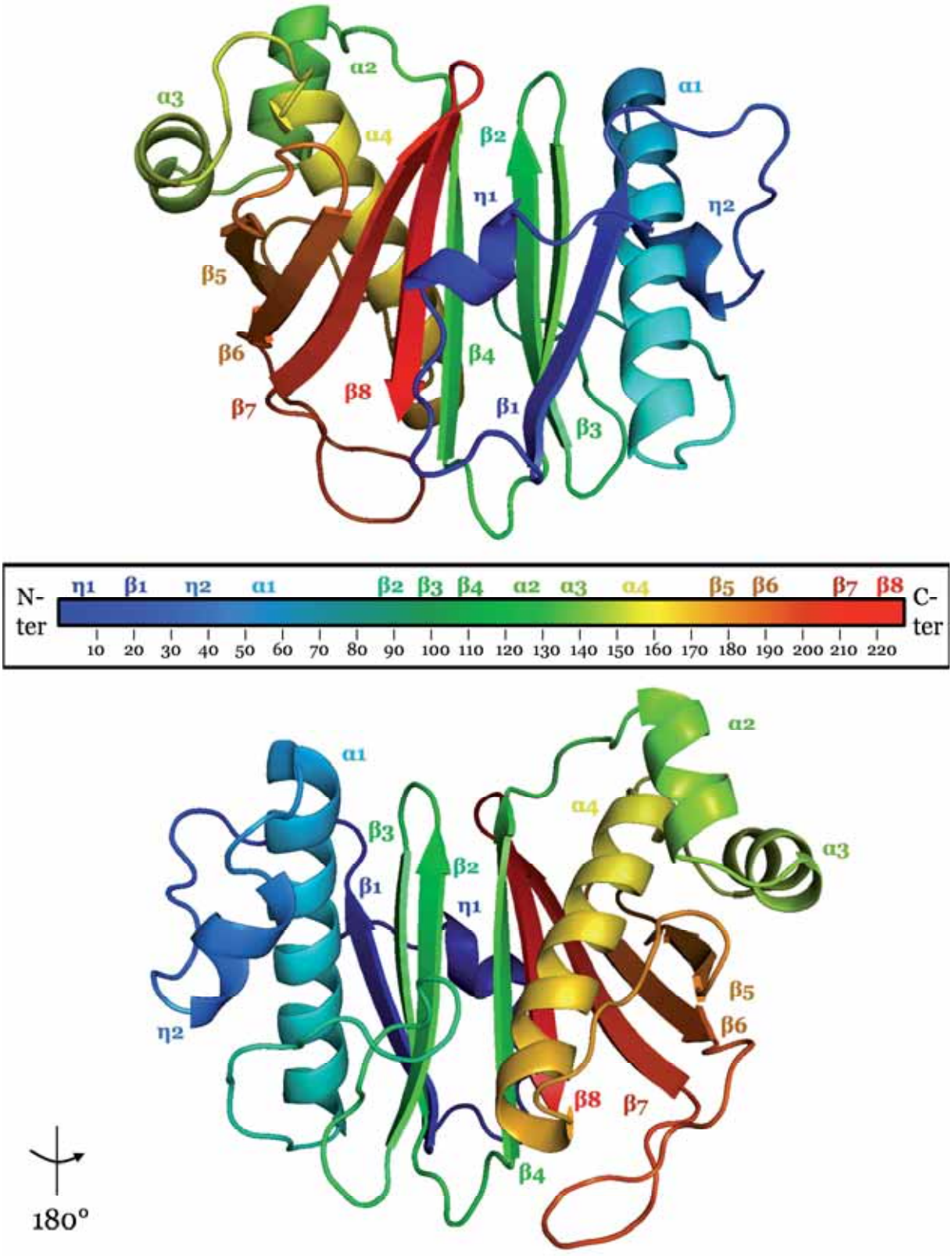


Figure 26. Overall PptT structure and secondary structure assignment.

η is for 3/10 helix, β ix for β-strand and α is for α-helix. PptT is 227 residues long. PptT structure starts at residue 3 and ends at residue 227.

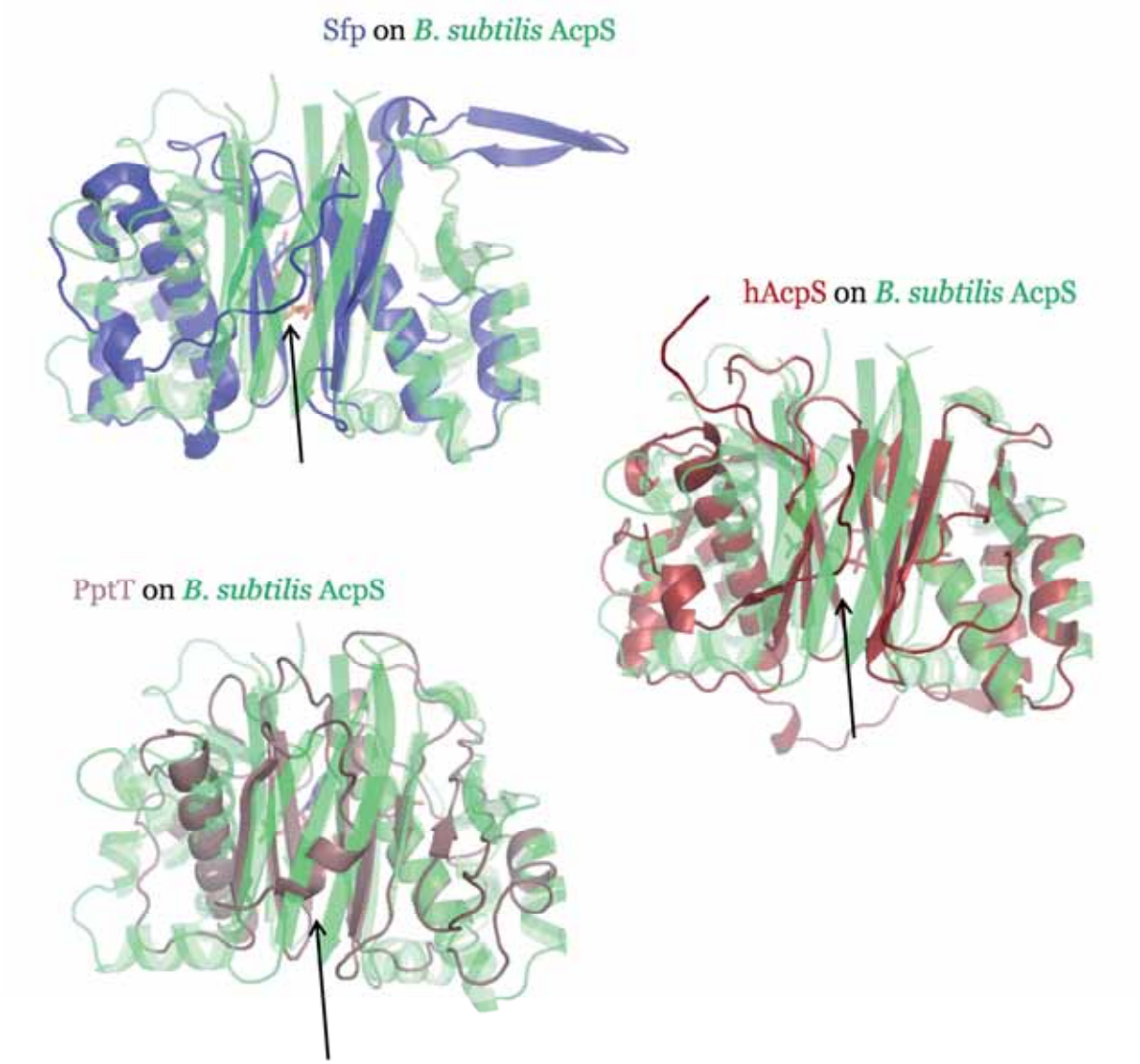


Figure 27. Superimposition of Sfp, hAcpS and PptT on *B. subtilis* trimeric AcpS.

The structure of the group-II PPTases Sfp (PDB code: 1QR0) in blue, hAcpS (PDB code: 2C43) in red and PptT in brown, have been superimposed on the structure of the trimeric group-I PPTase AcpS from *B. subtilis,* which is colored transparent green. In the representation of the trimeric AcpS, only the β-sheet from the third monomer was drawn for clarity. Superimpositions show the role of extensions in group-II PPTases in closing the hydrophobic core. These extensions superimpose with the β-sheet of the third monomer of AcpS and are located at the C-terminus of Sfp and hAcpS, or at the N-terminus of PptT (black arrows).

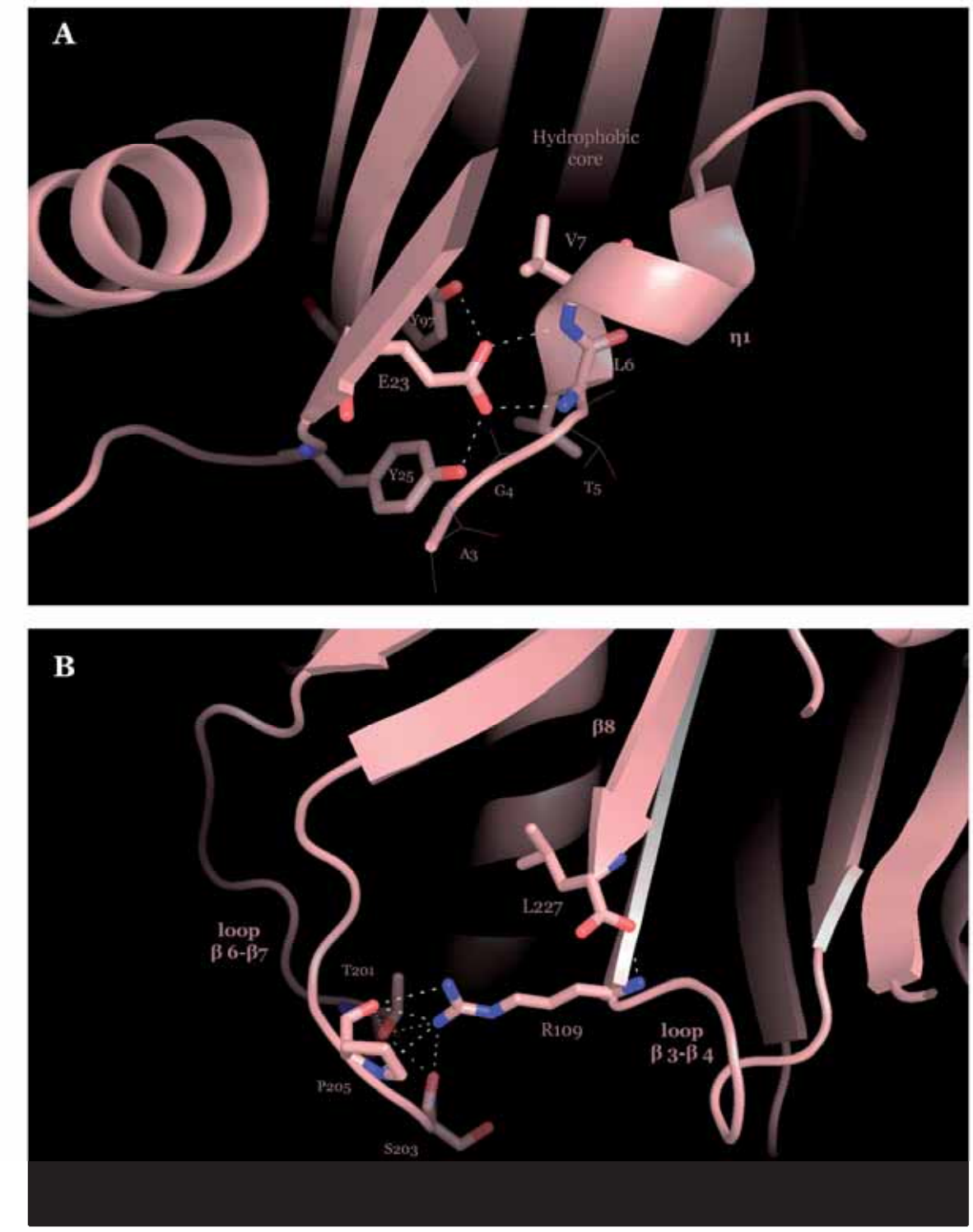


Figure 28. Role of the N- and C-terminal residues in the PptT structure.

(A) Role of N-terminal residues. Residues involved in interactions are shown in sticks, whereas residues not involved are shown in lines. The first two residues of PptT are not visible in electron density and hence have not been built. A3, G4, and T5 do not seem to play any role in maintaining the PptT structure. On the contrary, L6 and V7 main-chain nitrogen atoms are involved in a hydrogen bond network with E23, and their hydrophobic side chains point towards the hydrophobic core of the PptT structure. (B) Role of L227 in stabilizing the β3-β4 and β6-β7 loops via R109.

When searching for more soluble constructs of PptT, we have tested a combination of N- and C-terminally truncated versions of the enzyme (see **Rottier, Faille et al., Figure 2**). Using the split GFP technology, we have established that the deletion of the first five residues only slightly hampered PptT solubility. On the contrary, deleting the last residue L227 was sufficient to observe a large decrease in solubility. These results can easily be interpreted when examining the PptT structure. Indeed, the first five residues do not take part into any secondary structure and are not involved in any interaction with the remaining part of the protein (**Figure 28**). Deleting the residues L6 and V7 would probably have a more dramatic effect on PptT solubility since, in addition to their implication in the hydrogen bonding network through the side chain of E23, they prevent solvent from entering the hydrophobic core of the protein. The large negative effect on solubility caused by the deletion of the last residue L227 can also be explained. The L227 carboxy-terminus is located at the extremity of β8 and is part of the C-terminal β-sheet (**Figure 29**). It also helps stabilize the β6-β7 and β3-β4 loops via an extensive hydrogen bond network involving R109.

Interestingly, in Sfp [97] and hAcpS [92], closing of the solvent access to the hydrophobic core does not involve any N-terminal residues but instead, 30 C-terminal residues absent from the PptT sequence (**Figures 25 and 27**). Other structural differences appear when comparing PptT to Sfp and hAcpS. The positions of the strands β5 and β6 in PptT, although similar to those in hAcpS (β7 and β8) largely differ from those in Sfp (β7 and β8). In PptT, they are stacked against the long α4 helix whereas in Sfp they come out of the structure and participate in crystal packing by interacting with several symmetry-related molecules. Their position in Sfp might not then reflect the solution structure but show the flexibility of this structural element, to which no function has been attributed yet. It also illustrates that PptT ressembles more hAcpS than Sfp. Finally, PptT also bears a 16 residues insertion compared to both Sfp and hAcpS. Located between L196 and R211, this insertion creates a small cavity on top of the protein that does not exist in the two homologues.

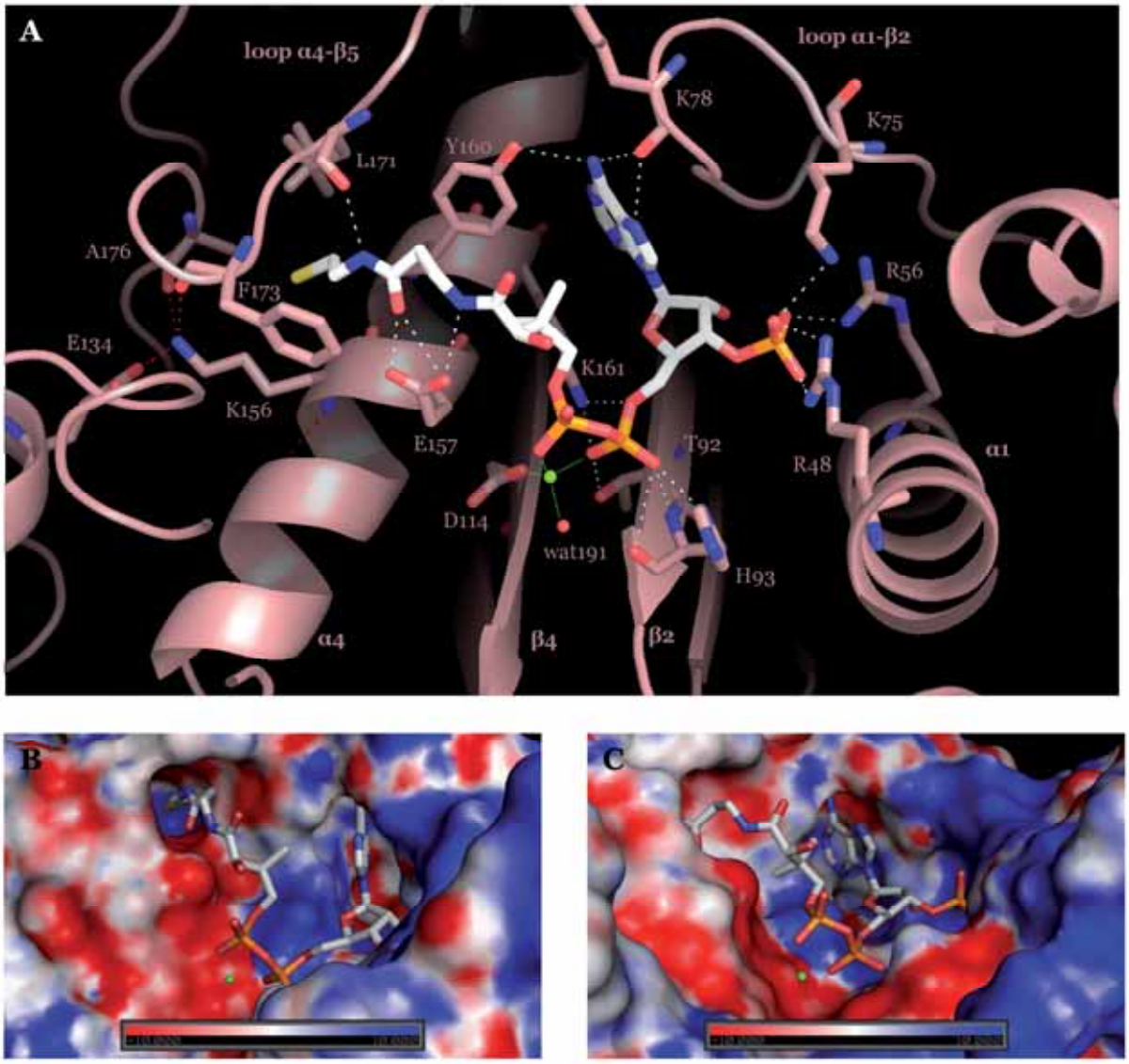


Figure 29. Coenzyme-A binding in PptT.

(A) PptT residues and CoA are represented in sticks. Interactions between PptT and CoA are represented in white. Interactions involving K156 seem to stabilize the cavity in which the P-pant enters and are represented in red. The Mg^{2+} ion and its coordinations are represented in green. (B-C) Views of the electrostatic potential surface of the CoA binding pocket in the PptT structure.

c) Coenzyme A Binding

CoA binds in a cleft delineated by the core elements α1, α4, β2 and β4 and the loops α1-β2 (residues 67-86) and α4-β5 (residues 168-175) (**Figure 29**). It is stabilized by numerous polar interactions mainly involving side chains from the core elements (**Figure 29A**). A magnesium ion has been found to coordinate with the CoA pyrophosphate in addition to the side chain of D114 and a water molecule (wat191), a feature already observed in both Sfp and hAcpS structures. Interestingly, in the structures of these homologues, the Mg^{2+} coordinates with an additional glutamic acid. In PptT, E157 lies in a position similar to the glutamic acid in Sfp and hAcpS, but it participates to the stabilization of the P-pant moiety. A possible rotamer of the E157 side chain would allow the carboxylate to reach the coordination sphere of Mg^{2+} but there is no evidence from electron density that this is the case. Despite the low sequence identity between homologues, residues whose sides chains are involved in polar interactions with CoA are well conserved (**Figure 25**). Indeed, to the exception of Y160, the five other residues are either similar or identical and participate in the same interactions with CoA as in the Sfp and hAcpS structures. A conserved lysine residue (K156, **Figure 29A**) seems to be responsible for the relative positioning of the α4 core helix and the loop α4-β5 that form the buried cavity in which the P-pant enters in the hAcpS and PptT structures. In the Sfp structure, the P-pant is not visible in the electron density and probably solvent exposed (**Figure 16**), and no similar buried cavity could be observed.

CoA binding is also driven by electrostatic patches (**Figure 29B-C**). Along with Mg^{2+}, the P-pant part lies on an electronegative surface, to the exception of the bottom of the buried cavity, which is rather electropositive. On the contrary, the adenosine phosphate lies on an electropositive surface, especially the 5'-phosphate which interacts with the side chains of R48, R56 and K75 (**Figure 29A**). Only the adenosine N6 is stabilized by the side chain of Y160. Similar electronegative and electropositive patches are found in Sfp, but in this case the P-pant part of CoA is not visible in the structure. Also, the shape of the CoA-binding pocket largely differs from that of PptT and is more solvent-exposed. In hAcpS, both shape and electrostatic surface are similar to those found in PptT. However, the P-pant moiety is also buried

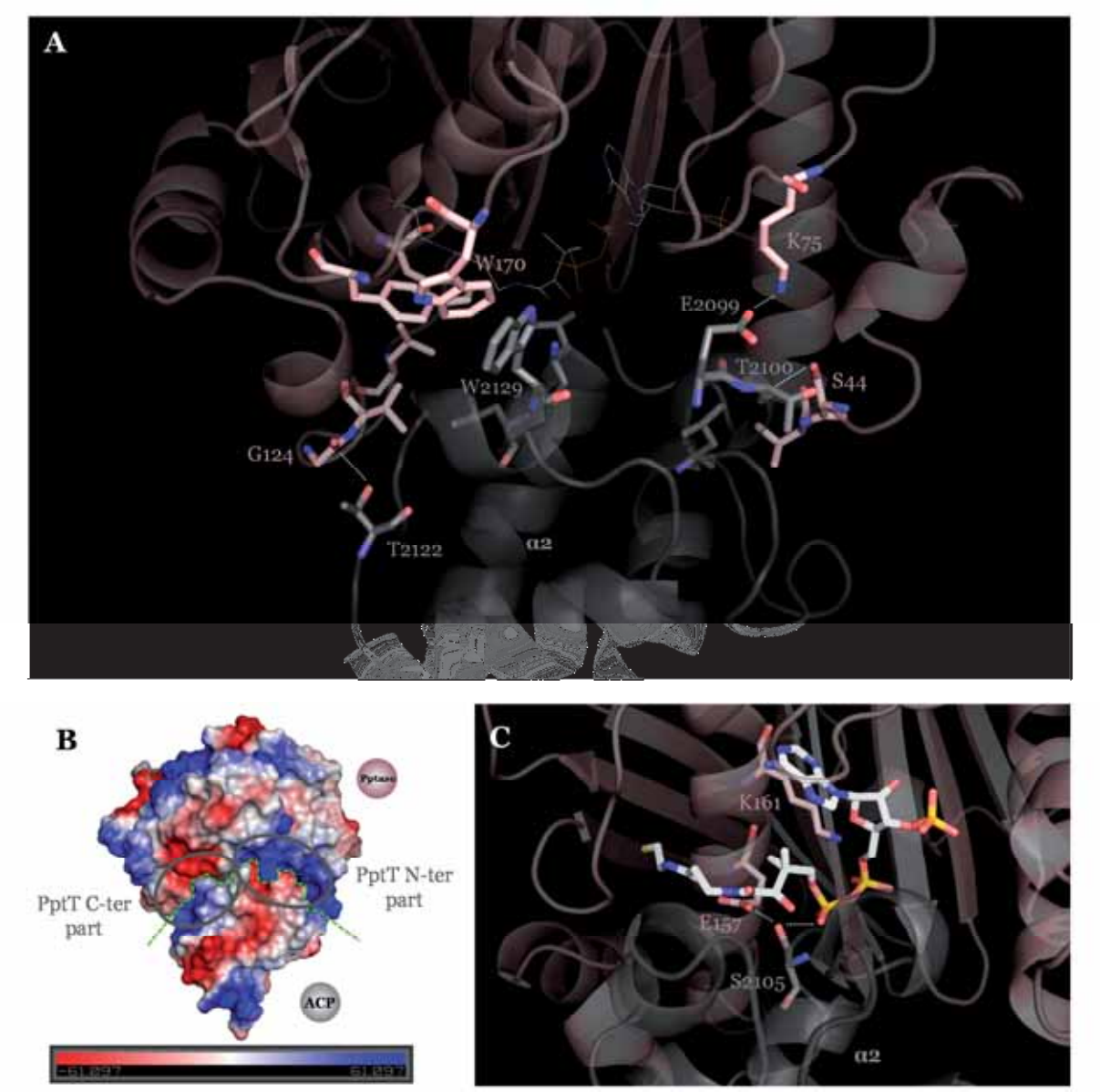


Figure 30. Docking model of the PptT-PpsC ACP complex.

(A) PptT is represented in brown and ACP is represented in gray. Residues involved in interactions between the two proteins are represented in sticks. Labeled residues could be responsible for polar interactions (white dashed lines). Non-labled residues may participate in hydrophobic interactions. CoA is represented in lines. (B) Electrostatic potential surfaces of PptT and ACP separated by a green dashed line showing the complementarity between the two proteins. (C) The putative catalytic residues on both PptT and ACP are represented. Interactions (white dashed lines) that may form between PptT E157 and ACP S64 (2.5 Å), and between S64 and pyrophosphate (3.4 Å) are compatible with the catalytic mechanism proposed for hAcpS.

in a cavity but in a different conformation and not stabilized by any polar interaction. Also, the electrostatic surface where lies the adenosine moiety is of opposite charge in hAcpS compared to PptT. As to why CoA is required for the stability of PptT and not for hAcpS, still remains an enigma.

d) Docking of PpsC ACP Domain and PptT Catalytic Mechanism

To date the only structure involving a group II PPTase in complex with a CP domain, is between hAcpS and the human ACP (hACP). This complex helped the authors to describe a detailed mechanism of P-pant transfer (see **section I.B.2.c**). Nonetheless, if the catalytic mechanism involves highly conserved residues and is thus probably similar in all group II PPTases, the interaction between hAcpS and the ACP may not be universal considering the low amino-acid identity among both group II PPTases and CPs. We have therefore modeled a complex between the PpsC ACP domain and PptT by computational docking. We used the same PpsC ACP domain model generated by MODELLER, as in **section II.A.4**. In the hAcpS-hACP complex (PDB code: 2CG5; **Figure 16C**) and hAcpS (PDB code: 2C43) structures, the CoA positions identically. We therefore used the position of the CoA in our PptT structure as a constraint to guide molecular docking. More specifically, in all binding modes calculated by PatchDock, the distance between the oxygen atom of the ACP serine and the phosphorus of the P-pant phosphate was not allowed to be greater than 5 Å of the. After refinement with FireDock, only one docking solution still satisfied this unique condition.

In the resulting complex, the oxygen atom of S2105 is 3.1 Å away from the P-pant phosphore (**Figure 30C**). Additional evidences indicate that our docking model could be relevant for the description of PptT-ACP interactions. Indeed, the PptT glutamic acid (E157), which probably activates the ACP serine (S2105) is 2.7 Å away from the serine (**Figure 30C**) and in a suitable position to catalyze deprotonation, as seen in the hAcpS-hACP structure (**Figure 16**). Furthermore, a putative Π-stacking

interaction is observed between PptT W170 and W2129 of PpsC ACP (**Figure 30A**). It is noteworthy that in the PptT experimental structure a well-defined imidazole positions identically to the pyrrole ring of ACP W2129. Likewise, the main-chain oxygen of ACP T2100 which forms a polar interaction with S44 of PptT superposes well with the oxygen of a phosphate molecule involved in an interaction with S44 in the PptT experimental structure.

Interestingly, we have found similar interactions in the model of the PptT-ACP complex as those occurring in the structure of the hAcpS-hACP. These interactions mainly comprise hydrophobic contacts between the α1 and α4 long helices of PptT and the α1 and α2 helices of the ACP domain. A few polar interactions also seem to participate to PptT-ACP docking, namely G124-T2122, K75-E2099, S44-T2100. PptT and the PpsC ACP domain also present electrostatic patches of opposite charges that may drive their interaction (**Figure 30B**). Furthermore, two main contact areas can be observed in the PptT-ACP model. One involves a negatively charged patch at the C-terminus of PptT and the electropositive ACP helix α2. The other one involves a positively charged patch at the N-terminus of PptT and the electronegative α1-α2 loop of the ACP domain.

The group II PPTase catalytic mechanism has been well described for hAcpS (**Figure 16C**). Our PptT-PpsC ACP docking model, which as explained above is structurally similar to the hAcpS-hACP structure, is also compatible with the catalytic mechanism described for the latter. In hAcpS, this mechanism involves two residues playing the role of acid/base catalysts. These two residues are strictly conserved in PptT (E157 and K161) and are indeed in a suitable position to catalyze the P-pant transfer onto the catalytic S2105 of ACP. As in hAcpS-hACP, S2105 benefits from the helix-dipole effect of the α2 helix (**Figure 30C**), and the side chain of E157, located 2.7 Å away from the catalytic serine, would abstract a proton from it. K161 would then protonate the CoA α-phosphate and the the serine hydroxylate would attack the β-phosphate, which is only 3.4 Å away from S2105 in our model. It would result in the thioester attachment of the P-pant moiety onto the ACP S2105 and the release of an adenosine di-phosphate molecule

e) Experimental Procedures

PRODUCTION AND PURIFICATION

The protocol described in **Rottier, Faille *et al.***, was applied for the structural study of PptT:

Frozen cells expressing the full-length PptT from a pTet vector missing the GFP11 tag were used to start an overnight 3 ml LB-spectinomycin (112 µg/ml) culture at 32 °C prior to inoculation in baffled flasks containing 500 ml of the same media. Cells were allowed to grow for approximately 2 h at 37 °C before temperature was dropped from 37 °C to 20 °C. When OD_{600} reached 0.5–0.7, cells were induced with anhydrotetracycline at a final concentration of 8 µg/ml and grown for an additional 18 h prior to harvesting by centrifugation at 4000*g* for 30 min and storage at –80 °C. The pelleted cells were suspended in 40 ml of buffer A (100 mM Tris, 300 mM NaCl, 10 mM $MgCl_2$, 50 µM CoA pH 8.0) and lysed by sonication (6 cycles of 30 s pulse, 50% amplitude, power 5) prior to centrifugation at 20,000*g* for 1 h. The supernatant was filtered (0.2 µm) and purified on an ÄKTA purifier system (GE Healthcare) following a two-step procedure that combined an immobilized metal ion affinity chromatography (IMAC) and a size-exclusion chromatography (SEC). The lysate was loaded onto a 1 ml HisTrap HP (GE Healthcare) affinity column. The N-terminally His-tagged PptT was eluted from the column with a step of buffer A supplemented with 120 mM imidazole. The eluted protein fractions were pooled and concentrated down to 2 ml using a Vivaspin 20 centrifugal concentrator (Sartorius), prior to injection into a HiLoad 16/60 Superdex 200 (GE Healthcare) pre-equilibrated with buffer B (50 mM Tris pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 50 µM CoA).

CRYSTALLIZATION

Initial crystals were obtained from 0.2 M lithium sulfate, 0.1 M phosphate-citrate pH 4.2, 20% (w/v) PEG 1000 while screening approximately a thousand conditions through QIAGEN NeXtal commercial kits using a Nanodrop ExtY crystallization robot. Crystallization was further optimized using microseeding

techniques in Linbro 24 well plates. Several crystals from the initial condition were crushed in 200 µl of reservoir solution and serially diluted. Final crystallization setup consisted in 0.5 µl of each dilution, 2 µl protein and 1.5 µl 0.2 M lithium sulfate, 0.1 M phosphate-citrate pH 4.0, 22% (w/v) PEG 1000 and yielded thick platelets of 0.2-0.4 mm length. These were then handled in nylon loops and flash frozen in liquid nitrogen. All chemicals used were provided by Sigma-Aldrich.

DATA COLLECTION, PHASING AND STRUCTURE DETERMINATION

A native dataset was first collected on PROXIMA 1 beamline at the SOLEIL synchrotron in Gif-sur-Yvette, France. As molecular replacement was not efficient to obtain the structure of PptT, we collected a highly redundant dataset consisting in 13 scans of 120° rotation each, on the ID-29-1 beamline at the ESRF in Grenoble, France. The diffraction data were collected at a wavelength of 2 Å in order to accumulate the anomalous signal from the sulfur and phosphorus atoms. Data were processed with XDS, and scans were scaled together with XSCALE. Strong anomalous signal was detected and the SHELX software suite allowed us to identify nine of the eleven sulfur and phosphorus atoms present in the PptT-CoA complex, and obtain a density map using the single wavelength anomalous diffraction (SAD) phasing method. PHENIX AutoBuild was then used to automatically build the PptT structure. CoA was manually added and the structure was finally refined using REFMAC5 and COOT. Statistics of the datasets are detailed in **Table 7**.

PPTT-ACP DOCKING

Except from CoA, hetero atoms were first removed from the PptT refined structure. The resulting pdb file was then uploaded and set as the receptor molecule for molecular docking using the PatchDock webserver (http://bioinfo3d.cs.tau.ac.il/PatchDock/). The PpsC ACP domain model generated by MODELLER described in section II.A.4 was used as the ligand molecule. A 5 Å distance constraint between the side-chain oxygen atom of ACP S64 and the β-phosphate phosphorus atom of CoA was set and the docking was run with default

values. The twenty best results were then streamed to the FireDock webserver (http://bioinfo3d.cs.tau.ac.il/FireDock/) and ran with default values. The second result in terms of global binding energy (-27.88 kcal/mol) was then considered as the most relevant docking model and is the one described in section II.B.3.d.

III. CONCLUSION & PERSPECTIVES

A. CONTEXT

Although polyketides are widely used as drug molecules in diverse pharmaceutical applications accounting for several millions of dollars in industry, many of them remain to be discovered and characterized. The recent availability of numerous sequenced genomes should accelerate this process. New challenges also concern the direct engineering of polyketide synthases, especially modular polyketide synthases. In this exciting research field, advances in biochemical characterization of polyketide synthases have recently allowed the synthesis of new polyketide compounds. Also, the expensive chemical synthesis of polyketides is beginning to be replaced by natural biosynthesis. Nonetheless, in both cases, the lack of details in the polyketide biosynthesis mechanism remains the major bottleneck to take these strategies one step beyond. For that reason, efforts are now focusing on deciphering the complex mechanism used by polyketide synthases and other enzymes to synthesize these interesting molecules. Despite great advances during the last decade, there are still a number of outstanding issues that need to be addressed, including the substrate specificities of each catalytic domain and the architecture of these multi-domain enzymes. During my PhD thesis, I have undertaken the functional and structural characterization of the mono-modular polyketide synthase PpsC and its activator PptT. These two enzymes are responsible for the synthesis of polyketide-derived lipids that provide its virulence to *Mycobacterium tuberculosis* (*Mtb*). While PpsC is essential for the production of phtiocerol dimycocerosates, PptT catalyzes the post-translational modification of all the polyketide synthases and is thus essential to the biosynthesis of all polyketide-derived molecules in *Mtb*. In the long run, this study may help understand the mechanism of polyketide biosynthesis and assist the design of specific inhibitors of these two enzymes with the hope of developing new antituberculosis drugs.

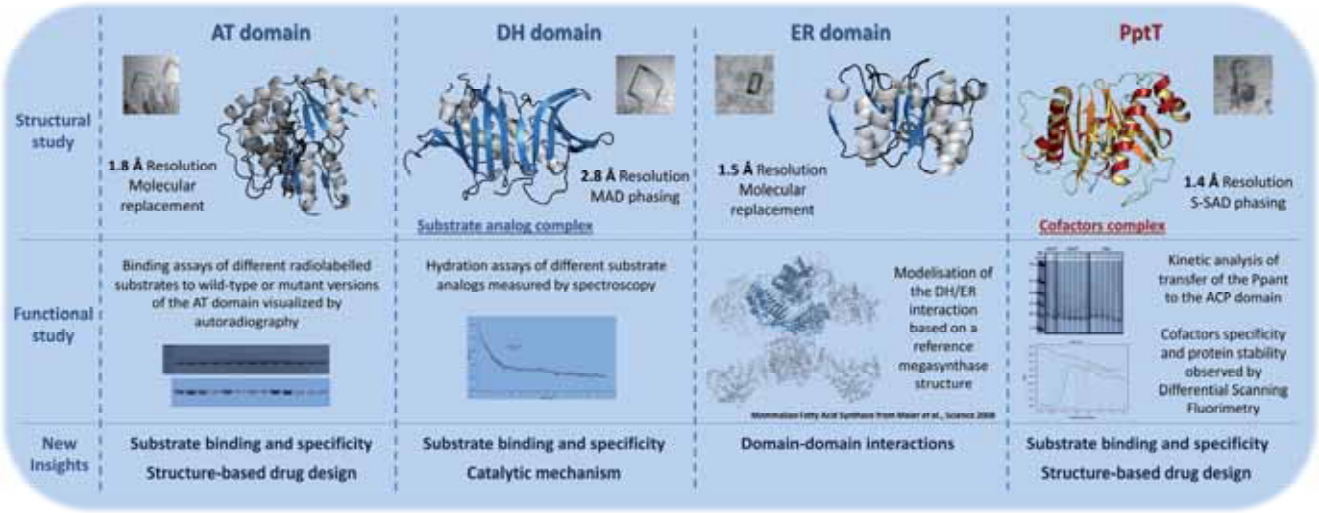


Figure 31. Summary of the results of the functional and structural study of PpsC and PptT.

B. MAIN RESULTS

In this study, we have first used the recently developed *domain trapping* strategy to identify soluble fragments representing each domain of the polyketide synthase PpsC from *Mycobacterium tuberculosis*. These fragments have then been produced and purified in *Escherichia coli* for further structural and functional characterization. Crystallization trials realized on 19 fragments spanning the full-length PpsC have allowed us to obtain crystals for 6 of them and solve the structures of the acyltransferase domain, the dehydratase domain, and the nucleotide-binding sub-domain of the enoylreductase domain from PpsC. In addition, we have solved the structure of the dehydratase domain in complex with crotonyl-coenzyme A that we have identified, using enzymatic tests, as a suitable substrate for the catalytic activity of this domain. The structure of the acyltransferase domain provides us with useful information about substrate specificity. A detailed analysis of the structure and function of this domain will be the point of a review together with structures of acyltransferases from other PKSs solved in our group. Using the structure of the dehydratase/crotonyl-coenzyme A complex, we have characterized the interactions that drive the substrate specificity of the dehydratase domain. Finally, both enoylreductase and dehydratase structures have helped us understand how the β-carbon processing domains assemble into a polyketide synthase. Based on functional and structural similarities with the previously characterized mammalian fatty acid synthase, we have indeed established a model of the assembly of these two domains that we extended to the acyl carrier protein domain. We have proposed that type-I mono-modular polyketide synthases assemble in a way similar to mammalian fatty acid synthases (mFAS), to the contrary of multi-modular polyketide synthases for which a different assembly mode seems to prevail.

In parallel, we applied the split GFP technology to diagnose the importance of the co-factors CoA and Mg^{2+} on the solubility of the *Mtb* 4'-phosphopantetheinyl transferase PptT *in vivo* and *in vitro*. We are now able to obtain this enzyme in a soluble and stable form without the need for a maltose binding protein fusion. We

have also confirmed using biophysical techniques the importance of co-factors for the in vitro stability of PptT. Crystallization trials on this enzyme have resulted in the formation of clusters of platelets that were further optimized using microseeding techniques to yield high resolution diffracting crystals. The PptT structure has then been solved using the single anomalous wavelength diffraction method. Analysis of the structure led to the classification of PptT in the group-II PPTase family of enzymes, represented by the well-characterized Sfp. We have also found an interesting docking solution with a model of the acyl carrier protein domain of PpsC that allowed us to propose a catalytic mechanism based on the previously described complex between the human Acp synthase (AcpS) and the acyl carrier protein domain of mFAS.

C. PERSPECTIVES

1. FURTHER APPLICATIONS OF THE *DOMAIN TRAPPING* STRATEGY

The *domain trapping* strategy has allowed us to identify soluble fragments representing each domain of PpsC. Our fragment study, along with published results on domains of other PKSs, has highlighted the limitations of a structural study on individual domains. For example, the KS domain has only been structurally characterized when expressed in fusion with the AT domain. The *domain trapping* method should allow the identification of fragments encompassing both the KS and the AT domains of PpsC, provided that a DNA library of fragments with a suitable size has been generated and that iPCR primers are located in the linker region between these two domains. Obviously, such strategy can be envisioned for the other domains of PpsC. The successful application of the *domain trapping* strategy on PpsC has also encouraged us to apply it to other PKSs. In particular, such approach could be used on PpsB and PpsD, which we believe are interacting with PpsC in order to stream the growing polyketide. Our domain trapping method would notably be useful to identify the boundaries of the docking domains and help the structural study of complexes between the KS and ACP domains of these PKSs.

2. SUBSTRATE SPECIFICITY OF POLYKETIDE SYNTHASES DOMAINS

Modifying the substrate specificity of the different catalytic domains is a key step in an attempt to produce new polyketides from engineered PKSs. In this respect, AT domains from several PKSs have been mutated to accommodate malonyl-CoA instead of methylmalonyl-CoA. Unfortunately, the residues identified as responsible for this specificity switch are not strictly conserved and thus only concern a restricted number of PKSs. Another solution to alter elongation unit specificity of a PKS is the replacement of its AT domain by another AT domain possessing a different specificity, a strategy known as domain swapping, but it often comes at the cost of decreased product titres. In most cases, the only remaining solution is to directly mutate the AT domain. Therefore, the residues responsible for the specificity of additional AT domains must be identified. Preliminary results accumulated in our group on the functional and structural study of the AT domains from Mas, PpsA, PpsC (unpublished results), and Pks13 [135], indicate that substrate specificity of these domains can also be changed when mutating specific residues. Therefore, our study will certainly allow deciphering the determinants of elongation unit specificity for an increasing number of PKSs.

We have also participated to the characterization of the dehydratase domain specificity. No structural information was available concerning a complex with a substrate that could actually be (de)hydrated. Instead, only mechanistic inhibitors were co-crystallized with FAS-II dehydratases. Our functional and structural work on the dehydratase domain of PpsC and its substrate analogue crotonyl-coenzyme A provides useful information on the substrate specificity of dehydratases and their catalytic mechanism. Nonetheless, similar experiments should also be carried out with substrates that are more similar to the physiological growing polyketides. For example, the DH domain of PpsC catalyzes the dehydration of the β-hydroxyl from a substrate longer than crotonyl by 20 carbons and which bears two additional hydroxyl substitutions at positions C5 and C7. These chemical groups probably enhance the substrate affinity for the DH domain through interacting residues that need to be characterized.

Our strategy for the structural study of the dehydratase domain in complex with its substrate can potentially be applied to any dehydratase domain or stand-alone enzyme. This can be achieved either by co-crystallization or by soaking in a solution of commercially available crotonyl-coenzyme A, to the condition that the strictly conserved catalytic histidine at position 959 has been mutated into a phenylalanine. Applying this strategy to various dehydratases should (i) facilitate the identification of residues responsible for the specificity of these enzymes toward their substrates and (ii) allow engineering variants of the dehydratase domain that keep a high affinity for non-natural substrates, a feature interesting for high-throughput production of non-natural polyketides.

3. DOMAIN-DOMAIN INTERACTIONS AND ASSEMBLY OF THE BETA-CARBON PROCESSING DOMAINS

Very little information is available concerning domain-domain interactions in megasynthases. Also, no complex involving an acyl carrier protein domain and another catalytic domain from a polyketide synthase has ever been reported. The structure of the DH domain of PpsC in complex with crotonyl-coenzyme A has allowed us to propose a docking model between *holo*-ACP and the DH domain. A possible way to validate our model could be the mutation of putative interface residues. For example, our model helped identifying residues which might be involved in hydrogen bonding interactions mediating the binding of the ACP domain onto the DH domain. With the possibility of measuring the catalytic activity of the DH domain on substrates tethered to the ACP domain, we could compare the activity of DH and/or ACP mutants with their corresponding wild-type versions. Another perspective for the study of the DH/ACP interaction is to use our docking model as a basis for a rational engineering approach to tighten up the two domains by di-sulfide cross-linking between residues mutated into cysteines. Using a web program

intended to design disulfide bonds from structural information [http://cptweb.cpt.wayne.edu/DbD2/], at least two pairs of residues that can fulfill this role have been identified.

We have also used our DH/ACP docking model to generate an assembly model of the DH-ER-ACP domains in PpsC, based on the mammalian FAS assembly. In our model, the ACP interacts with both the DH and ER domains, thus revealing the possibility of obtaining a stable complex between these three domains. To do so, we would proceed in two steps: (i) loading of a crotonyl-P-pant molecule onto the ACP domain with the help of PptT, and (ii) forming a stable complex with a DH-ER fragment bearing an inactive version of the DH domain. In addition, we are still working on the optimization of the crystallization conditions of the fragment 6 (DH-ER-KR tri-domain), whose structure would definitely unravel the mechanism by which the β-carbon processing domains assemble in mono-modular PKSs.

4. ANTITUBERCULOSIS DRUG DESIGN

Both PpsC and PptT have arisen as promising drug targets against tuberculosis. Indeed, when PpsC is not expressed, DIM are not produced, which results in great attenuation of the virulence of *Mtb*. However, all the PpsC domains cannot be considered equally. On the one hand, the domains responsible for condensation (i.e. KS and AT) are obviously essential to the activity of the enzyme and thus represent potential drug targets. On the other hand, we do not have any evidence about the therapeutic value of inhibiting the β-carbon processing domains. For example inhibiting the DH domain of PpsC could either result in the formation of a DIM variant in which the β-carbon processed by PpsC (C9 for DIM A/B or C10 for DIM A'/B') bears a hydroxyl substitution, or may not be processed by the downstream enzymes because of its unusual structure. As far as we know, there is no clue indicating how a growing polyketide would be processed in such a pathway. Therefore, in our study the structure of the AT domain appears as the most promising basis for the rational design of inhibitors. In the context of an ANR funded research

project involving several research groups in our institute, whose focus is the characterization of PKSs involved in *Mtb* virulence, AT domains from PpsA, Mas, and Pks13 have also been characterized. Preliminary results concerning the search for inhibitors of the AT domain of Pks13 indicate that several molecules are able to bind this domain and lead to the inhibition of its activity. The results of this project will certainly help the design of inhibitors of the AT domains of the PpsA, PpsC, and Mas enzymes. Furthermore, the recent development of the PICT platform ("*Plateforme Intégrée de Criblage de Toulouse*", or Integrated Screening Platform of Toulouse) will greatly facilitate the identification, development, and characterization of inhibitor compounds.

Inhibiting PptT would also be of great therapeutic value. Indeed, the study performed by our collaborators on the conditional expression of PptT has shown the importance of this enzyme for replication of the tubercle bacillus in acute and chronic phases of mycobacterial infection. Besides, PptT has the advantage of certainly being "druggable", since several submicromolar inhibitors have already been found for one of its closest homologue, *B. subtilis* Sfp. In an attempt to search for inhibitors of PptT, a miniaturized assay suitable for high-throughput screening has been developed. In collaboration with the pharmaceutical company Sanofi, this approach has already led to the identification of several hit compounds which inhibit PptT activity. In the near future, the binding of these molecules onto PptT will be characterized by X-ray crystallography using the structure of PptT we have recently solved. The structures of the resulting complexes will provide a rational basis for the design of new molecules with enhanced efficiency.

IV. REFERENCES

[1] J. N. Collie and W. S. Myers, “The formation of orcinol and other condensation products from dehydroacetic acid,” *Journal of Chemical Society*, no. 63, p. 122, 1893.

[2] J. N. Collie, “Derivatives of the multiple keten group,” *Journal of Chemical Society*, vol. 91, p. 1806, 1907.

[3] A. Birch, R. Massy-Westropp, and C. Moye, “Studies in relation to biosynthesis. VII. 2-Hydroxy-6-methylbenzoic acid in Penicillium griseofulvum Dierckx,” *Australian Journal of Chemistry*, vol. 8, p. 539, 1955.

[4] D. E. Cane and C. C. Yang, “Biosynthetic origin of the carbon skeleton and oxygen atoms of nargenicin A1,” *Journal of the American Chemical Society*, vol. 106, no. 3, pp. 784–787, Feb. 1984.

[5] D. E. Cane, P. C. Prabhakaran, W. Tan, and W. R. Ott, “Macrolide biosynthesis. 6 Mechanism of polyketide chain elongation,” *Tetrahedron Letters*, vol. 32, no. 40, pp. 5457–5460, Sep. 1991.

[6] X. Xie and Y. Tang, “Efficient synthesis of simvastatin by use of whole-cell biocatalysis.,” *Applied and environmental microbiology*, vol. 73, no. 7, pp. 2054–60, Apr. 2007.

[7] J. Cortes, K. E. Wiesmann, G. a Roberts, M. J. Brown, J. Staunton, and P. F. Leadlay, “Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage.,” *Science (New York, N.Y.)*, vol. 268, no. 5216, pp. 1487–9, Jun. 1995.

[8] M. Oliynyk, M. J. Brown, J. Cortés, J. Staunton, and P. F. Leadlay, “A hybrid modular polyketide synthase obtained by domain swapping.,” *Chemistry & biology*, vol. 3, no. 10, pp. 833–9, Oct. 1996.

[9] S. Gaisser, L. Kellenberger, A. L. Kaja, A. J. Weston, R. E. Lill, G. Wirtz, S. G. Kendrew, L. Low, R. M. Sheridan, B. Wilkinson, I. S. Galloway, K. Stutzman-Engwall, H. a McArthur, J. Staunton, and P. F. Leadlay, “Direct production of ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of Streptomyces avermitilis ATCC31272.,” *Organic & biomolecular chemistry*, vol. 1, no. 16, pp. 2840–7, Aug. 2003.

[10] K. Patel, M. Piagentini, A. Rascher, Z.-Q. Tian, G. O. Buchanan, R. Regentin, Z. Hu, C. R. Hutchinson, and R. McDaniel, “Engineered biosynthesis of geldanamycin analogs for Hsp90 inhibition.,” *Chemistry & biology*, vol. 11, no. 12, pp. 1625–33, Dec. 2004.

[11] C. M. Starks, Y. Zhou, F. Liu, and P. J. Licari, “Isolation and characterization of new epothilone analogues from recombinant Myxococcus xanthus fermentations.,” *Journal of natural products*, vol. 66, no. 10, pp. 1313–7, Oct. 2003.

[12] H. G. Menzella, R. Reid, J. R. Carney, S. S. Chandran, S. J. Reisinger, K. G. Patel, D. a Hopwood, and D. V Santi, “Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes.,” *Nature biotechnology*, vol. 23, no. 9, pp. 1171–6, Sep. 2005.

[13] M. Sosio, F. Giusino, C. Cappellano, E. Bossi, a M. Puglia, and S. Donadio, “Artificial chromosomes for antibiotic-producing actinomycetes.,” *Nature biotechnology*, vol. 18, no. 3, pp. 343–5, Mar. 2000.

[14] P. Rugbjerg, M. Naesby, U. H. Mortensen, and R. J. Frandsen, “Reconstruction of the biosynthetic pathway for the core fungal polyketide scaffold rubrofusarin in Saccharomyces cerevisiae.,” *Microbial cell factories*, vol. 12, no. 1, p. 31, Apr. 2013.

[15] H. S. Raper, “The condensation of acetaldehyde and its relation to the biochemical synthesis of fatty acids,” *Journal of Chemical Society*, vol. 91, p. 1831, 1907.

[16] R. S. Flugel, “Holo-(Acyl Carrier Protein) Synthase and Phosphopantetheinyl Transfer in Escherichia coli,” *Journal of Biological Chemistry*, vol. 275, no. 2, pp. 959–968, Jan. 2000.

[17] R. Lambalot, A. Gehring, and R. Flugel, “A new enzyme superfamily—the phosphopantetheinyl transferases,” *Chemistry & biology*, 1996.

[18] C. T. Walsh, H. C. P. Synthase, and R. H. Lambalot, “Cloning, Overproduction, and Characterization of the Escherichia coli Holo-acyl Carrier Protein Synthase,” *Journal of Biological Chemistry*, vol. 270, no. 42, pp. 24658–24661, Oct. 1995.

[19] J. K. Hiltunen, M. S. Schonauer, K. J. Autio, T. M. Mittelmeier, A. J. Kastaniotis, and C. L. Dieckmann, “Mitochondrial fatty acid synthesis type II: more than just fatty acids.,” *The Journal of biological chemistry*, vol. 284, no. 14, pp. 9011–5, Apr. 2009.

[20] D. Yu, F. Xu, J. Zeng, and J. Zhan, “Type III polyketide synthases in natural product biosynthesis.,” *IUBMB life*, vol. 64, no. 4, pp. 285–95, Apr. 2012.

[21] C. Hertweck, A. Luzhetskyy, Y. Rebets, and A. Bechthold, “Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork.,” *Natural product reports*, vol. 24, no. 1, pp. 162–90, Feb. 2007.

[22] B. Shen, “Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms,” *Current Opinion in Chemical Biology*, vol. 7, no. 2, pp. 285–295, Apr. 2003.

[23] J. F. Aparicio, I. Molnár, T. Schwecke, a König, S. F. Haydock, L. E. Khaw, J. Staunton, and P. F. Leadlay, “Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: analysis of the enzymatic domains in the modular polyketide synthase.,” *Gene*, vol. 169, no. 1, pp. 9–16, Feb. 1996.

[24] L. Tang, Y. J. Yoon, C. Y. Choi, and C. R. Hutchinson, “Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by Amycolatopsis mediterranei.,” *Gene*, vol. 216, no. 2, pp. 255–65, Aug. 1998.

[25] R. Kong, X. Liu, C. Su, C. Ma, R. Qiu, and L. Tang, “Elucidation of the biosynthetic gene cluster and the post-PKS modification mechanism for fostriecin in Streptomyces pulveraceus.,” *Chemistry & biology*, vol. 20, no. 1, pp. 45–54, Jan. 2013.

[26] J. Piel, “A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 22, pp. 14002–7, Oct. 2002.

[27] Y.-Q. Cheng, G.-L. Tang, and B. Shen, “Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 6, pp. 3149–54, Mar. 2003.

[28] S. L. Ward, Z. Hu, A. Schirmer, R. Reid, W. P. Revill, C. D. Reeves, O. V Petrakovsky, S. D. Dong, and L. Katz, “Chalcomycin biosynthesis gene cluster from Streptomyces bikiniensis: novel features of an unusual ketolide produced through expression of the chm polyketide synthase in Streptomyces fradiae.,” *Antimicrobial agents and chemotherapy*, vol. 48, no. 12, pp. 4703–12, Dec. 2004.

[29] N. Palaniappan, B. S. Kim, Y. Sekiyama, H. Osada, and K. a Reynolds, “Enhancement and selective production of phoslactomycin B, a protein phosphatase IIa inhibitor, through identification and engineering of the corresponding biosynthetic gene cluster.,” *The Journal of biological chemistry*, vol. 278, no. 37, pp. 35552–7, Sep. 2003.

[30] J. Ligon, S. Hill, J. Beck, R. Zirkle, I. Molnár, J. Zawodny, S. Money, and T. Schupp, “Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from Sorangium cellulosum So ce26.,” *Gene*, vol. 285, no. 1–2, pp. 257–67, Feb. 2002.

[31] B. Julien, Z.-Q. Tian, R. Reid, and C. D. Reeves, “Analysis of the ambruticin and jerangolid gene clusters of Sorangium cellulosum reveals unusual mechanisms of polyketide biosynthesis.,” *Chemistry & biology*, vol. 13, no. 12, pp. 1277–86, Dec. 2006.

[32] R. a Butcher, F. C. Schroeder, M. a Fischbach, P. D. Straight, R. Kolter, C. T. Walsh, and J. Clardy, “The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 5, pp. 1506–9, Jan. 2007.

[33] N. Brendel, L. P. Partida-Martinez, K. Scherlach, and C. Hertweck, “A cryptic PKS-NRPS gene locus in the plant commensal Pseudomonas fluorescens Pf-5

codes for the biosynthesis of an antimitotic rhizoxin complex.," *Organic & biomolecular chemistry*, vol. 5, no. 14, pp. 2211–3, Jul. 2007.

[34] D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik, and R. Müller, "Stereochemical determination and complex biosynthetic assembly of etnangien, a highly potent RNA polymerase inhibitor from the myxobacterium Sorangium cellulosum.," *Journal of the American Chemical Society*, vol. 130, no. 43, pp. 14234–43, Oct. 2008.

[35] R. Carvalho, R. Reid, N. Viswanathan, H. Gramajo, and B. Julien, "The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation.," *Gene*, vol. 359, pp. 91–8, Oct. 2005.

[36] N. Gaitatzis, B. Silakowski, B. Kunze, G. Nordsiek, H. Blöcker, G. Höfle, and R. Müller, "The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase.," *The Journal of biological chemistry*, vol. 277, no. 15, pp. 13082–90, Apr. 2002.

[37] C. Olano, B. Wilkinson, C. Sánchez, S. J. Moss, R. Sheridan, V. Math, A. J. Weston, A. F. Braña, C. J. Martin, M. Oliynyk, C. Méndez, P. F. Leadlay, and J. A. Salas, "Biosynthesis of the angiogenesis inhibitor borrelidin by Streptomyces parvulus Tü4055: cluster analysis and assignment of functions.," *Chemistry & biology*, vol. 11, no. 1, pp. 87–97, Jan. 2004.

[38] A. S. Eustáquio, R. P. McGlinchey, Y. Liu, C. Hazzard, L. L. Beer, G. Florova, M. M. Alhamadsheh, A. Lechner, A. J. Kale, Y. Kobayashi, K. a Reynolds, and B. S. Moore, "Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 30, pp. 12295–300, Jul. 2009.

[39] S. Rachid, L. Huo, J. Herrmann, M. Stadler, B. Köpcke, J. Bitzer, and R. Müller, "Mining the cinnabaramide biosynthetic pathway to generate novel proteasome inhibitors.," *Chembiochem : a European journal of chemical biology*, vol. 12, no. 6, pp. 922–31, Apr. 2011.

[40] Y. Tang, C.-Y. Kim, I. I. Mathews, D. E. Cane, and C. Khosla, "The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 30, pp. 11124–9, Jul. 2006.

[41] Y. Tang, A. A. Y. Chen, C. C.-Y. Kim, D. D. E. Cane, C. Khosla, and A. Manuscript, "Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase.," *Chemistry & biology*, vol. 14, no. 8, pp. 931–43, Aug. 2007.

[42] G. Pappenberger, J. Benz, B. Gsell, M. Hennig, A. Ruf, M. Stihle, R. Thoma, and M. G. Rudolph, "Structure of the human fatty acid synthase KS-MAT didomain

as a framework for inhibitor design.," *Journal of molecular biology*, vol. 397, no. 2, pp. 508–19, Mar. 2010.

[43] A. Witkowski, A. K. Joshi, and S. Smith, "Characterization of the beta-carbon processing reactions of the mammalian cytosolic fatty acid synthase: role of the central core.," *Biochemistry*, vol. 43, no. 32, pp. 10458–66, Aug. 2004.

[44] A. Witkowski, A. A. K. Joshi, and S. Smith, "Mechanism of the beta-ketoacyl synthase reaction catalyzed by the animal fatty acid synthase.," *Biochemistry*, vol. 41, no. 35, pp. 10877–87, Sep. 2002.

[45] C. Khosla, R. S. Gokhale, J. R. Jacobsen, and D. E. Cane, "Tolerance and specificity of polyketide synthases.," *Annual review of biochemistry*, vol. 68, pp. 219–53, Jan. 1999.

[46] T. Maier, M. Leibundgut, and N. Ban, "The crystal structure of a mammalian fatty acid synthase.," *Science (New York, N.Y.)*, vol. 321, no. 5894, pp. 1315–22, Sep. 2008.

[47] S. Kapur, B. Lowry, S. Yuzawa, S. Kenthirapalan, A. Y. Chen, D. E. Cane, and C. Khosla, "Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 11, pp. 4110–5, Mar. 2012.

[48] J. Wang, S. M. Soisson, K. Young, W. Shoop, S. Kodali, A. Galgoci, R. Painter, G. Parthasarathy, Y. S. Tang, R. Cummings, S. Ha, K. Dorso, M. Motyl, H. Jayasuriya, J. Ondeyka, K. Herath, C. Zhang, L. Hernandez, J. Allocco, A. Basilio, J. R. Tormo, O. Genilloud, F. Vicente, F. Pelaez, L. Colwell, S. H. Lee, B. Michael, T. Felcetto, C. Gill, L. L. Silver, J. D. Hermes, K. Bartizal, J. Barrett, D. Schmatz, J. W. Becker, D. Cully, and S. B. Singh, "Platensimycin is a selective FabF inhibitor with potent antibiotic properties.," *Nature*, vol. 441, no. 7091, pp. 358–61, May 2006.

[49] M. Moche, "Structure of the Complex between the Antibiotic Cerulenin and Its Target, beta -Ketoacyl-Acyl Carrier Protein Synthase," *Journal of Biological Chemistry*, vol. 274, no. 10, pp. 6031–6034, Mar. 1999.

[50] a C. Price, K. H. Choi, R. J. Heath, Z. Li, S. W. White, and C. O. Rock, "Inhibition of beta-ketoacyl-acyl carrier protein synthases by thiolactomycin and cerulenin. Structure and mechanism.," *The Journal of biological chemistry*, vol. 276, no. 9, pp. 6551–9, Mar. 2001.

[51] L. Serre, E. C. Verbree, Z. Dauter, A. R. Stuitje, and Z. S. Derewenda, "The Escherichia coli malonyl-CoA:acyl carrier protein transacylase at 1.5-A resolution. Crystal structure of a fatty acid synthase component.," *The Journal of biological chemistry*, vol. 270, no. 22, pp. 12961–4, Jun. 1995.

[52] A. T. Keatinge-Clay, A. a Shelat, D. F. Savage, S. C. Tsai, L. J. W. Miercke, J. D. O'Connell, C. Khosla, and R. M. Stroud, "Catalysis, specificity, and ACP docking

site of Streptomyces coelicolor malonyl-CoA:ACP transacylase.," *Structure (London, England : 1993)*, vol. 11, no. 2, pp. 147–54, Feb. 2003.

[53] C. Oefner, H. Schulz, A. D'Arcy, and G. E. Dale, "Mapping the active site of Escherichia coli malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography.," *Acta crystallographica. Section D, Biological crystallography*, vol. 62, no. Pt 6, pp. 613–8, Jun. 2006.

[54] V. Alekseyev, C. Liu, and D. Cane, "Solution structure and proposed domain–domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase," *Protein science*, pp. 2093–2107, 2007.

[55] A. Busche, D. Gottstein, C. Hein, N. Ripin, I. Pader, P. Tufar, E. B. Eisman, L. Gu, C. T. Walsh, D. H. Sherman, F. Löhr, P. Güntert, and V. Dötsch, "Characterization of molecular interactions between ACP and halogenase domains in the Curacin A polyketide synthase.," *ACS chemical biology*, vol. 7, no. 2, pp. 378–86, Feb. 2012.

[56] A. Roujeinikova, W. J. Simon, J. Gilroy, D. W. Rice, J. B. Rafferty, and A. R. Slabas, "Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates.," *Journal of molecular biology*, vol. 365, no. 1, pp. 135–45, Jan. 2007.

[57] E. Płoskoń, C. J. Arthur, S. E. Evans, C. Williams, J. Crosby, T. J. Simpson, and M. P. Crump, "A mammalian type I fatty acid synthase acyl carrier protein domain does not sequester acyl chains.," *The Journal of biological chemistry*, vol. 283, no. 1, pp. 518–28, Jan. 2008.

[58] S. Kapur, A. Y. Chen, D. E. Cane, and C. Khosla, "Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 51, pp. 22066–71, Dec. 2010.

[59] S. Anand and D. Mohanty, "Modeling holo-ACP:DH and holo-ACP:KR complexes of modular polyketide synthases: a docking and molecular dynamics study.," *BMC structural biology*, vol. 12, no. 1, p. 10, Jan. 2012.

[60] A. Keatinge-clay, "Crystal structure of the erythromycin polyketide synthase dehydratase.," *Journal of molecular biology*, vol. 384, no. 4, pp. 941–53, Dec. 2008.

[61] J. Zheng, C. a Taylor, S. K. Piasecki, and A. T. Keatinge-Clay, "Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase.," *Structure (London, England : 1993)*, vol. 18, no. 8, pp. 913–22, Aug. 2010.

[62] J. Zheng, S. K. Piasecki, and A. T. Keatinge-Clay, "Structural Studies of an A2-Type Modular Polyketide Synthase Ketoreductase Reveal Features Controlling α-Substituent Stereochemistry.," *ACS chemical biology*, Jun. 2013.

[63] A. T. Keatinge-Clay, “A tylosin ketoreductase reveals how chirality is determined in polyketides.,” *Chemistry & biology*, vol. 14, no. 8, pp. 898–908, Aug. 2007.

[64] A. T. Keatinge-Clay and R. M. Stroud, “The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases.,” *Structure (London, England : 1993)*, vol. 14, no. 4, pp. 737–48, Apr. 2006.

[65] H. Zhou, Z. Gao, K. Qiao, J. Wang, J. C. Vederas, and Y. Tang, “A fungal ketoreductase domain that displays substrate-dependent stereospecificity.,” *Nature chemical biology*, vol. 8, no. 4, pp. 331–3, Apr. 2012.

[66] D. H. Kwan, M. Tosin, N. Schläger, F. Schulz, and P. F. Leadlay, “Insights into the stereospecificity of ketoreduction in a modular polyketide synthase.,” *Organic & biomolecular chemistry*, vol. 9, no. 7, pp. 2053–6, Apr. 2011.

[67] D. D. L. Akey, J. J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick, and J. L. Smith, “Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway,” *Structure*, vol. 18, no. 1, pp. 94–105, Jan. 2010.

[68] M. Leesong, B. S. Henderson, J. R. Gillig, J. M. Schwab, and J. L. Smith, “Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site.,” *Structure (London, England : 1993)*, vol. 4, no. 3, pp. 253–64, Mar. 1996.

[69] B. Kusebauch, B. Busch, K. Scherlach, M. Roth, and C. Hertweck, “Functionally distinct modules operate two consecutive alpha,beta-->beta,gamma double-bond shifts in the rhizoxin polyketide assembly line.,” *Angewandte Chemie (International ed. in English)*, vol. 49, no. 8, pp. 1460–4, Feb. 2010.

[70] E. Nordling, H. Jörnvall, and B. Persson, “Medium-chain dehydrogenases/reductases (MDR),” *European Journal of Biochemistry*, vol. 269, no. 17, pp. 4267–4276, Sep. 2002.

[71] J. Zheng, D. C. Gay, B. Demeler, M. a White, and A. T. Keatinge-Clay, “Divergence of multimodular polyketide synthases revealed by a didomain structure.,” *Nature chemical biology*, vol. 8, no. 7, pp. 615–21, Jul. 2012.

[72] B. Ames and C. Nguyen, “Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis,” *Proceedings of the National Academy of Sciences of the United States of America*, 2012.

[73] J. Staunton, P. Caffrey, and J. Aparicio, “Evidence for a double-helical structure for modular polyketide synthases,” *Nature Structural & …*, 1996.

[74] J. F. Aparicio, P. Caffrey, a F. Marsden, J. Staunton, and P. F. Leadlay, “Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase.,” *The Journal of biological chemistry*, vol. 269, no. 11, pp. 8524–8, Mar. 1994.

[75] I. Fujii, N. Yoshida, S. Shimomaki, H. Oikawa, and Y. Ebizuka, “An iterative type I polyketide synthase PKSN catalyzes synthesis of the decaketide alternapyrone with regio-specific octa-methylation.,” *Chemistry & biology*, vol. 12, no. 12, pp. 1301–9, Dec. 2005.

[76] D. Edwards, B. Marquez, and L. Nogle, “Structure and Biosynthesis of the Jamaicamides, New Mixed Polyketide-Peptide Neurotoxins from the Marine Cyanobacterium Lyngbya majuscula,” *Chemistry & Biology*, vol. 11, pp. 817–833, 2004.

[77] E. Brignole, S. Smith, and F. Asturias, “Conformational flexibility of metazoan fatty acid synthase enables catalysis,” *Nature structural & molecular …*, vol. 16, no. 2, pp. 190–197, 2009.

[78] V. S. Rangan, a K. Joshi, and S. Smith, “Mapping the functional topology of the animal fatty acid synthase by mutant complementation in vitro.,” *Biochemistry*, vol. 40, no. 36, pp. 10792–9, Sep. 2001.

[79] a. Witkowski, “Dibromopropanone Cross-linking of the Phosphopantetheine and Active-site Cysteine Thiols of the Animal Fatty Acid Synthase Can Occur Both Inter- and Intrasubunit. REEVALUATION OF THE SIDE-BY-SIDE, ANTIPARALLEL SUBUNIT MODEL,” *Journal of Biological Chemistry*, vol. 274, no. 17, pp. 11557–11563, Apr. 1999.

[80] C. Kao, G. Luo, and L. Katz, “Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase,” *Journal of the American …*, no. 16, pp. 9105–9106, 1995.

[81] R. S. R. Gokhale, J. Lau, D. D. E. Cane, and C. Khosla, “Functional orientation of the acyltransferase domain in a module of the erythromycin polyketide synthase,” *Biochemistry*, vol. 2960, no. 97, pp. 2524–2528, Feb. 1998.

[82] R. S. Gokhale, “Dissecting and Exploiting Intermodular Communication in Polyketide Synthases,” *Science*, vol. 284, no. 5413, pp. 482–485, Apr. 1999.

[83] F. T. Wong, A. Y. Chen, D. E. Cane, and C. Khosla, “Protein-protein recognition between acyltransferases and acyl carrier proteins in multimodular polyketide synthases.,” *Biochemistry*, vol. 49, no. 1, pp. 95–102, Jan. 2010.

[84] P. Kumar, Q. Li, D. E. Cane, and C. Khosla, “Intermodular communication in modular polyketide synthases: structural and mutational analysis of linker mediated protein-protein recognition.,” *Journal of the American Chemical Society*, vol. 125, no. 14, pp. 4097–102, Apr. 2003.

[85] R. Broadhurst and D. Nietlispach, “The structure of docking domains in modular polyketide synthases,” *Chemistry & biology*, vol. 10, pp. 723–731, 2003.

[86] J. Zheng, C. D. Fage, B. Demeler, D. W. Hoffman, and A. T. Keatinge-Clay, “The Missing Linker: A Dimerization Motif Located within Polyketide Synthase Modules.,” *ACS chemical biology*, Mar. 2013.

[87] J. Copp and B. Neilan, “The phosphopantetheinyl transferase superfamily: phylogenetic analysis and functional implications in cyanobacteria,” *Applied and environmental microbiology*, vol. 72, no. 4, 2006.

[88] L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, and C. T. Walsh, “Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases.,” *Biochemistry*, vol. 37, no. 6, pp. 1585–95, Feb. 1998.

[89] A. K. Joshi, L. Zhang, V. S. Rangan, and S. Smith, “Cloning, expression, and characterization of a human 4’-phosphopantetheinyl transferase with broad substrate specificity.,” *The Journal of biological chemistry*, vol. 278, no. 35, pp. 33142–9, Aug. 2003.

[90] K. D. Parris, L. Lin, a Tam, R. Mathew, J. Hixon, M. Stahl, C. C. Fritz, J. Seehra, and W. S. Somers, “Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites.,” *Structure (London, England : 1993)*, vol. 8, no. 8, pp. 883–95, Aug. 2000.

[91] M. R. Mofid, R. Finking, L. O. Essen, and M. a Marahiel, “Structure-based mutational analysis of the 4’-phosphopantetheinyl transferases Sfp from Bacillus subtilis: carrier protein recognition and reaction mechanism.,” *Biochemistry*, vol. 43, no. 14, pp. 4128–36, Apr. 2004.

[92] G. Bunkoczi, S. Pasta, A. Joshi, X. Wu, K. L. Kavanagh, S. Smith, and U. Oppermann, “Mechanism and substrate recognition of human holo ACP synthase,” *Chemistry & biology*, vol. 14, no. 11, pp. 1243–1253, Nov. 2007.

[93] C. T. Walsh, a M. Gehring, P. H. Weinreb, L. E. Quadri, and R. S. Flugel, “Post-translational modification of polyketide and nonribosomal peptide synthases.,” *Current opinion in chemical biology*, vol. 1, no. 3, pp. 309–15, Oct. 1997.

[94] N. Barekzi, “Genetic characterization of pcpS, encoding the multifunctional phosphopantetheinyl transferase of Pseudomonas aeruginosa,” *Microbiology*, vol. 150, no. 4, pp. 795–803, Apr. 2004.

[95] H. D. Mootz, R. Finking, and M. a Marahiel, “4’-phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis.,” *The Journal of biological chemistry*, vol. 276, no. 40, pp. 37289–98, Oct. 2001.

[96] N. Y. Chirgadze, S. L. Briggs, K. a McAllister, a S. Fischl, and G. Zhao, “Crystal structure of Streptococcus pneumoniae acyl carrier protein synthase: an essential enzyme in bacterial fatty acid biosynthesis.,” *The EMBO journal*, vol. 19, no. 20, pp. 5281–7, Oct. 2000.

[97] K. Reuter, M. R. Mofid, M. a Marahiel, and R. Ficner, “Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4’-phosphopantetheinyl transferase superfamily.,” *The EMBO journal*, vol. 18, no. 23, pp. 6823–31, Dec. 1999.

[98] J. Ku, R. G. Mirmira, L. Liu, and D. V Santi, “Expression of a functional non-ribosomal peptide synthetase module in Escherichia coli by coexpression with a phosphopantetheinyl transferase.,” *Chemistry & biology*, vol. 4, no. 3, pp. 203–7, Mar. 1997.

[99] J. T. Kealey, L. Liu, D. V Santi, M. C. Betlach, and P. J. Barr, “Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 2, pp. 505–9, Jan. 1998.

[100] F. Fichtlscherer, C. Wellein, M. Mittag, and E. Schweizer, “A novel function of yeast fatty acid synthase. Subunit alpha is capable of self-pantetheinylation.,” *European journal of biochemistry / FEBS*, vol. 267, no. 9, pp. 2666–71, May 2000.

[101] S. Jenni, M. Leibundgut, D. Boehringer, C. Frick, B. Mikolásek, and N. Ban, “Structure of fungal fatty acid synthase and implications for iterative substrate shuttling.,” *Science (New York, N.Y.)*, vol. 316, no. 5822, pp. 254–61, Apr. 2007.

[102] K. a McAllister, R. B. Peery, T. I. Meier, a S. Fischl, and G. Zhao, “Biochemical and molecular analyses of the Streptococcus pneumoniae acyl carrier protein synthase, an enzyme essential for fatty acid biosynthesis.,” *The Journal of biological chemistry*, vol. 275, no. 40, pp. 30864–72, Oct. 2000.

[103] C. Chalut, “The nonredundant roles of two 4′-phosphopantetheinyl transferases in vital processes of Mycobacteria,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 22, pp. 8511–8516, 2006.

[104] C. Leblanc, T. Prudhomme, G. Tabouret, A. Ray, S. Burbaud, S. Cabantous, L. Mourey, C. Guilhot, and C. Chalut, “4’-Phosphopantetheinyl transferase PptT, a new drug target required for Mycobacterium tuberculosis growth and persistence in vivo.,” *PLoS pathogens*, vol. 8, no. 12, p. e1003097, Dec. 2012.

[105] S. Donadio, M. Staver, and J. McAlpine, “Modular organization of genes required for complex polyketide biosynthesis,” *Science*, vol. 1, 1991.

[106] M. Oliynyk, M. Samborskyy, J. B. Lester, T. Mironenko, N. Scott, S. Dickens, S. F. Haydock, and P. F. Leadlay, “Complete genome sequence of the

erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338.," *Nature biotechnology*, vol. 25, no. 4, pp. 447–53, Apr. 2007.

[107] G. L. Challis and D. a Hopwood, "Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by Streptomyces species.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100 Suppl , pp. 14555–61, Nov. 2003.

[108] S. Omura, H. Ikeda, J. Ishikawa, a Hanamoto, C. Takahashi, M. Shinose, Y. Takahashi, H. Horikawa, H. Nakazawa, T. Osonoe, H. Kikuchi, T. Shiba, Y. Sakaki, and M. Hattori, "Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 21, pp. 12215–20, Oct. 2001.

[109] H. Ikeda, J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, and S. Omura, "Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis.," *Nature biotechnology*, vol. 21, no. 5, pp. 526–31, May 2003.

[110] T. M. Daniel, "The history of tuberculosis.," *Respiratory medicine*, vol. 100, no. 11, pp. 1862–70, Nov. 2006.

[111] M. C. Gutierrez, S. Brisse, R. Brosch, M. Fabre, B. Omaïs, M. Marmiesse, P. Supply, and V. Vincent, "Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis.," *PLoS pathogens*, vol. 1, no. 1, p. e5, Sep. 2005.

[112] G. M. Taylor, D. B. Young, and S. A. Mays, "Genotypic analysis of the earliest known prehistoric case of tuberculosis in Britain.," *Journal of clinical microbiology*, vol. 43, no. 5, pp. 2236–40, May 2005.

[113] H. Herzog, Basel, "History of Tuberculosis," *Respiration*, vol. 65, no. 1, pp. 5–15, 1998.

[114] M. a Behr, "BCG--different strains, different vaccines?," *The Lancet infectious diseases*, vol. 2, no. 2, pp. 86–92, Feb. 2002.

[115] B. B. Trunz, P. Fine, and C. Dye, "Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness.," *Lancet*, vol. 367, no. 9517, pp. 1173–80, Apr. 2006.

[116] H. McShane, "Tuberculosis vaccines: beyond bacille Calmette-Guerin.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 366, no. 1579, pp. 2782–9, Oct. 2011.

[117] A. A. Velayati, M. R. Masjedi, P. Farnia, P. Tabarsi, J. Ghanavi, A. H. Ziazarifi, and S. E. Hoffner, "Emergence of new forms of totally drug-resistant

tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran.," *Chest*, vol. 136, no. 2, pp. 420–5, Aug. 2009.

[118] S. Loewenberg, "India reports cases of totally drug-resistant tuberculosis," *The Lancet*, vol. 379, no. 9812, p. 205, Jan. 2012.

[119] D. Yee, C. Valiquette, M. Pelletier, I. Parisien, I. Rocher, and D. Menzies, "Incidence of serious side effects from first-line antituberculosis drugs among patients treated for active tuberculosis.," *American journal of respiratory and critical care medicine*, vol. 167, no. 11, pp. 1472–7, Jun. 2003.

[120] T. Törün, G. Güngör, I. Ozmen, Y. Bölükbaşi, E. Maden, B. Biçakçi, G. Ataç, T. Sevim, and K. Tahaoğlu, "Side effects associated with the treatment of multidrug-resistant tuberculosis.," *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, vol. 9, no. 12, pp. 1373–7, Dec. 2005.

[121] B. Villemagne, C. Crauste, M. Flipo, A. R. Baulard, B. Déprez, and N. Willand, "Tuberculosis: the drug development pipeline at a glance.," *European journal of medicinal chemistry*, vol. 51, pp. 1–16, May 2012.

[122] E. STACKEBRANDT, F. A. RAINEY, and N. L. WARD-RAINEY, "Proposal for a New Hierarchic Classification System, Actinobacteria classis nov.," *International Journal of Systematic Bacteriology*, vol. 47, no. 2, pp. 479–491, Apr. 1997.

[123] E. Tortoli, "The new mycobacteria: an update.," *FEMS immunology and medical microbiology*, vol. 48, no. 2, pp. 159–78, Nov. 2006.

[124] S. Mostowy, J. Inwald, S. Gordon, C. Martin, R. Warren, K. Kremer, D. Cousins, and M. A. Behr, "Revisiting the evolution of Mycobacterium bovis.," *Journal of bacteriology*, vol. 187, no. 18, pp. 6386–95, Sep. 2005.

[125] R. Brosch, S. V Gordon, T. Garnier, K. Eiglmeier, W. Frigui, P. Valenti, S. Dos Santos, S. Duthoy, C. Lacroix, C. Garcia-Pelayo, J. K. Inwald, P. Golby, J. N. Garcia, R. G. Hewinson, M. a Behr, M. a Quail, C. Churcher, B. G. Barrell, J. Parkhill, and S. T. Cole, "Genome plasticity of BCG and impact on vaccine efficacy.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 13, pp. 5596–601, Mar. 2007.

[126] S. Cole, R. Brosch, J. Parkhill, and T. Garnier, "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence," *Nature*, vol. 396, no. NOVEMBER, 1998.

[127] J.-C. Camus, M. J. Pryor, C. Médigue, and S. T. Cole, "Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv.," *Microbiology (Reading, England)*, vol. 148, no. Pt 10, pp. 2967–73, Oct. 2002.

[128] M. Daffé and P. Draper, “The envelope layers of mycobacteria with reference to their pathogenicity,” *Advances in Microbial Physiology*, vol. 29, pp. 131–203, 1997.

[129] B. Zuber, M. Chami, C. Houssin, J. Dubochet, G. Griffiths, and M. Daffé, “Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state.,” *Journal of bacteriology*, vol. 190, no. 16, pp. 5672–80, Aug. 2008.

[130] M. Jackson, G. Stadthagen, and B. Gicquel, “Long-chain multiple methyl-branched fatty acid-containing lipids of Mycobacterium tuberculosis: biosynthesis, transport, regulation and biological activities.,” *Tuberculosis (Edinburgh, Scotland)*, vol. 87, no. 2, pp. 78–86, Mar. 2007.

[131] D. E. Minnikin, L. Kremer, L. G. Dover, and G. S. Besra, “The methyl-branched fortifications of Mycobacterium tuberculosis.,” *Chemistry & biology*, vol. 9, no. 5, pp. 545–53, May 2002.

[132] K. Takayama, L. Wang, and H. L. David, “Effect of isoniazid on the in vivo mycolic acid synthesis, cell growth, and viability of Mycobacterium tuberculosis.,” *Antimicrobial agents and chemotherapy*, vol. 2, no. 1, pp. 29–35, Jul. 1972.

[133] D. M. E. Bowdish, K. Sakamoto, M.-J. Kim, M. Kroos, S. Mukhopadhyay, C. a Leifer, K. Tryggvason, S. Gordon, and D. G. Russell, “MARCO, TLR2, and CD14 are required for macrophage cytokine responses to mycobacterial trehalose dimycolate and Mycobacterium tuberculosis.,” *PLoS pathogens*, vol. 5, no. 6, p. e1000474, Jun. 2009.

[134] D. Portevin, C. De Sousa-D’Auria, C. Houssin, C. Grimaldi, M. Chami, M. Daffé, and C. Guilhot, “A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 1, pp. 314–9, Jan. 2004.

[135] F. Bergeret, S. Gavalda, C. Chalut, W. Malaga, A. Quémard, J.-D. Pedelacq, M. Daffé, C. Guilhot, L. Mourey, and C. Bon, “Biochemical and structural study of the atypical acyltransferase domain from the mycobacterial polyketide synthase Pks13.,” *The Journal of biological chemistry*, vol. 287, no. 40, pp. 33675–90, Sep. 2012.

[136] S. Gavalda, M. Léger, B. van der Rest, A. Stella, F. Bardou, H. Montrozier, C. Chalut, O. Burlet-Schiltz, H. Marrakchi, M. Daffé, and A. Quémard, “The Pks13/FadD32 crosstalk for the biosynthesis of mycolic acids in Mycobacterium tuberculosis.,” *The Journal of biological chemistry*, vol. 284, no. 29, pp. 19255–64, Jul. 2009.

[137] P. Brodin, Y. Poquet, F. Levillain, I. Peguillet, G. Larrouy-Maumus, M. Gilleron, F. Ewann, T. Christophe, D. Fenistein, J. Jang, M.-S. Jang, S.-J. Park, J. Rauzier, J.-P. Carralot, R. Shrimpton, A. Genovesio, J. a Gonzalo-Asensio, G. Puzo, C. Martin, R. Brosch, G. R. Stewart, B. Gicquel, and O. Neyrolles, “High

content phenotypic cell-based visual screen identifies Mycobacterium tuberculosis acyltrehalose-containing glycolipids involved in phagosome remodeling.," *PLoS pathogens*, vol. 6, no. 9, p. e1001100, Jan. 2010.

[138] C. Passemar, A. Arbués, W. Malaga, I. Mercier, F. Moreau, L. Lepourry, O. Neyrolles, C. Guilhot, and C. Astarie-Dequeker, "Multiple deletions in the polyketide synthase gene repertoire of Mycobacterium tuberculosis reveal functional overlap of cell envelope lipids in host-pathogen interactions.," *Cellular microbiology*, vol. 1, Sep. 2013.

[139] J. D. Mougous, C. J. Petzold, R. H. Senaratne, D. H. Lee, D. L. Akey, F. L. Lin, S. E. Munchel, M. R. Pratt, L. W. Riley, J. a Leary, J. M. Berger, and C. R. Bertozzi, "Identification, function and structure of the mycobacterial sulfotransferase that initiates sulfolipid-1 biosynthesis.," *Nature structural & molecular biology*, vol. 11, no. 8, pp. 721–9, Aug. 2004.

[140] T. D. Sirakova, a K. Thirumala, V. S. Dubey, H. Sprecher, and P. E. Kolattukudy, "The Mycobacterium tuberculosis pks2 gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis.," *The Journal of biological chemistry*, vol. 276, no. 20, pp. 16833–9, May 2001.

[141] K. Bhatt, S. S. Gurcha, A. Bhatt, G. S. Besra, and W. R. Jacobs, "Two polyketide-synthase-associated acyltransferases are required for sulfolipid biosynthesis in Mycobacterium tuberculosis.," *Microbiology (Reading, England)*, vol. 153, no. Pt 2, pp. 513–20, Feb. 2007.

[142] P. Kumar, M. W. Schelle, M. Jain, F. L. Lin, C. J. Petzold, M. D. Leavell, J. a Leary, J. S. Cox, and C. R. Bertozzi, "PapA1 and PapA2 are acyltransferases essential for the biosynthesis of the Mycobacterium tuberculosis virulence factor sulfolipid-1.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 27, pp. 11221–6, Jul. 2007.

[143] J. C. Seeliger, C. M. Holsclaw, M. W. Schelle, Z. Botyanszki, S. a Gilmore, S. E. Tully, M. Niederweis, B. F. Cravatt, J. a Leary, and C. R. Bertozzi, "Elucidation and chemical modulation of sulfolipid-1 biosynthesis in Mycobacterium tuberculosis.," *The Journal of biological chemistry*, vol. 287, no. 11, pp. 7990–8000, Mar. 2012.

[144] R. Saavedra, E. Segura, R. Leyva, L. A. Esparza, and L. U. Z. M. Lo, "Mycobacterial Di- O -Acyl-Trehalose Inhibits Mitogen- and Antigen-Induced Proliferation of Murine T Cells In Vitro," vol. 8, no. 6, pp. 1081–1088, 2001.

[145] R. Saavedra, E. Segura, E. P. Tenorio, and L. M. López-Marín, "Mycobacterial trehalose-containing glycolipid with immunomodulatory activity on human CD4+ and CD8+ T-cells.," *Microbes and infection / Institut Pasteur*, vol. 8, no. 2, pp. 533–40, Feb. 2006.

[146] S. a Gilmore, M. W. Schelle, C. M. Holsclaw, C. D. Leigh, M. Jain, J. S. Cox, J. a Leary, and C. R. Bertozzi, "Sulfolipid-1 biosynthesis restricts Mycobacterium

tuberculosis growth in human macrophages.," *ACS chemical biology*, vol. 7, no. 5, pp. 863–70, May 2012.

[147] V. S. Dubey, T. D. Sirakova, and P. E. Kolattukudy, "Disruption of msl3 abolishes the synthesis of mycolipanoic and mycolipenic acids required for polyacyltrehalose synthesis in Mycobacterium tuberculosis H37Rv and causes cell aggregation.," *Molecular microbiology*, vol. 45, no. 5, pp. 1451–9, Sep. 2002.

[148] S. K. Hatzios, M. W. Schelle, C. M. Holsclaw, C. R. Behrens, Z. Botyanszki, F. L. Lin, B. L. Carlson, P. Kumar, J. a Leary, and C. R. Bertozzi, "PapA3 is an acyltransferase required for polyacyltrehalose biosynthesis in Mycobacterium tuberculosis.," *The Journal of biological chemistry*, vol. 284, no. 19, pp. 12745–51, May 2009.

[149] J. J. De Voss, K. Rutter, B. G. Schroeder, H. Su, Y. Zhu, and C. E. Barry, "The salicylate-derived mycobactin siderophores of Mycobacterium tuberculosis are essential for growth in macrophages.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 3, pp. 1252–7, Feb. 2000.

[150] S. C. Andrews, A. K. Robinson, and F. RodrÃguez-QuiÃ±ones, "Bacterial iron homeostasis," *FEMS Microbiology Reviews*, vol. 27, no. 2–3, pp. 215–237, Jun. 2003.

[151] D. Schnappinger, S. Ehrt, M. I. Voskuil, Y. Liu, J. a Mangan, I. M. Monahan, G. Dolganov, B. Efron, P. D. Butcher, C. Nathan, and G. K. Schoolnik, "Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment.," *The Journal of experimental medicine*, vol. 198, no. 5, pp. 693–704, Sep. 2003.

[152] A. M. Talaat, R. Lyons, S. T. Howard, and S. A. Johnston, "The temporal expression profile of Mycobacterium tuberculosis infection in mice.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 13, pp. 4602–7, Mar. 2004.

[153] L. E. Quadri, J. Sello, T. a Keating, P. H. Weinreb, and C. T. Walsh, "Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin.," *Chemistry & biology*, vol. 5, no. 11, pp. 631–45, Nov. 1998.

[154] M. Goren, O. Brokl, and W. Schaefer, "Lipids of putative relevance to virulence in Mycobacterium tuberculosis: phthiocerol dimycocerosate and the attenuation indicator lipid," *Infection and immunity*, vol. 9, no. 1, pp. 150–158, 1974.

[155] E. Kondo and K. Kanai, "A suggested role of a host-parasite lipid complex in mycobacterial infection.," *Japanese journal of medical science & biology*, 1976.

[156] J. S. Cox, B. Chen, M. McNeil, and W. R. Jacobs, "Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice.," *Nature*, vol. 402, no. 6757, pp. 79–83, Nov. 1999.

[157] L. Camacho and D. Ensergueix, “Identification of a virulence gene cluster of Mycobacterium tuberculosis by signature-tagged transposon mutagenesis,” *Molecular …*, vol. 34, pp. 257–267, 1999.

[158] L. R. Camacho, P. Constant, C. Raynaud, M. a Laneelle, J. a Triccas, B. Gicquel, M. Daffe, and C. Guilhot, “Analysis of the phthiocerol dimycocerosate locus of Mycobacterium tuberculosis. Evidence that this lipid is involved in the cell wall permeability barrier.,” *The Journal of biological chemistry*, vol. 276, no. 23, pp. 19845–54, Jun. 2001.

[159] C. Rousseau, N. Winter, E. Pivert, Y. Bordat, O. Neyrolles, P. Avé, M. Huerre, B. Gicquel, and M. Jackson, “Production of phthiocerol dimycocerosates protects Mycobacterium tuberculosis from the cidal activity of reactive nitrogen intermediates produced by macrophages and modulates the early immune response to infection.,” *Cellular microbiology*, vol. 6, no. 3, pp. 277–87, Mar. 2004.

[160] K. Pethe, D. L. Swenson, S. Alonso, J. Anderson, C. Wang, and D. G. Russell, “Isolation of Mycobacterium tuberculosis mutants defective in the arrest of phagosome maturation.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 37, pp. 13642–7, Sep. 2004.

[161] M. a Kirksey, A. D. Tischler, R. Siméone, K. B. Hisert, S. Uplekar, C. Guilhot, and J. D. McKinney, “Spontaneous phthiocerol dimycocerosate-deficient variants of Mycobacterium tuberculosis are susceptible to gamma interferon-mediated immunity.,” *Infection and immunity*, vol. 79, no. 7, pp. 2829–38, Jul. 2011.

[162] C. Astarie-Dequeker, L. Le Guyader, W. Malaga, F.-K. Seaphanh, C. Chalut, A. Lopez, and C. Guilhot, “Phthiocerol dimycocerosates of M. tuberculosis participate in macrophage invasion by inducing changes in the organization of plasma membrane lipids.,” *PLoS pathogens*, vol. 5, no. 2, p. e1000289, Feb. 2009.

[163] M. B. Reed, P. Domenech, C. Manca, H. Su, A. K. Barczak, B. N. Kreiswirth, G. Kaplan, and C. E. Barry, “A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response.,” *Nature*, vol. 431, no. 7004, pp. 84–7, Sep. 2004.

[164] Z. Hasan, I. Zaidi, B. Jamil, M. A. Khan, A. Kanji, and R. Hussain, “Elevated ex vivo monocyte chemotactic protein-1 (CCL2) in pulmonary as compared with extra-pulmonary tuberculosis.,” *BMC immunology*, vol. 6, p. 14, Jan. 2005.

[165] C. Manca, M. B. Reed, S. Freeman, B. Mathema, B. Kreiswirth, C. E. Barry, and G. Kaplan, “Differential monocyte activation underlies strain-specific Mycobacterium tuberculosis pathogenesis.,” *Infection and immunity*, vol. 72, no. 9, pp. 5511–4, Sep. 2004.

[166] L. Tsenova, E. Ellison, R. Harbacheuski, A. L. Moreira, N. Kurepina, M. B. Reed, B. Mathema, C. E. Barry, and G. Kaplan, “Virulence of selected Mycobacterium tuberculosis clinical isolates in the rabbit model of meningitis

is dependent on phenolic glycolipid produced by the bacilli.," *The Journal of infectious diseases*, vol. 192, no. 1, pp. 98–106, Jul. 2005.

[167] O. a Trivedi, P. Arora, A. Vats, M. Z. Ansari, R. Tickoo, V. Sridharan, D. Mohanty, and R. S. Gokhale, "Dissecting the mechanism and assembly of a complex virulence mycobacterial lipid.," *Molecular cell*, vol. 17, no. 5, pp. 631–43, Mar. 2005.

[168] G. Stadthagen, J. Korduláková, R. Griffin, P. Constant, I. Bottová, N. Barilone, B. Gicquel, M. Daffé, and M. Jackson, "p-Hydroxybenzoic acid synthesis in Mycobacterium tuberculosis.," *The Journal of biological chemistry*, vol. 280, no. 49, pp. 40699–706, Dec. 2005.

[169] P. Constant, E. Perez, W. Malaga, M.-A. Lanéelle, O. Saurel, M. Daffé, and C. Guilhot, "Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the Mycobacterium tuberculosis complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift ," *The Journal of biological chemistry*, vol. 277, no. 41, pp. 38148–58, Oct. 2002.

[170] O. Trivedi, P. Arora, V. Sridharan, and R. Tickoo, "Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria," *Nature*, vol. 428, no. March, pp. 4–6, 2004.

[171] K. C. Onwueme, J. a Ferreras, J. Buglino, C. D. Lima, and L. E. N. Quadri, "Mycobacterial polyketide-associated proteins are acyltransferases: proof of principle with Mycobacterium tuberculosis PapA5.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 13, pp. 4608–13, Mar. 2004.

[172] a Rao and a Ranganathan, "Interaction studies on proteins encoded by the phthiocerol dimycocerosate locus of Mycobacterium tuberculosis.," *Molecular genetics and genomics : MGG*, vol. 272, no. 5, pp. 571–9, Dec. 2004.

[173] K. C. Onwueme, C. J. Vos, J. Zurita, C. E. Soll, and L. E. N. Quadri, "Identification of phthiodiolone ketoreductase, an enzyme required for production of mycobacterial diacyl phthiocerol virulence factors.," *Journal of bacteriology*, vol. 187, no. 14, pp. 4760–6, Jul. 2005.

[174] E. Pérez, P. Constant, F. Laval, A. Lemassu, M.-A. Lanéelle, M. Daffé, and C. Guilhot, "Molecular dissection of the role of two methyltransferases in the biosynthesis of phenolglycolipids and phthiocerol dimycoserosate in the Mycobacterium tuberculosis complex.," *The Journal of biological chemistry*, vol. 279, no. 41, pp. 42584–92, Oct. 2004.

[175] M. Jain and J. J. S. Cox, "Interaction between polyketide synthase and transporter suggests coupled synthesis and export of virulence lipid in M. tuberculosis.," *PLoS pathogens*, vol. 1, no. 1, p. e2, Sep. 2005.

[176] G. Sulzenbacher, S. Canaan, Y. Bordat, O. Neyrolles, G. Stadthagen, V. Roig-Zamboni, J. Rauzier, D. Maurin, F. Laval, M. Daffé, C. Cambillau, B. Gicquel, Y. Bourne, and M. Jackson, "LppX is a lipoprotein required for the translocation of phthiocerol dimycocerosates to the surface of Mycobacterium tuberculosis.," *The EMBO journal*, vol. 25, no. 7, pp. 1436–44, Apr. 2006.

[177] M. L. Schaeffer, G. Agnihotri, H. Kallender, P. J. Brennan, and J. T. Lonsdale, "Expression, purification, and characterization of the Mycobacterium tuberculosis acyl carrier protein, AcpM.," *Biochimica et biophysica acta*, vol. 1532, no. 1–2, pp. 67–78, May 2001.

[178] H. C. Wong, G. Liu, Y.-M. Zhang, C. O. Rock, and J. Zheng, "The solution structure of acyl carrier protein from Mycobacterium tuberculosis.," *The Journal of biological chemistry*, vol. 277, no. 18, pp. 15874–80, May 2002.

[179] O. Dym, S. Albeck, Y. Peleg, A. Schwarz, Z. Shakked, Y. Burstein, and O. Zimhony, "Structure-function analysis of the acyl carrier protein synthase (AcpS) from Mycobacterium tuberculosis.," *Journal of molecular biology*, vol. 393, no. 4, pp. 937–50, Nov. 2009.

[180] a K. Azad, T. D. Sirakova, N. D. Fernandes, and P. E. Kolattukudy, "Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria.," *The Journal of biological chemistry*, vol. 272, no. 27, pp. 16741–5, Jul. 1997.

[181] a K. Azad, T. D. Sirakova, L. M. Rogers, and P. E. Kolattukudy, "Targeted replacement of the mycocerosic acid synthase gene in Mycobacterium bovis BCG produces a mutant that lacks mycosides.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 10, pp. 4787–92, May 1996.

[182] S. Cabantous and G. Waldo, "In vivo and in vitro protein solubility assays using split GFP," *Nature Methods*, vol. 3, no. 10, pp. 845–854, 2006.

[183] S. Cabantous, T. C. Terwilliger, and G. S. Waldo, "Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein.," *Nature biotechnology*, vol. 23, no. 1, pp. 102–7, Jan. 2005.

[184] R. M. MCCUNE and R. TOMPSETT, "Fate of Mycobacterium tuberculosis in mouse tissues as determined by the microbial enumeration technique. I. The persistence of drug-susceptible tubercle bacilli in the tissues despite prolonged antimicrobial therapy.," *The Journal of experimental medicine*, vol. 104, no. 5, pp. 737–62, Nov. 1956.

[185] A. Yasgar, T. Foley, A. Jadhav, J. Inglese, M. Burkart, and A. Simeonov, "A Strategy to Discover Inhibitors of Bacillus subtilis Surfactin- type Phosphopantetheinyl Transferase," *Molecular biosytems*, vol. 6, no. 2, pp. 365–375, 2010.

[186] T. Bergfors, "Seeds to crystals," *Journal of Structural Biology*, vol. 142, no. 1, pp. 66–76, Apr. 2003.

[187] J. R. Luft and G. T. DeTitta, “A method to produce microseed stock for use in the crystallization of biological macromolecules.,” *Acta crystallographica. Section D, Biological crystallography*, vol. 55, no. Pt 5, pp. 988–93, May 1999.