Higher-order assemblies of biosynthetic multienzymes

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Florian Xavier Delbart

2024

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

First supervisor: Prof. Dr. Timm Maier
First supervisor: Prof. Dr. Marek Basler
Second supervisor: Prof. Dr. Richard Neher
External expert: Prof. Dr. Jan-Willem Veening

Basel, den 21 Juni 2022

Prof. Dr. Marcel Mayor
Dean of the Faculty

Preamble to my thesis

This thesis contains my conceptual and experimental work done in the research groups of Prof. Dr. Timm Maier and Prof. Dr. Marek Basler in the focal areas of structural biology & biophysics and infection biology, of the Biozentrum, University of Basel, Switzerland.

The thesis contains the work done in both groups. It starts with a general introduction on multienzyme complexes involved in natural product synthesis and metabolism. It elaborates on the enzymes' ability to oligomerise into larger filaments and recent advances made in the fields. The result section consists of research progress made in imaging large multienzymes forming natural products in their native bacterial hosts (chapter 2) and their *in vitro* characterisation (chapter 3). Chapter 4 focusses on a human metabolic multienzyme and elucidates its activation and regulation mechanism. It is written as a manuscript ready for submission. Finally, I discuss the results obtained during this thesis and provide an outlook for future research on these topics.

I Abstract

This work provides a biochemical and structural analysis of two giant enzyme systems relevant for human health. Most enzymes have a single active site and turnover freely diffusing substrates. However, for biosynthetic pathways involving unstable intermediates, multienzymes have evolved that carry multiple enzymatic sites for subsequent reaction steps and use carrier proteins to shuttle covalently tethered substrates between active sites. Two prominent multienzyme systems are bacterial modular polyketide synthases (modPKSs) and human acetyl-CoA carboxylases (ACC). ACC catalyses the carboxylation of acetyl-CoA to malonyl-CoA, which serves as a substrate for fatty acid biosynthesis by fatty acid synthase. ModPKS are prokaryotic proteins that share a common evolutionary ancestor with FAS. They use malonyl-CoA and other acyl-CoA building blocks to generate highly complex bioactive natural products, the polyketides, with great potential as drug candidates.

Chapters 2 and 3, respectively, report the *in vivo* and *in vitro* analysis of the mupirocin PKS from *Pseudomonas fluorescens*. For *in vivo* analysis, I established methods for knocking-out, mutating and fluorescent protein labelling of mupirocin PKS proteins and analysed mupirocin production as a functional read-out. The results demonstrate that different labelled mupirocin PKS proteins localize to the cell pole and maintain product formation, while the determinants for localization remain to be defined. To further analyse protein interactions involved in this process, we aimed at assembling the entire ~1.2 MDa mupirocin PKS from individual proteins recombinantly expressed in insect cells. Due to challenges cloning large and repetitive genes, full reconstruction of the entire mupirocin PKS was not achieved. Biophysical and structural analysis by cryo-electron microscopy revealed that isolated mupirocin PKS proteins are predominantly homodimeric and lack higher order assemblies. The methods and materials derived here still provide an important stepping stone towards *in vivo* studies of modPKS assembly dynamics combined with *in vitro* reconstitution of an entire modPKS.

In chapter 4, I analyse the assembly of recombinant human ACC into active filaments triggered by its principal activator citrate. It has long been known to cause filament formation of the otherwise dimeric ACC, but the mechanistic basis for citrate action remained unknown. Here, we combine structural analysis of ACC filaments by cryo-EM and of ACC fragments by crystallography with biophysical and biochemical characterization of mutants. We demonstrate in atomic detail how citrate binds at the interface of two non-catalytic domains to ACC, aligning domains forming an active state that is conformationally locked by filament formation. Overall, we solve a 60-year puzzle of regulation of a key human enzyme and provide new paradigms for enzyme regulation and citrate sensing.

II Table of Contents

III List of Figures

IV List of Tables

V Abbreviations

AA	Amino acids
ACC	Acetyl-CoA carboxylase
ACP	Acyl carrier protein
ACYL	ATP-citrate lyase
ADP	Adenosine diphosphate
AT	Acetyltransferase
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BC	Biotin carboxylase
BCCP	Biotin carboxyl carrier protein
bp	Basepairs
BRCA	Breast cancer type 1 susceptibility protein
BRCT	Breast cancer type 1 susceptibility protein C Terminus
BT	BC-CT interaction domain
CD	Central domain
CoA	Coenzyme A
cryo-ET	Cryo-electron tomography
cryo-EM	Cryo-electron microscopy
CT	Carboxyl transferase
CV	Column volume
DEBS	6-Deoxyerythronolide B synthase
DH	Dehydratase
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
E. coli	*Escherichia coli*
e.g.	*exempli gratia*
EM	Electron microscopy
ER	Enoyl reductase
FAS	Fatty acid synthase
gDNA	genomic DNA

HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
i.e.	*id est*
iPKS	Iterative polyketide synthase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilo basepairs
kDa	Kilo Dalton
KO	Knockout
KR	Ketoreductase
KS	Ketosynthase
LB	Luria Bertrani broth
LB-agar	Luria Bertrani agar
LD	Linker domain
LINKS	Laterally-INteracting Ketosynthase Sequence
MCS	Multiple cloning site
MDa	Mega Dalton
mFAS	Mammalian fatty acid synthase
mmp	Mupirocin multifunctional protein
mNG	mNeonGreen
modPKS	Modular polyketide synthase
mRNA	messenger RNA
MS	Mass spectrometry
mSc	mScarlet-I
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NMR	Nuclear magnetic resonance
NRPS	Nonribosomal peptide synthetase
OE-PCR	Overlap extension PCR
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol

PKS	Polyketide synthases
Ppant	Phosphopantetheinyl
Ppt	Phosphopantetheinyl transferases
PTM	Post-translational modification
RNA	Ribonucleic acid
RT	Room Temperature
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SOB	Super optimal broth
SOC	Super Optimal broth with Catabolite repression
TCA	Tri-carboxylic acid
TCEP	Tris(2-carboxyethyl)phosphine
TE	Thioesterase
TEM	Transmission electron microscope
TEV	Tobacco etch virus
Tris	Tris(hydroxymethyl)aminomethane
TTC	Triphenyl tetrazolium chloride
WHO	World health organization
WT	Wild-type
β-ME	β-Mercaptoethanol

1 Introduction

Cells are the basis of life and need to sustain chemical reactions to live and grow. Metabolism is the sum of various chemical processes occurring in a living organism. Their main purpose is to convert energy from nutrients into energy enabling them to grow and proliferate, by generating cellular building blocks and eliminating waste. These processes are carried out by specific proteins and their function is well-regulated. One of the crucial functions of proteins is to catalyse chemical reactions that wouldn't spontaneously occur under cellular conditions [1]. They are involved in almost every process, from cell replication, chemical catalysis, molecule transport to a structural part of the cell providing physical structure to our cells and bodies. Correct functioning of proteins is therefore essential for health and dysfunctions lead to diseases.

Proteins consist of linear polymers (peptides) of 20 different amino acids (also termed residues), connected by peptide bonds. Each amino acid has specific chemical properties determined by its sidechain. The sequence of residues is called the primary structure. Based on the chemical properties of the amino acid sidechains, these can form local structural elements of α-helices or β-strands, which is called the secondary structure. The tertiary structure refers to the three-dimensional arrangement of these peptides, and the quaternary structure refers to the association of proteins into complexes [2]. Proteins are synthesized by the ribosome during translation of the amino acid sequence from messenger RNA (mRNA). The growing peptide folds co-translationally or post-translationally into its three-dimensional structure. Regions that fold independently of other parts of the protein are called domains. The primary structure *i.e.* the sequence of amino acids determines the native structure of the protein in its physiological condition, as stated by Anfinsen's dogma [3]. The sequence of the protein is encoded into the DNA sequence that is transcribed into mRNA and used as a template for protein synthesis. The human genome contains approximately 20000–25000 protein coding genes [4] whereas bacteria such as *E. coli* containing only about 6000 protein coding genes [5].

1.1 Enzymes in metabolism

Enzymes catalyse chemical reactions by lowering the required activation energy or stabilizing a chemical transition state [1]. Enzymes bind to their substrate in their active site, which provides the chemical environment to assist in the chemical reaction and convert the substrate into its product. A series of successive chemical reactions, where one enzymatic product is the substrate of another enzyme, is called a metabolic pathway. Controlling this flux of metabolites, substrates

and products of enzymatic reactions is essential for cell survival and growth. Regulation of these pathways happens both transcriptionally and post-translationally. At the transcriptional level, the amount of enzyme produced is regulated by transcription factors. A classic example of this regulation is the lactose operon, where genes encoding proteins involved in lactose catabolism are only expressed in the presence of lactose under low glucose conditions [6]. This is a slow mechanism of regulation does not allow for rapid changes due to time and energy requirements to produce new enzymes. Short-term modulation of enzyme function occurs at the post-translational level. Here, the amount of active enzyme, *i.e.* the amount of enzyme able to catalyse substrate turnover, is regulated by changes protein localisation, or induction of protein degradation. The activation state of enzymes is tightly modulated through binding of small molecules, association with a binding partner, oligomerisation or post-translational modifications (PTMs) [7].

1.2 Multi-subunit enzyme complexes and multienzymes

The multiple reactions of a metabolic pathway are catalysed by enzymes present as either monomers or protein complexes or protein-RNA complexes [8]. An enzyme assembly composed of different subunits contributing to its enzymatic activity is called a multi-subunit enzyme complex (Figure 1.1) [9]. These complexes can be a transient or permanent state [10]. Multienzymes contain multiple catalytic domains on a single polypeptide. In cases where multienzymes form higher oligomers these are called multienzyme complexes (Figure 1.1). Mammalian fatty acid synthase (mFAS), polyketide synthases (PKS) and eukaryotic acetyl-CoA carboxylase (ACC) belong to this category [11-13].

The advantage of multi-subunit enzyme complexes and multienzymes is the short distance between active sites, enabling efficient substrate channelling and increasing the local substrate concentration. However, multi-subunit enzyme complexes require strong binding affinities between subunit interfaces to prevent dissociation of the complex. The advantage of multienzymes is a fixed stoichiometry of enzymatic domains [10]. Additionally, binding interfaces between enzymatic domains were not selected for high affinity, because the presence of multiple interfaces in the same protein provides avidity [14]. However, these lower affinities often result in dissociation of enzymatic domains during structural studies, where single domains or truncations of multienzymes are usually studied in isolation.

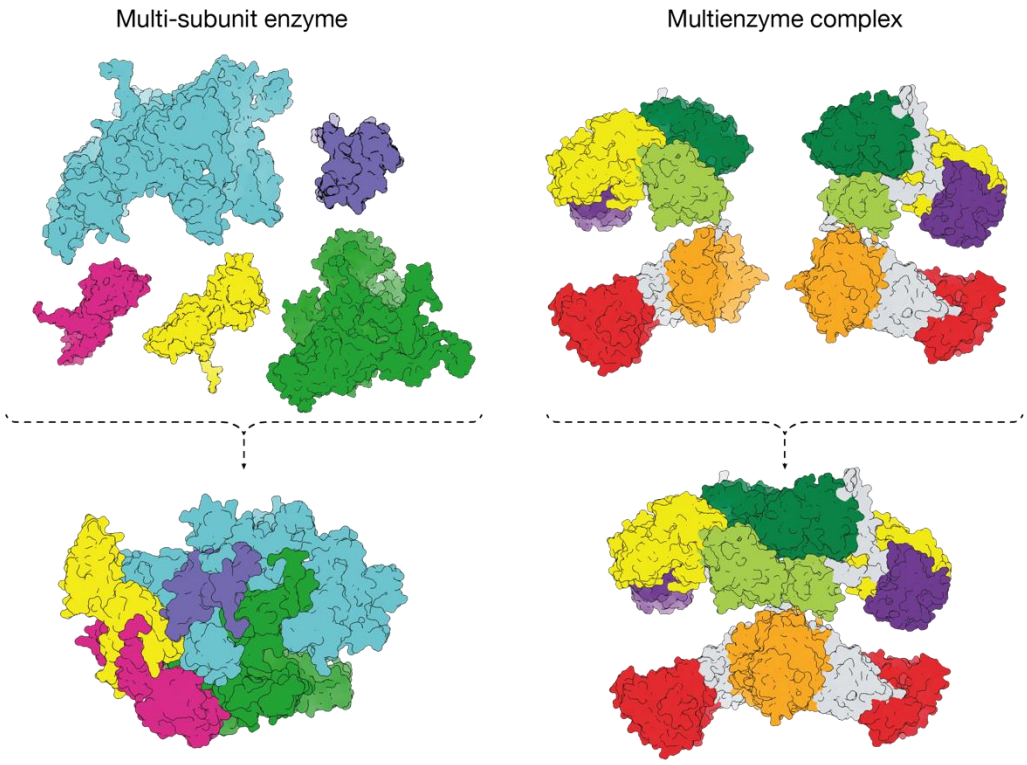

Figure 1.1: Multi-subunit enzyme and multienzyme complex
A multi-subunit enzyme consists of distinct proteins functioning as one complex with catalytic activity. The individual proteins are shown in different colours. Here, RNA-polymerase is shown as an example (PDB: 1HQW). A multienzyme complex consists of multienzymes associating into an enzymatic complex. Mammalian fatty acid synthase is shown as an example and the enzymatic domains are coloured according to their colour code (PDB: 2VZ8, [15]).

1.3 Carrier proteins in metabolic pathways

Some multienzyme complexes, such as PKS, ACC and mFAS, use carrier proteins to shuttle substrates between their enzymatic domains [16]. Others, such as CAD, have the substrate diffusing between the enzymatic domains [17]. Carrier proteins are small domains consisting of 70 to 80 amino acids (AA) and have a molecular weight (MW) between 8 to 10 kDa. Carrier proteins can be present as isolated proteins freely diffusing in the cytosol or form an integral part of the multienzyme. Multiple carrier proteins can be present on the same multienzyme [18]. One of the disadvantages is reduced product turnover because of the decreased diffusion coefficient of larger carrier protein than compared to free-standing molecules. However, this is compensated by linkers, typically 20 to 40 aa in length, connecting the carrier protein to the multienzyme and decreasing its degrees of freedom thereby increasing local substrate concentration. Advantages of carrier proteins are substrate specificity, limit substrate diffusion by restricting the substrate to a specific multienzyme, prevent side reactions of the substrate and improving solubility [16].

All carrier proteins are post-translationally modified with a cofactor from their *apo* form to the the *holo* form. The cofactor enables the carrier proteins to covalently tether to their substrate (Figure 1.2). Three cofactors, phosphopantetheine, lipoyl and biotin, define three classes of enzymes [16]. These cofactors are 15–20 Å long and deliver the substrate deep into the active site.

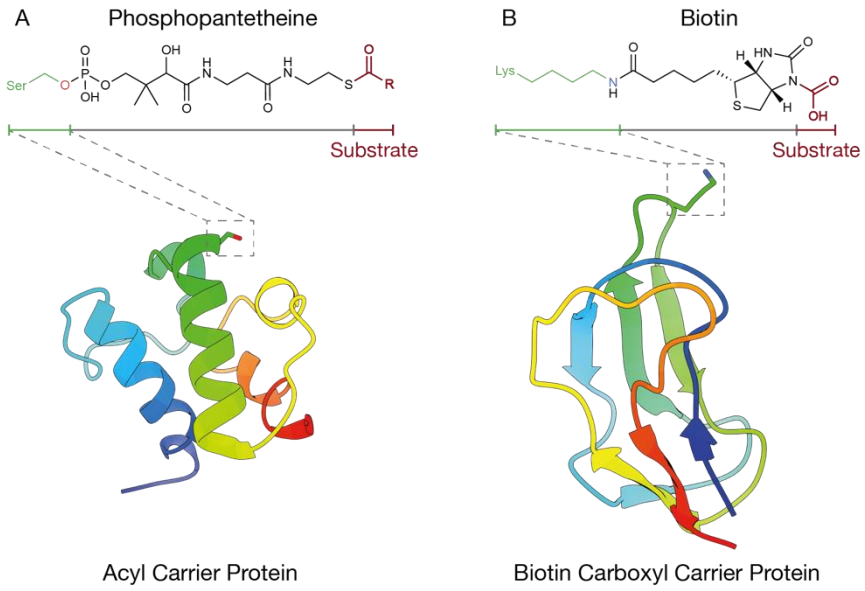

Figure 1.2: Carrier proteins with their cofactors
A: Acyl carrier protein (ACP) with the serine to which the phosphopantetheine cofactor binds shown as stick. (ACP from the mupirocin PKS of *Pseudomonas fluorescens* PDB: 2L22) B: Biotin carboxyl carrier protein (BCCP) and the lysine attached to the biotin cofactor shown as stick. The carboxyl substrate is bound to the biotin moiety. (BCCP from human ACC PDB: 2KCC). Carrier proteins are coloured rainbow from N- (blue) to C-terminus (red).

mFAS as well as PKS are pantetheine-dependent enzymes with a phosphopantetheine co-factor attached to the acyl carrier protein (ACP) [15]. The pyruvate dehydrogenase complex (PDC) has a lipoic acid moiety on its lipoyl cofactor domain. It enables binding of the substrate pyruvate and is in the class of lipoic acid dependent enzymes [19]. ACC belongs to the biotin-dependent carboxylases and has a biotin moiety attached to its biotin carboxyl carrier protein (BCCP), at a highly conserved Met-Lys-Met sequence [20]. The biotin cofactor allows the BCCP to transport carboxyl groups between the different reaction sites [21].

In PKS and FAS, a phosphopantetheinyl transferase (Ppt) transfers the phosphopantetheine (Ppant) moiety from CoA to a conserved serine at the end of helix 2 [22, 23]. This modification can be exploited to load ACPs of recombinantly expressed PKS with reactive derivatives of Ppant [24-26]. The ACP loaded with the reactive warhead, called *crypto*-ACP, then reacts and crosslinks to the active site of the target protein.

1.4 Polyketides

In natural environments organisms are competing for resources and space. Microorganisms have evolved different strategies to compete. One of these approaches is the synthesis of small molecules that inhibit the growth or even kill other competing species [27]. An important family of these molecules are polyketides [28-30]. Their extensive chemical diversity, functionality and bioactivity makes them an attractive source for therapeutics (Figure 1.3) [31-33]. Thus, many polyketides or derivatives thereof are already in clinical use as antibiotics (*e.g.* erythromycin A, rifampicin, mupirocin), cholesterol lowering drugs (*e.g.* statins), immunosuppressiva (Rapamycin), cancer chemotherapeutics (*e.g.* doxorubicin, epothilone), antifungals (amphotericin B) and antiparasitics (avermectin) [34]. Polyketides are a diverse group of secondary metabolites. Despite their diverse structures (Figure 1.1), polyketides share common properties and synthesis. Polyketides are synthesized by polyketide synthases (PKSs) by Claisen condensation of small acyl-building blocks. These are further functionalised, yielding e continuous chain of carbon atoms with substitutions consisting of carbonyl or hydroxyl groups, double bonds and fully saturated carbon-carbon bonds (Figure 1.4). Additionally, many PKS gene clusters contain tailoring enzymes which further modify the product [29].

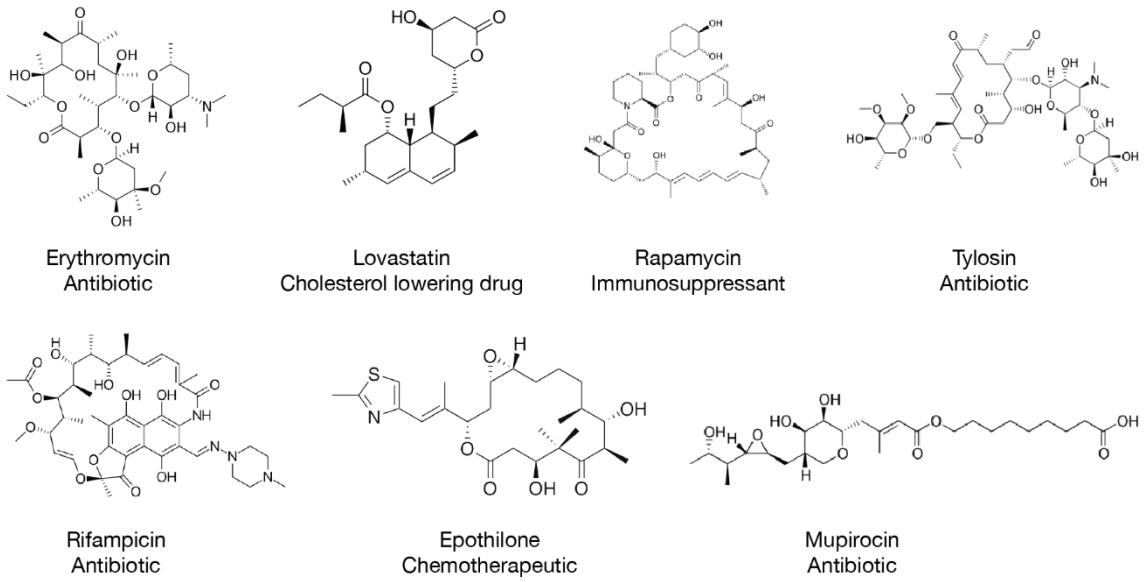

Figure 1.3: Examples of polyketide drugs in clinical use
Polyketides are chemically diverse compounds. Name and therapeutic use of polyketide drugs indicated.

1.4.1 Modular PKS

Polyketides are synthesized by PKS that are classified into three classes. Type I PKS are large multienzymes consisting of several enzymatic domains with defined functions, separated by short

linker regions. Type II PKS produce aromatic compounds by stand-alone enzymatic domains. Type I and II PKS are both ACP dependent. Type III PKS, are ACP independent, performing iteratively condensation reactions [35, 36].

Type I PKS are further divided into iterative (iPKS) and modular PKS (modPKS) [37, 38]. In iPKS, the same set of domains is repeatedly used for substrate extension and modification. modPKS consists of an assembly line composed of multiple modules. Each module catalyses one defined step of chain elongation and modification before the substrate is transferred to the next module by the carrier protein.

Polyketide synthesis starts with loading of an acyl-CoA-derivative (*e.g.* malonyl-CoA) by an acyl-transferase (AT) onto the phosphopantetheine moiety of the ACP. Next, the acyl-CoA derivative is transferred to the active site cysteine of the ketosynthase (KS) domain in a step called chain translocation. The ACP is then again loaded with an acyl-CoA derivative (the extender unit), which is placed in the KS active site. The growing polyketide is formed by subsequent condensation of the acyl-building blocks in the KS active site in a decarboxylative Claisen-condensation, the principal elongation step (Figure 1.4) [39]. The elongation product can then be transferred to optional modifying domains catalysing reductions, elimination of water, methylations or isomerizations. The principal set of enzymes includes a ketoreductase (KR) that reduces the β-carbonyl to a hydroxyl and introduces a stereocenter, a dehydratase (DH) domain forming an α,β-double bond by eliminating water and an enoylreductase (ER) reducing the product to a fully saturated carbon chain. In iPKS, the same cycle is repeated until product release while in modPKS the product is transferred to the downstream module for a following round of elongation and optional modification. The final product is released from the ACP by the thioesterase domain (TE, Figure 1.4) [29]. Due to the directed transfer of the substrate from one module to the next one, modPKS are compared to molecular assembly lines, analogous to an assembly line in a car factory [40, 41]. Typically, two to four modules are present on the same polypeptide, and distinct polypeptides are linked through N- and C-terminal docking domains [42, 43]. The order of modules in modPKS corresponds to the direction of product synthesis, which is called collinearity [44]. This collinearity allows prediction of the polyketide structure based on the PKS sequence inferred from phylogenetic analyses [45-48].

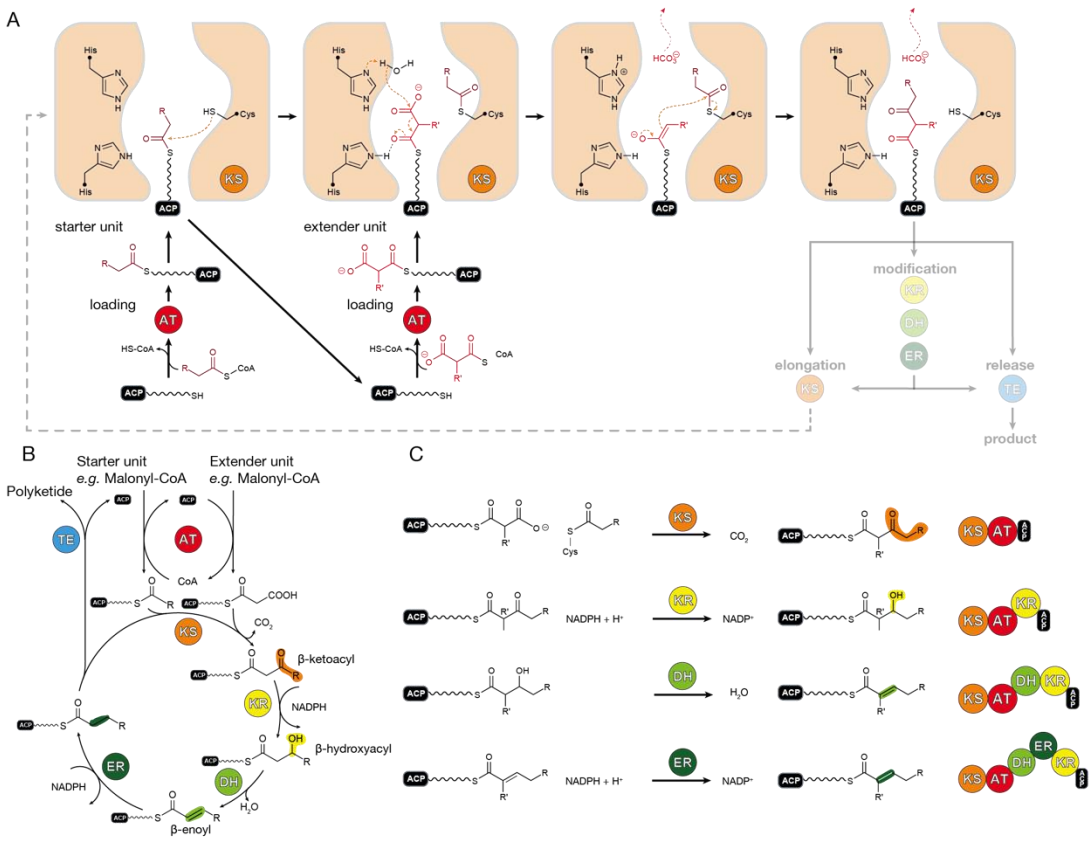

Figure 1.4: Scheme of the condensing and canonical modifying reactions of PKS
A: Scheme of the condensing reaction in the KS active site. Synthesis starts with loading of a starter unit by the AT. Next, the AT loads an extender unit, that is condensed in the KS active site with release of CO_2 as HCO3-. B: Schematic representation of a single cycle of polyketide synthesis and modification. C, most common reactions catalysed by PKS, the product and module composition generating such a product is shown on the right. Reaction scheme modified from [49]

The prototype for modPKS is the 6-deoxyerythronolide B (DEBS) assembly line that was the first discovered modPKS [12]. It's involved in erythromycin synthesis, by producing the macrolide intermediate of erythromycin A. The genes *eryAI-III* encode three assembly line proteins (DEBS1-3), that each contain 2 modules (Figure 1.5) [48]. For the DEBS assembly line, this collinearity even extends to the genome level [48]. However, discoveries of other assembly lines showed that this colinear organization of open reading frames (ORFs) doesn't always apply, as is the case for the rapamycin or mupirocin PKS [50, 51]. Two evolutionary distinct families of modPKS exist based on whether the AT domains are an integral part of the module or encoded as a free-standing domain [37, 38, 52]. The former are referred to as *cis*-AT PKS, while the latter are named *trans*-AT PKS.

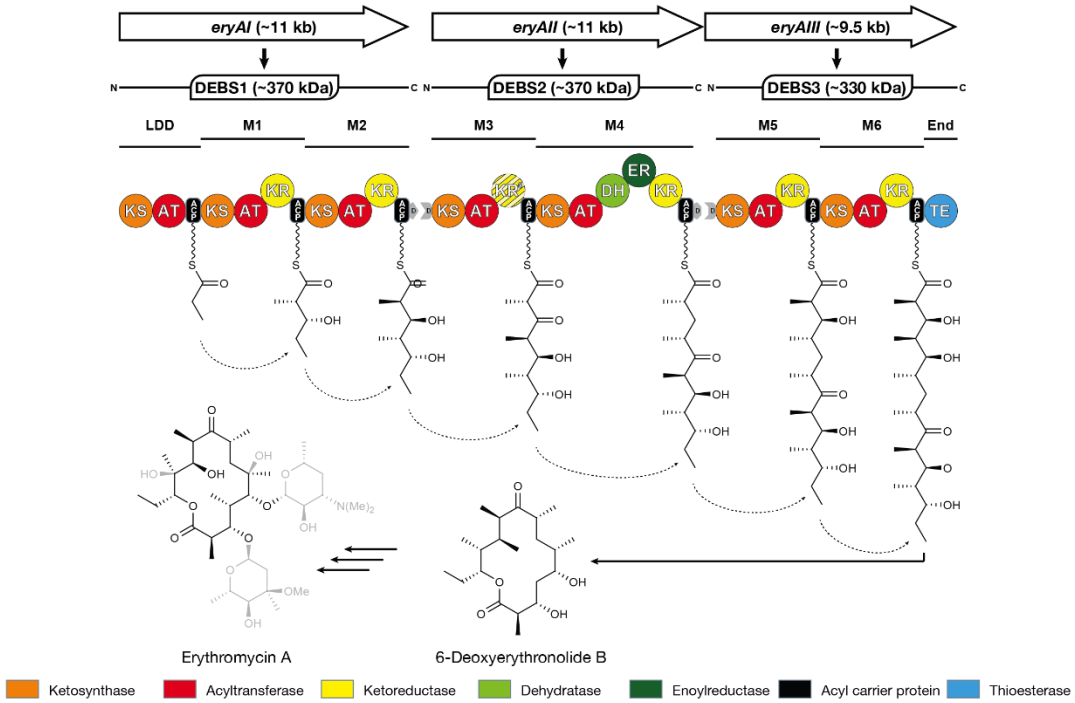

Figure 1.5: Schematic representation of the 6-deoxyerythronolide B (DEBS) *cis*-AT PKS
The colinear organisation of the cluster is shown at the genetic- (arrows at the top), protein- and module level. Protein docks through N- and C-terminal docking domain. Each catalytic domain is represented as a sphere and performs one step in the biosynthesis of 6-deoxyerythronolide B, with substrate transfer to the next one in an assembly line manner. The thioesterase (TE) domain releases 6-deoxyerythronolide B, that is modified by tailoring enzymes to form erythromycin A. The acyltransferase (AT) domain is part of each module, as is the case for *cis*-AT PKS. Scheme modified from [49]. M: module, LDD: Loading didomain.

Because of the sequential nature of polyketide biosynthesis in modPKS there must be directionality in product synthesis. Efficient substrate handover to the next module requires that each downstream KS has higher substrate affinity than the previous KS. Otherwise, assembly line stalling or stuttering would happen [53, 54]. Biochemical experiments confirmed that the downstream KS indeed selects for the right product by binding to it with higher affinity and preventing retrotransfer of the product to the previous module [55]. Evolutionary analysis also confirmed that this directionality is partially encoded in the KS sequences. The downstream KS does not only recognize the substrate, but also the ACP of the previous module. Co-evolution of the KS with the modifying region and ACP of the preceding module is observed. This led to the suggested re-definition of a module, being composed of the modifying region of the upstream module with the KS of the downstream module [56-58]. However, this module definition is not frequently used in literature, therefore we adhere to the traditional module definition.

Another mechanism to control the growth of polyketide intermediates was proposed. The KS acts as a turnstile, that remains closed until the product from the previous catalytic cycle has been

transferred to the next module. Then the KS opens to accommodate binding of the previous ACP with its product [59].

1.4.2 Bioengineering of PKS

Due to their complex chemistry, chemical synthesis of polyketides has been very challenging, requiring multiple steps resulting in low product yields [60-63]. The domain organization of modules and the order of modules determines the synthesized product [44]. This offers tremendous potential for genetic manipulation to engineer PKS and synthesize novel "unnatural" products, with optimised properties [64, 65]. PKS have already been engineered with the goal of increasing product yield, producing novel compounds with therapeutic potential [37, 66, 67] or synthetic biofuels [68, 69]. Using a cut-and-paste approach, several domains [70, 71] or modules [72, 73] were exchanged and active sites were mutated [74]. However, so far most of these PKS re-engineering attempts have had limited success or failed [34]. The main findings were absence of the desired product or only moderate yield.

One of the initial findings is that the KS moiety selects for the correct substrate [44]. Attempts to feed non-natural substrates to the KS result in "stalling" of the assembly line with intermediates of the polyketide being released, called derailing products [53, 54]. To solve this issue, mutations of some of the residues lining the KS active site were introduced enabling them to accept non-natural intermediates [74-76]. Still, lack of structural understanding of the mechanism of substrate recognition renders these attempts challenging.

Another challenge are the protein-protein interactions required for product synthesis. The ACP must interact with all enzymatic domains of that module, before handing over the substrate to the KS of the downstream module. In studies where intact modules have been exchanged, the N- or C-terminal docking domains were retained [72, 73]. These domains dock the different assembly line proteins from N- to C-terminus in the direction of product synthesis, but have weak affinities as shown *in vitro* ($K_d \approx 1–100$ µM) [77-80]. In an approach to reconstitute the DEBS assembly line *in vitro,* DEBS1 containing modules 1 and 2 (Figure 1.5) was truncated between both modules. To reconstitute the complex, docking domains were added at the N- and C-termini of the constructs [81]. Retaining these interactions after swapping modules is important, yet the additional challenge is forcing the ACP interaction with a non-native interface. One of the attempts is the use of synthetic zipper domains, termed SYNZIPs, which form coiled-coil interactions between the preceding and next modules [82]. These would force the ACP to remain in the proximity of the domains required for product synthesis. This approach was proven successful for

modules 5 and 6 of the DEBS assembly line, there remain challenges translating this approach as a general engineering strategy for modPKS. This solution doesn't solve the issue of the ACP having to interact with non-native interfaces and product yield varies between modules used.

Progress in this field is more challenging than initially thought. Main challenges remain impaired protein activity due to mutations, unknown appropriate expression host, limits of traditional cloning methods to engineer genes of this size and reduced product yield. Additionally, a major limitation is the lack of knowledge on higher order assemblies and substrate transfer between and within these large proteins [34, 73, 83, 84].

1.4.3 Structures of polyketide synthase modules

PKS are evolutionary related to fatty acid synthases (FAS), both functionally and structurally [40, 47]. Most insights into the biochemistry of polyketide synthesis are derived from FAS [85]. With advance in genetics in the 1980s, the first PKS gene clusters were identified with the actinorhodin PKS from *Streptomyces coelicolor* [86]. In the 1990s the with DEBS cluster in *Saccharopolyspora erythraea* was characterized [12]. The cluster is involved in synthesis of the macrolide precursor of the antibiotic erythromycin A. The cluster was organized into six repeating units encoding enzymes with FAS-like properties. Each of these units was designated as a module, and two modules are part of the same protein. Based on this observation, a model for polyketide synthesis was proposed. Each module catalyses one step of product extension before transferring the substrate to the next module. Six of these steps are required to form the final polyketide product, that is subsequently released by the TE domain [12, 48]. DEBS module 4 most closely matches the domain organization of mammalian fatty acid synthase 8mFAS), with 23 % sequence identity. The other modules correspond to smaller mFAS-like variants because of different domain organisations in the modifying region [12, 48]. While similar both in catalytic activity and at the sequence level, this sequential organization highlights one of the main differences between modPKS and mFAS. For fatty acid synthesis, malonyl-CoA precursors are iteratively condensed to form palmitic acid [15]. In modPKS, each module typically only performs one step of product extension and modification before substrate handover to the next one [40, 41].

In the first decade of the 2000s, structural insights into PKS domains or multidomains and of mFAS were obtained. Initial low-resolution findings from EM and small-angle X-ray scattering (SAXS) were compared with high resolution crystal structures. Several structures of isolated PKS domains and multidomain fragments have been solved [39, 87-95]. These PKS structures

resemble the general architecture of mammalian fatty acid synthase and confirm that domains are structurally conserved (Figure 1.6) [15, 40].

Insights into the multidomain architecture of PKS were obtained with the crystal structures of a 194 KDa fragment from DEBS module 5 [96] and a year later with module 3 [97]. These models consist of the homodimeric KS-AT didomain and the flanking linkers, both models being very similar. Both the KS and AT active sites are visible and separated by 70-80 Å. This distance is too large to be reached by a static ACP, highlighting the requirement of flexibility and domain rearrangements for catalysis. Throughout polyketide or fatty acid synthesis, the growing chain remains attached to the enzyme during synthesis. During the whole reaction cycle the ACP must change its position to reach every active site, until final product release. Therefore, flexibility and conformational changes are an inherent property of this family of enzymes.

A large N-terminal linker extends from the KS dimer interface, that is absent in mFAS. This N-terminal linker is presumed to have evolved to facilitate interactions with and transfer from the previous module in the assembly line [96]. Later, these regions have been confirmed as being docking domains, responsible for docking the assembly line proteins together [78, 98]. Overall, the domain organisation is consistent with the structure of mFAS [15, 99].

The first structural insights in the overall architecture of a *trans*-AT PKS were obtained in 2013 with the crystal structure of the RhiE module from *Burkholderia rhizoxinica* consisting of a KS and a branching domain [100]. The structure revealed a conserved KS core with the next domain of the module on top of it. In 2014 the cryo-EM structure of module 5 from the pikromycin PKS assembly line was revealed [87]. The model consists of a KS-AT-KR-ACP and was solved by cryo-EM at final resolutions of 7.3-9.5 Å. While resolution was too low for *de novo* model building, the maps allowed rigid body fitting of the homologous KS, AT, KR and ACP domain structures from DEBS. Fitting of the domains revealed several clashes and overlaps. The overall architecture has a horseshoe- or arch-shape and differs considerably with the extended DEBS modules 3 and 5 structures [96, 97] and the X-shaped mFAS model (Figure 1.6) [15]. The authors suggest that refolding of the protein is required to enable this conformation. An alternative entrance to the catalytic site was also suggested. However, this would require dissociation of the post-AT linker from the KS. Yet, crystal structures reveal that this linker is tightly embedded in a hydrophobic core and plays an important role in fixing the relative positions of the AT and KS domains [96]. A hybrid model of the mycocerosic acid-like synthase (MAS) PKS from *Mycobacterium smegmatis*

serves as a prototype for a fully reducing PKS module [39]. The model revealed, by combining crystal structures with small-angle X-ray scattering (SAXS) experiments, that the PKS is present in its extended conformation similar to the structure of mFAS (Figure 1.6).

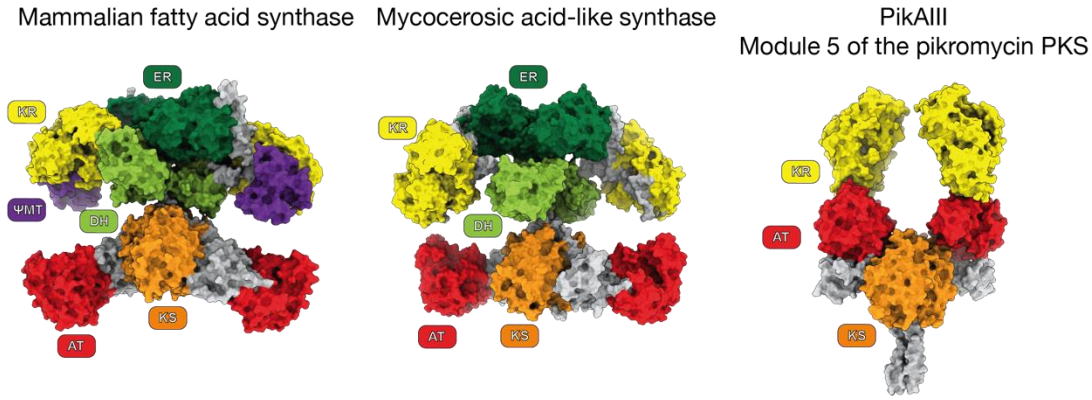

Figure 1.6: Architectures of mFAS and PKS modules
Structures of mFAS [15], MAS [39] and PikAIII [87] show the similarity and divergence in architecture with mFAS, respectively. Enzymatic domains are labelled according to their colour code. Both mFAS and MAS have an X-shaped architecture, while PikAIII is arch-shaped.

1.4.4 Higher order assemblies of PKS

PKSs are dimeric multienzymes forming gigantic molecular assembly lines, with modPKS belong to some of the largest protein complexes known. Assembly lines commonly reach MDa sizes, as is the case for the DEBS PKS that measures 2 MDa [18, 48, 101]. While the structures of enzymatic domains and even modules are already well understood, information on the higher order organisation of PKS modules remains elusive. Each polypeptide can consist of multiple modules and multiple polypeptides interact through N- and C-terminal docking domains to form one giant assembly line. These interactions are considered weak and of transient nature taking into account the average protein concentration in *E. coli* cells of 1 nM – 10 µM [5, 102]. A requirement for assembly line formation *in vivo* would be high local concentrations of the assembly line proteins. For the PksR protein of the bacillaene *trans*-AT PKS assembly line, quantitative western blots indicated the presence of 50 – 150 copies/cell. This corresponds to a concentration of 50 -150 nM inside the cell, relative to *E. coli* [5, 103]. These concentrations are too low for stable association of the assembly line proteins. However, the bacillaene PKS was visualized in *Bacillus subtilis* cells by electron microscopy (EM) with immunogold labelling. A single, dense region with high local concentrations of assembly line proteins corresponding to masses of 10-100 MDa is present at the cell membrane [103]. The clustering of this complex at the membrane leads to a cellular organelle-like factory producing the polyketide product

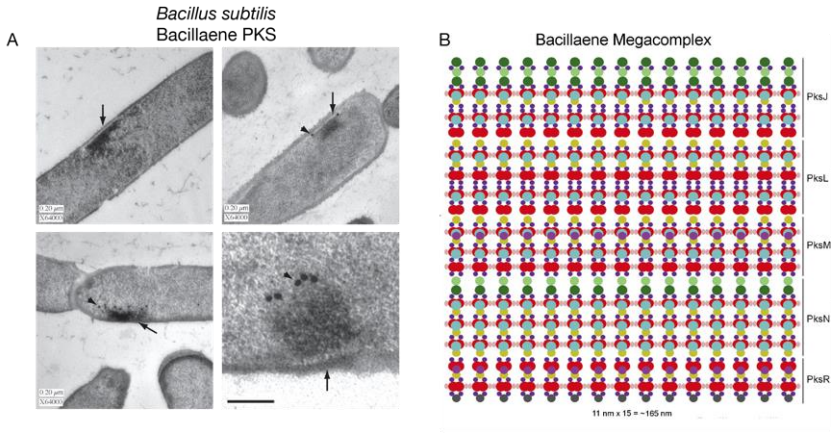

Figure 1.7: Megacomplex formation of the bacillaene *trans*-AT modPKS of *B. subtilis*
A: Electron microscopy images of *B. subtilis* cells reveal a large, dense mass at the membrane. Immunogold labelling confirmed that this is the bacillaene PKS. B: Suggested model of PKS megacomplex formation, where the individual assembly line proteins interact vertically and laterally. Adapted from Straight *et al.* [103] and Gay *et al.* [104] with permission.

The biological advantage of such megacomplex formation and its role in product synthesis is so far unknown. It's hypothesized to sequester *trans*-acting enzymes and to increase the local concentrations of components, limiting diffusion distances. Due to the multiple active sites and interactions required for polyketide synthesis, limiting the diffusion distances is hypothesized to increase efficiency. Its presence at the membrane points towards efficient secretion of the product, potentially coupled to a membrane transporter. Indeed, polyketide products are often secreted from the producer cells through a dedicated transporter [105]. For example, the mycrocystin secondary metabolite cluster encodes a dedicated ABC transporter responsible for product export [106]. Another example where membrane vicinity can be beneficial is in the kalimantacin polyketide cluster from *Pseudomonas fluorescens*. There, an inactive polyketide is released by the assembly line and upon product export in the environment it is oxidized by a periplasmic enzyme gaining its toxic activity [107]. In *Mycobacterium ulcerans* the mycolactone polyketide synthase produces mycolic acids, which are part of the mycobacterial cell wall [108]. Confocal microscopy revealed that a minimal trimodular construct of this PKS is present at the cell membrane. Membrane association makes sense in the context of mycolactone synthesis, yet the exact mechanism of membrane association is unknown. Again, association with a transporter is hypothesised. At present, it is unknown if other PKS assembly lines also localize at the membrane and if membrane proximity is a requirement for efficient product synthesis.

Docking domains docking the assembly line proteins together are known, however only these interactions are not sufficient to cluster the proteins together in these megacomplexes. At present, KS domains from both *cis*-AT and *trans*-AT modPKS have been determined [92, 109].

In *cis*-AT PKS the AT domains is connected to the KS through a 100 AA linker, termed linker domain (LD). While in *trans*-AT PKS the AT domain is free standing, the KS contains the LS without connection to an AT domain. Initially it was hypothesized to be involved in docking the free-standing AT domain [110, 111]. However, later it was found that docking of the AT domain to the KS of *trans*-AT PKS in an equivalent position as observed in *cis*-AT PS would clash with helix bundles of the LD [92]. Various structures of KS-LD domains from *trans*-AT PKS were crystallised making lateral interactions through their LD [104]. Initially, these were assumed to be a crystallisation artefact. However, additional structural information revealed that there is a common tendency of KS from *trans*-AT PKS to laterally stack through homotypic interactions between this region of multiple KSs [104]. The mechanism of PKS clustering was hypothesized to be mediated by lateral sequences flanking the KS domains, termed Laterally INteracting Ketosynthase Sequences (LINKS motif, Figure 1.8). The KSs would zip together laterally and form a two-dimensional arrangement. Based on sequence alignments and bioinformatic analysis, the LINKS-interaction is suggested to be unique for each KS. So far, there is no additional biological evidence for the formation of these *trans*-AT PKS complexes or filaments and their role in product synthesis remains unclear. However, local organisation of the PKS components may be required to increase the efficiency of product synthesis by overcoming slow diffusion rates [112]. Additional experiments combining both *in vivo* imaging and biophysical characterization of the proteins are required to assess the higher-order architecture of *trans*-AT PKS [14]. Only with detailed understanding of the critical determinants for product synthesis, efficient re-engineering of PKS to generate novel compounds is possible.

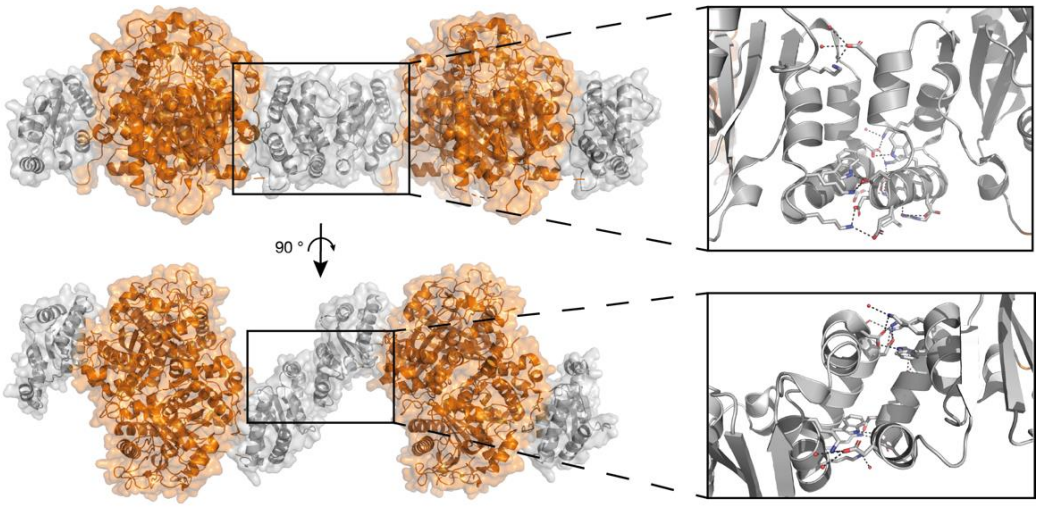

Figure 1.8: LINKS-interaction between two KS domains from *trans*-AT PKS
The KS core (orange) is flanked by linker domains (LD, grey). A three-helix bundle named LINKS-motif in the LD is able to zip KS together by forming homotypic interactions with the LINKS-motif of the neighbouring KS, potentially enabling large PKS complex formation [104]. KS of the bacillaene PKS, PDB: 4NA1, [92].

1.5 Acetyl-Coenzyme A Carboxylase

Acetyl-CoA carboxylase (ACC) is an enzyme present in all domains of life and catalyses the first committed step in fatty acid synthesis. Carboxylation of acetyl-CoA to form malonyl-CoA, that is used by fatty acid synthase for fatty acid synthesis (Figure 1.11).

1.5.1 Human ACC is a carrier protein-based multienzyme

In lower organisms such as archaea and bacteria, ACC is present as a multi-subunit enzyme. In eukaryotes, ACC is a multienzyme (Figure 1.9). Fungal ACC is a ~500 kDa homodimer [113]. In higher eukaryotes such as humans ACC is present in two isoforms, ACC1 and ACC2, sharing ~75 % sequence identity [114]. The main difference between ACC1 (~265 kDa) and ACC2 (~276 kDa) is an N-terminal sequence anchoring ACC2 to the mitochondrial membrane [115]. ACC1 is expressed in lipogenic tissues such as the liver and adipose [116]. Malonyl-CoA produced by ACC1 is directly used for incorporation into fatty acids by fatty acid synthase (FAS). ACC2 is uniquely expressed in oxidative tissues. Initially it was discovered in the rat heart and later also in skeletal muscles, where little to no *de novo* fatty synthesis occur [117, 118]. Malonyl-CoA produced by ACC2 inhibits fatty acid oxidation by inhibition of the mitochondrial carnitine palmitoyl transferase 1 (CPT1, Figure 1.11) [116].

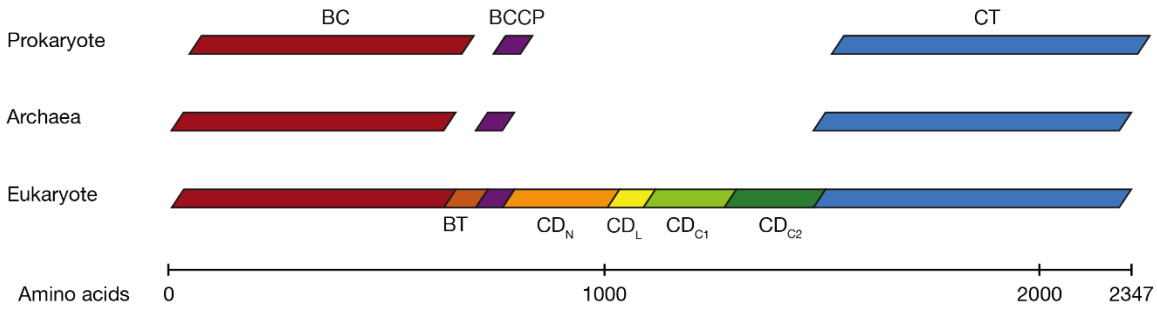

Figure 1.9: Organisation of ACC in the three domains of life
In prokaryotes and archaea, ACC is a multi-subunit enzyme with the biotin carboxyl carrier protein (BCCP) shuttling between the enzymatic biotin carboxylase (BC) and carboxyl transferase (CT) domains. In higher eukaryotes, ACC is a multienzyme with the BT-CT interaction (BT) and central domain (CD) connecting the BC and CT enzymatic domains. The CD is subdivided into the CD N-terminal (CD_N), CD linker (CD_L) and CD C-terminal 1 and 2 (CD_{C1} and CD_{C2}) subdomains.

Carboxylation of acetyl-CoA to malonyl-CoA occurs in two steps (Figure 1.10) [119]. First, the biotin cofactor of the Biotin Carboxyl Carrier Protein (BCCP) is carboxylated by the N-terminal Biotin Carboxylase (BC) domain. During the carboxylation reaction, bicarbonate donates CO_2 and ATP is consumed [120, 121]. Next, the BCCP translocates to the Carboxyl Transferase (CT) domain, where the carboxyl group is transferred to acetyl-CoA producing malonyl-CoA. Next to

these catalytic domains, eukaryotic ACCs also contain a BC-CT (BT) interaction domain and a central domain (CD), connecting the BC and CT domains. While both of these domains have no catalytic activity, they are flexible and serve as a hinge to position the BC and CT domains in the correct orientation for catalysis [122, 123].

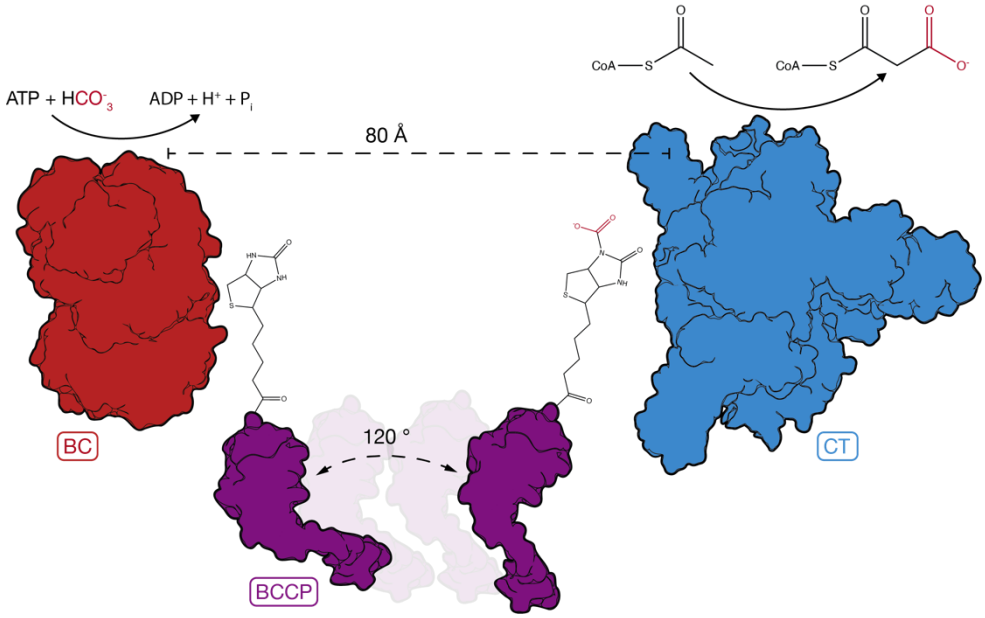

Figure 1.10: Carboxylation reaction of acetyl-CoA and domain movements of ACC
The BC domain carboxylates the biotin cofactor on the BCCP. The BCCP brings the carboxyl group to the CT active site, where acetyl-CoA is carboxylated to form malonyl-CoA. Both active sites are separated by 80 Å. This distance is bridged by the flexible BCCP shuttling between both domains. BC: Biotin Carboxylase, BCCP: Biotin Carboxyl Carrier Protein, CT: Carboxyl Transferase.

1.5.2 Regulation of eukaryotic ACC

Because ACC catalyses the rate-limiting step in fatty acid synthesis, its activity is tightly regulated, both transcriptionally and post-translationally. At the transcriptional level, ACC expression is regulated based on the cells nutritional status by Sterol Regulatory Element Binding Protein 1 (SREBP1) and Carbohydrate Response Element Binding Protein (ChREBP). This regulation is dependent on to the cellular nutritional state, sensed by mammalian target of rapamycin complex 1 and 2 (mTORC1 and mTORC2, Figure 1.11) [124, 125]. Due to the large size of ACC, regulation at the transcriptional level only does not allow for rapid adjustments in activity.

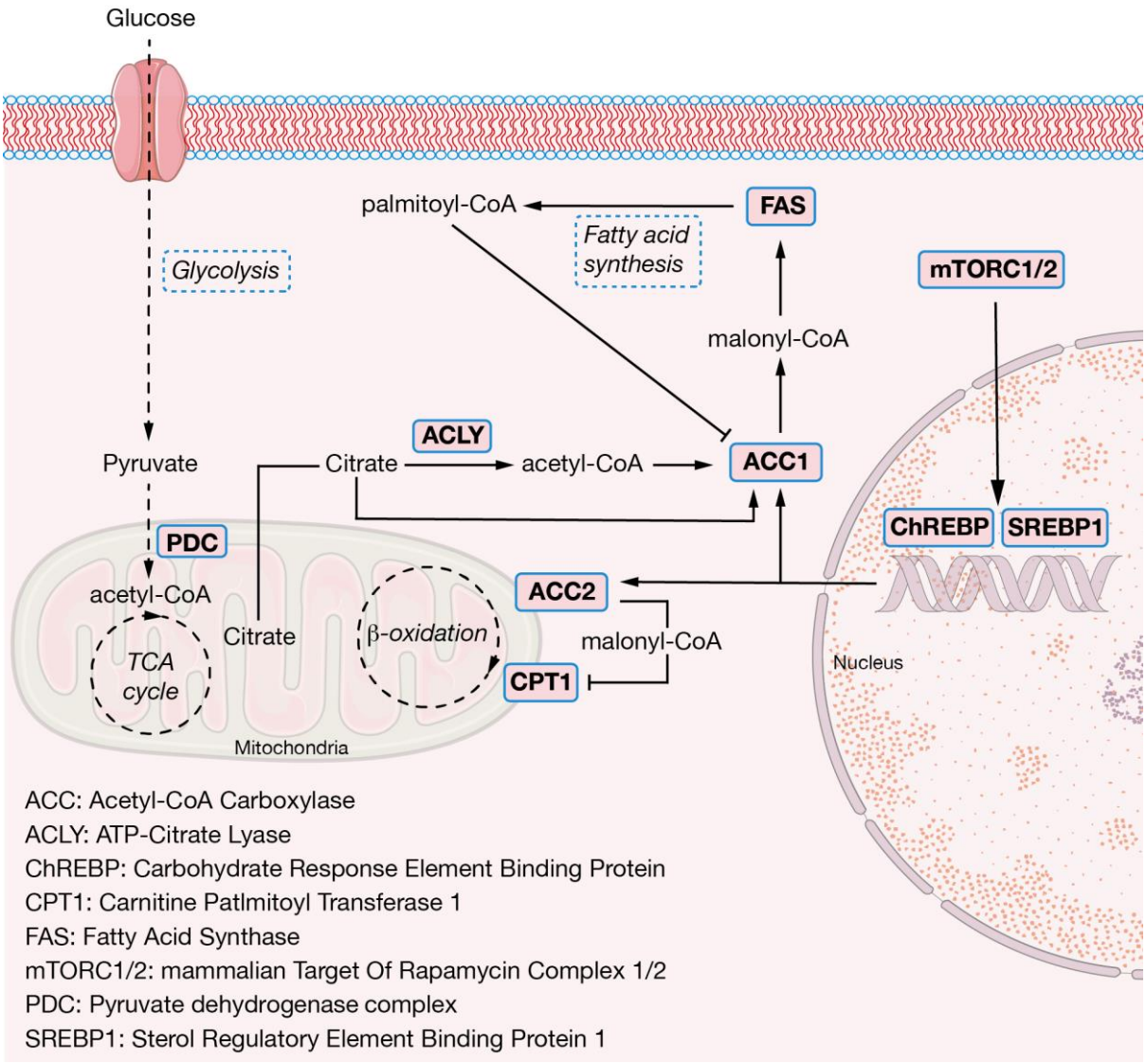

Figure 1.11: Role and regulation of human acetyl-CoA carboxylase in metabolism
Schematic representation of the central role of human ACC in fatty acid metabolism.

Almost 60 years ago it was shown that that adding citrate to ACC leads to increased malonyl-CoA production [126]. Activation of ACC from higher eukaryotes is correlated with formation of large filament-like oligomers as observed by electron microscopy. Later, it was shown that citrate addition results in ACC polymerisation into filaments [127]. These filaments were initially purified from avian livers and bovine adipose tissue. Citrate acts as the feed-forward allosteric activator of ACC in higher eukaryotes, but has no activating effect on ACC from microbes or plants [127]. Citrate is produced in the tricarboxylic acid (TCA) cycle. The enzyme ATP-citrate lyase (ACYL) converts citrate into acetyl-CoA (Figure 1.11). Acetyl-CoA is used as the substrate of ACC and carboxylated into malonyl-CoA, with citrate acting as the feedforward allosteric activator for catalysis. While the effect of citrate on the activity and polymerisation of ACC is known [7, 123, 127], how citrate binding leads to full activation and filament formation of ACC remains unclear.

Malonyl-CoA produced by ACC is subsequently used by mFAS to build fatty acids such as palmitate [128]. The product, palmitoyl-CoA acts as a feedback inhibitor on ACC, reducing its activity in high lipid conditions (Figure 1.11) [129, 130]. Because of its detergent properties it's suggested to be unspecific binding and a more general inhibition mechanism [131]. Still, addition of albumin to sequester non-specific binding of palmitoyl-CoA, did not abolish the binding of palmitoyl-CoA to ACC or it's inhibition of it. Even in the presence of citrate [132].

ACC is inactivated by phosphorylation by AMP-activated protein kinase (AMPK) and cAMP-dependent protein kinase (PKA). AMPK acts as an energy sensor by sensing AMP levels. With high ATP and low AMP concentrations, AMPK is inactivated, which in turn leads to ACC activation. As a result, fatty acid synthesis is stimulated and β-oxidation inhibited, to store the abundant energy in lipids. When ATP levels are low and AMP levels are high, AMPK is activated and thus inhibits ACC through phosphorylation. This promotes β-oxidation, to replenish the cells' energy stores (Figure 1.11) [133]. AMPK phosphorylates residues Ser 79, Ser 1200 and Ser 1215 and PKA residues and Ser 77 and Ser 1200 [134, 135]. Ser 79 is located N-terminally of the BC domain core. Phosphorylation of this residue leads to a 10 Å conformational change of the β19-β20 loop of the BC domain, that interferes with BCCP binding [113]. Interestingly, this site is also the target of the polyketide soraphen A inhibitor of ACC [113, 136]. To activate ACC, it is dephosphorylated by protein phosphatase 2A (PP2A) [137]. PP2A plays a role in cell cycle progression, and its activation of ACC results in fatty acid production required for membrane synthesis.

Additionally, ACC is also regulated by protein interactions. The tumour suppressor BRCA1 is involved in repair of DNA double strand breaks, ubiquitination, transcription regulation and cell cycle control [138]. Mutations in BRCA1 are associated with an increased risk of cancer and commonly found in breast and ovarian cancers. Women with germline mutations in the BRCA1 gene have an estimated 87 % risk of developing breast cancer before the age of 70 [139, 140]. The 220 kDa protein consists of an N-terminal RING domain, three nuclear localisation signals and two C-terminal BRCA1 C-terminus domains (BRCT) [141]. The BRCT repeats bind to phosphorylated by recognising phosphoserine or phosphothreonine motifs. Specifically for ACC, the BRCT domains binds to the phosphorylated Ser 1263 peptide [142, 143]. BRCT binding inhibits the dephosphorylation of Ser 79, keeping the BC domain in a catalytically inactive conformation. Additionally, binding of BRCT to ACC locks the enzyme in an open, inactive filament

formation. Distances between both the BC and CT active sites are too large to be reached by the BCCP in this open filament form [123].

The small cytosolic protein Mig21 binds to and induces polymerisation of both ACC1 and ACC2 *in vitro*. For ACC2, it is unknown if polymerisation For ACC1 it induces polymerisation without requiring citrate. It increases sensitivity to citrate, which in turn increases the enzymatic activity of ACC. It was shown to induce filament formation of both ACC1 and ACC2 *in vitro* [144]. However, filament formation of ACC2 *in vivo* when anchored to the mitochondrial membrane has not been shown. The prolyl isomerase 1 Pin1 is expressed in malignant tumours and its presence is associated with unfavourable outcomes. In human prostate cancer cells, Pin1 expression correlates with ACC1 expression. Pin1 binds to the C-terminal domain of ACC1, but not of ACC2. Binding prevents ACC1 protein degradation through the lysosomal pathway, resulting in increased protein levels [145]. Prolyl hydroxylase 3 (PHD3) is part of the nutrient-dependent signalling pathway. In responses to increased nutrient abundance it hydroxylates prolines of ACC2, thereby inhibiting fatty acid oxidation. Prolyl hydroxylase 3 inhibits fatty hydroxylates prolines 343, 450, and 2131 of ACC2, located in the BC, ATP binding and CT domain respectively [146]. While hydroxylation is one of the smallest post-translational modifications, it can favour the trans-configuration of the peptide backbone inducing considerable structural changes altering protein conformations and their activity [147]. Pro 450 of ACC2 is located near the adenine moiety of the ATP binding site of the BC domain. Hydroxylation of Pro 450 enhances ACC2 activity, likely by providing optimal ATP binding [146]. Due to its importance in lipid metabolism, controlling the activity of ACC is essential and dysregulation leads to pathologies.

1.5.3 Structures of ACC reveal its dynamic nature

Due to its importance in lipid metabolism and disease (*cfr.* 1.5.4) research on ACC has generated much attention. However, obtaining detailed structural insights has been challenging due to its large size and flexible nature. Structures of single enzymatic domains have increased our understanding. Only recently, structures of full-length ACC have been solved providing us with detailed insights into the multidomain arrangements of the enzyme.

The BC domain of *E. coli* was the first crystallised ACC domain [148]. It crystallised as a homodimer and contains ADP, biotin and bicarbonate in the active site (Figure 1.12 A, B). Crystal structures of the BC domains from yeast and human are monomeric (Figure 1.12 C,D). The dimer interface of *E. coli* BC differs from eukaryotic BC, suggesting that structural rearrangements are required for domain dimerization.

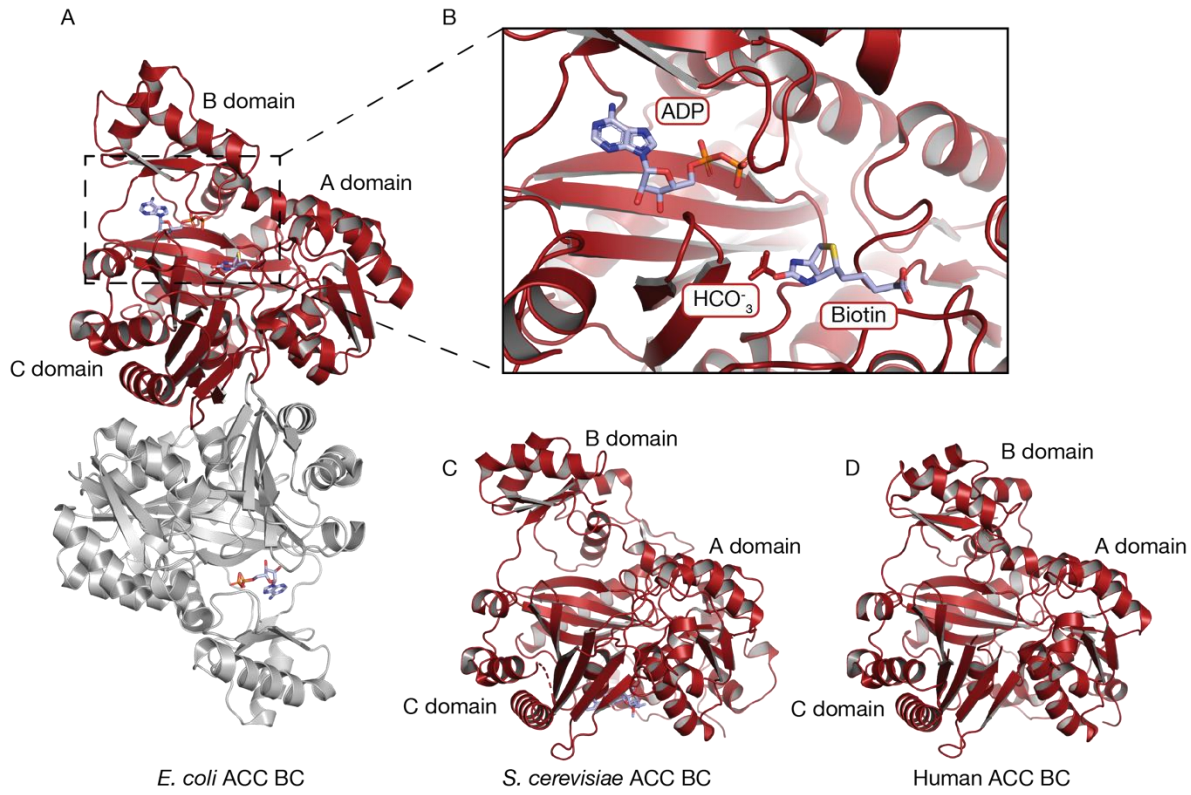

Figure 1.12: Crystal structures of ACC BC domains
A: Crystal structure of *E. coli* ACC BC crystallised as a homodimer in its active form (PDB: 3G8C, [148]). B: Active site of the BC homodimer containing ADP, Biotin and bicarbonate. Monomeric *S. cerevisiae* ACC BC (C, PDB: 1W96, [149]) and human ACC1 BC (D, PDB: 2YL2).

The BC consists of three subdomains. The A and C subdomains form the core with the flexible B domain on top of the active site. BC structures are conserved between prokaryotes and eukaryotes.

The CT is active as a homodimer, with each monomer consisting of an N- and C-domain (Figure 1.13 A). The fold of the CT domain is more between organisms than the BC domain. The CT active site is formed by a hydrophobic interface between both protomers (Figure 1.13 B) [150]. The CT is a common target for small molecule inhibitors, which prevent transfer of the carboxyl moiety to acetyl-CoA [151].

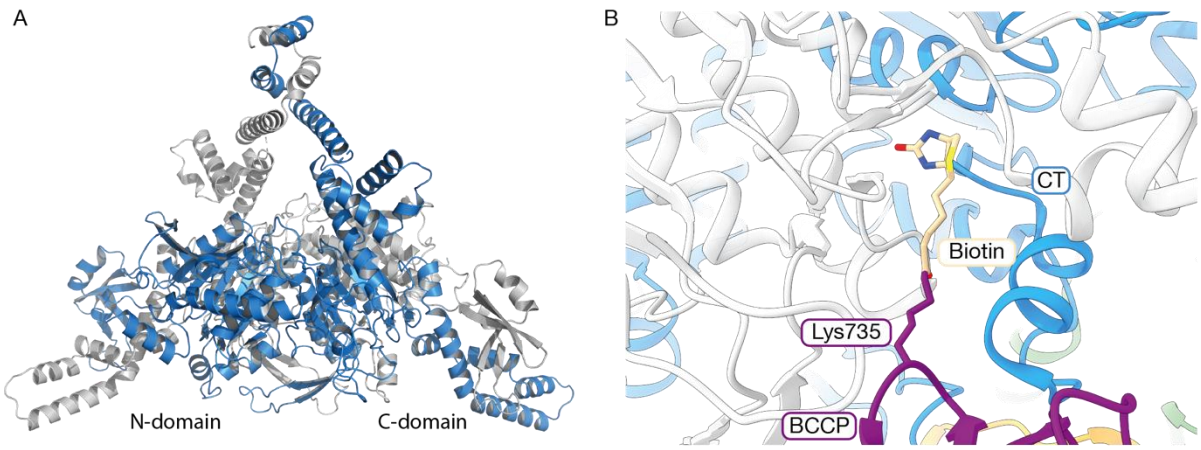

Figure 1.13: CT domain of ACC
A: CT domain dimer of human ACC. One monomer is shown in blue, the other in grey. (PDB: 4ASI) B: Biotin moiety of the BCCP in the active site of the yeast ACC CT domain (PDB: 5CSL). One monomer is coloured blue, the other in grey. The active site is formed at the interface of the dimer.

In 2015 the crystal structure of the full-length *Saccharomyces cerevisiae* ACC homodimer was solved (Figure 1.14 A) [113]. The dimer adopts a closed conformation, enabling the BCCP to reach the catalytic sites of the BC and CT domains. The BC domain is present as a homodimer, which corresponds to its active form, contrary to previously solved crystal structures (Figure 1.12 C). To reach this catalytically active conformation, the CD acts as a hinge bringing the BC domains together enabling dimerization. Flexibility of the CD is crucial in positioning the BC and CT domains for catalysis and therefore the target of regulation [122]. The central region is subdivided into five domains (AC1-5, corresponding to the CD N-terminal (CD_N), CD linker (CD_L) and CD C-terminal 1 and 2 (CD_{C1} and CD_{C2}) domains of human ACC). A regulatory loop is present in the CD_{C1}-CD_{C2} region of yeast that is the target of phosphorylation. After phosphorylation by Snf1, Ser1157 interacts with a positively charged pocket between CD_{C1}-CD_{C2}, locking the enzyme in the open catalytically inactive conformation [122, 152].

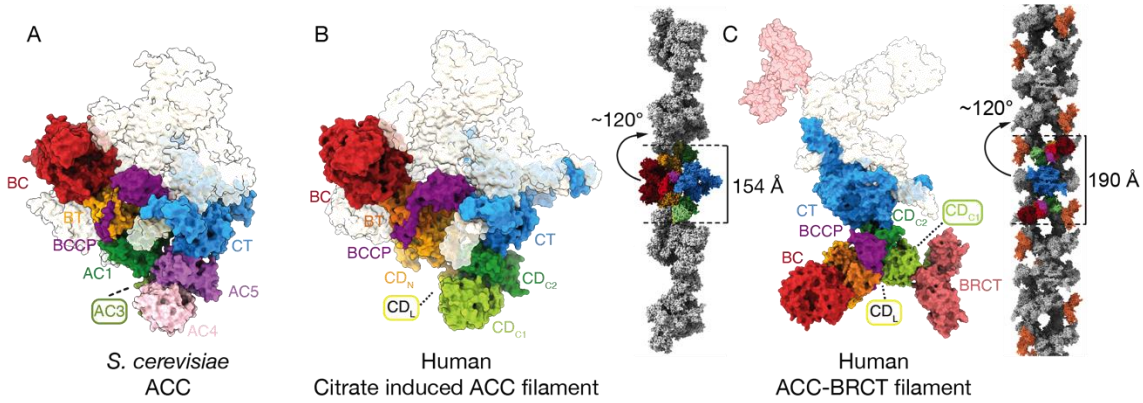

Figure 1.14: Structures of full-length of yeast and human ACC
A: Crystal structure of the full-length S. cerevisiae ACC homodimer. One monomer is coloured. The BC, BCCP and CT domains are coloured according to the human ACC colour code to show similarity, other domains are coloured according to the original publication (PDB: 5CSL, [113]). B: human ACC homodimer in its active, citrate bound conformation. ACC-citrate filaments are showed with one dimer coloured and the remaining ones in grey. The rise and twist between each homodimer are 154 Å and 120 °. C: Open inactive human ACC-BRCT filaments. The BRCT domain is bound to a phospholoop in the CD_{C1}. The rise and twist between each homodimer are 190 Å and 120 °.

Structural insights for full-length human ACC were obtained in 2018, with the cryo-EM structures of both active ACC-citrate and inactive ACC-BRCT filaments (Figure 1.14 B, C) [123]. In solution, human ACC is present as a flexible homodimer, as shown by negative stain EM. Due to this inherent flexibility, crystallisation attempts were unsuccessful. However, advances in cryo-EM enabled structure elucidation of human ACC.

Filaments of citrate induced ACC consist of stacked homodimers with a 120 ° twist and 154 Å rise, interacting through their CD. Comparting the interacting interfaces of the CD in the citrate-filament with the CD of yeast ACC in its active conformation revealed that filament formation of yeast ACC is not possible. In its citrate-filament form, the BC form dimer and the BCCP is able to reach both the BC and CT catalytic sites. Phosphorylation of Ser79 at the N-terminus of the BC prevents dimerization. In the ACC citrate filament, ACC is locked in its closed, catalytically most active form.

Binding of the BRCT domain to ACC, leads to inactive filament formation. BRCT is bound to phosphorylated Ser1263 and locks the enzyme in an open, inactive conformation [143]. In this filament form, the BC domain are monomers and the distances between the catalytic sites are too large to be reached by the BCCP [123].

Additional domains in eukaryotic ACCs are the BC-CT interaction domain (BT), which is wedged between the BC and BCCP (Figure 1.15 A). This domain mediates the interactions between the

BC and CT domains of the multienzyme, as both of these domains don't directly interact [113]. It consists of a long, central helix surrounded by a β–barrel with eight antiparallel β-strands. These interact with the N-terminally located BC domain and C-terminally placed BCCP.

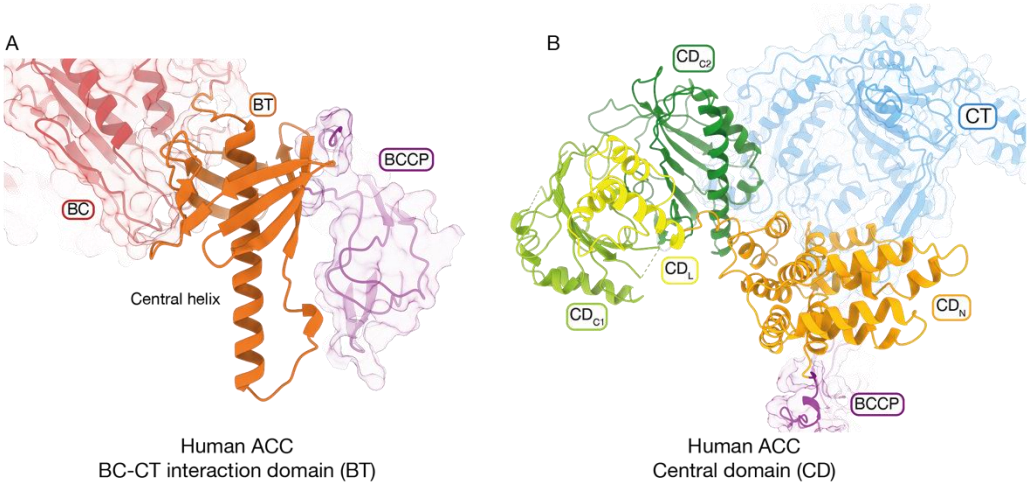

Figure 1.15: Structures of the BT and central domain of human ACC
A: BC-CT interaction domain (BT) of ACC flanked by the BC and BCCP. The central helix is surrounded by a β-barrel. B: Central domain of human ACC, with its four subdomains (CD_N-CD_L-CD_{C1}-CD_{C2}) coloured separately. It connects the BCCP with the CT.

The CD connects the BCCP to the C-terminally located CT domain (Figure 1.15 B). The CD provides the enzyme with the required flexibility to position the catalytic domains in the required position for catalysis. Contrary to the BC and CT domains, the CD provides minimal contributions to the dimer interface.

Detailed insights into the overall structure of human ACC are available and the reconstruction at a resolution of 5.4 Å allowed unambiguous placement of crystal structures of all domains. However, the mechanism by which citrate causes the large conformational rearrangements and filamentation of human ACC were not resolved.

1.5.4 The role of ACC in diseases and as a target for drug development

The utilisation of acetyl-CoA depends on the cellular energy state. In fed conditions, acetyl-CoA is used for fatty acid synthesis and inhibition of β-oxidation, via ACC1 and ACC2. During fasting, acetyl-CoA is generated through β-oxidation of fatty acids to generate energy [1]. Levels and activity of cellular ACC therefore display a proxy of the cellular metabolic state [153]. Expression of lipogenic proteins such ACC or FAS is correlated with the expression of many oncogenes [154-156]. These serve as diagnostic markers to assess the severity of cancers. ACC is overexpressed in liver, breast, prostate and colon cancer tissues, both at mRNA level and protein level [157-

161]. Dysregulation of ACC and subsequent changes in lipid metabolism contribute to the increased lipogenesis required for tumour cell growth.

Due to its central role in fatty acid synthesis and changes in expression levels of ACC in many cancers, ACC is a potential target for antineoplastic drugs [162, 163]. Furthermore, increased ACC activity is associated with type 2 diabetes, cardiovascular diseases, atherosclerosis and non-alcoholic fatty liver disease. The combination of these disorders is classified as metabolic syndrome [164, 165]. Targeting ACC as a means to decrease lipogenesis and thereby improve the patient's condition is and attractive therapeutic approach [165, 166].

As a proof of concept, knockdown of ACC by RNAi or chemicals inhibited cancer cell proliferation, reduced tumour cell invasiveness and induced cell death [167-170]. Due to the high sequence identity of ACC1 and 2 (~75 %), most ACC inhibitors are not specific and act on both isoforms [13]. Hepatic cells treated with an isoform-unselective inhibitor of ACC displayed increased lipid oxidation [171]. Rat models of fatty liver treated with a small molecule inhibitor of ACC1/2 had increased serum ketone levels, indicative of increased β-oxidation. However, there was no chronic reduction in hepatic triglyceride content and no significant effect on bodyweight or lipid mass [166]. This phenotype is also similar in mouse knockouts (KO) of ACC [172]. Chronic treatment with ACC1 inhibitors impairs glucose-induced insulin release, making it unattractive as a treatment for diabetes. However, ACC2 inhibition might still be a useful approach [173].

In mice with liver specific ACC1 KO, levels of hepatic triglycerides were elevated, without increase in liver mass, and β-oxidation was decreased [172]. These paradoxal findings suggest that there is a compensatory mechanism to replenish hepatic lipids. As a result of its function, ACC determines the cytosolic acetyl-CoA pool. Inhibition of ACC leads to accumulation of cellular acetyl-CoA. The same mice with liver specific ACC1 KO exhibited increased acetylation of cytosolic proteins. In another study performed in yeast, reduced ACC expression levels also led to an increase of protein acetylation [174]. Histone acetylation in particular leads to changes in the transcriptional landscape of cells. This mechanism is suggested to be a link between changes in the primary metabolism and epigenetic changes. However, it's not clear if this increased protein acetylation is specific and mediated by acetyltransferases, or if this happens non-enzymatically [172, 175]. These findings suggest that a desired ACC inhibitor should inhibit lipogenesis without triggering the compensatory increase in lipids and inhibition of β-oxidation, whilst at the same time not completely inhibit ACC to avoid the increase in cellular acetyl-CoA and subsequent protein acetylation.

The polyketide Soraphen A from the myxobacterial strain *Sorangium cellulosum* inhibits fungal ACC by preventing BC domain dimerization [113, 136]. Inhibition of plant ACC has found its use as a herbicide. These inhibitors target the CT and inhibit plant fatty acid synthesis by interfering with the transfer of the carboxyl moiety to acetyl-CoA [176, 177].

Overall, targeting lipid metabolism by ACC inhibition offers potential for the treatment of cancer and diseases classified under metabolic syndrome. Yet, ACC inhibition is not straightforward due to inhibitors targeting both isoforms, and the effectivity of inhibition must be tightly controlled. Insufficient inhibition risks to not improve the condition at all, excessive inhibition can lead to a compensatory increase in hepatic lipids and increased off-target protein acetylation due to acetyl-CoA accumulation.

1.6 Filament forming enzymes in biology

Filament formation was first reported in the late 1960s and 1970s for ACC and phosphofructokinase (PFK), and glutamine synthetase, respectively [127, 178, 179]. However, only until relatively recently new insights into protein oligomers and filamentation were gained. Oligomerisation happens frequently in biology, with about 35 % of all deposited proteins in the PDB being symmetric oligomers at the time of writing this thesis [180]. These go from small homodimers, to large oligomeric assemblies [181].

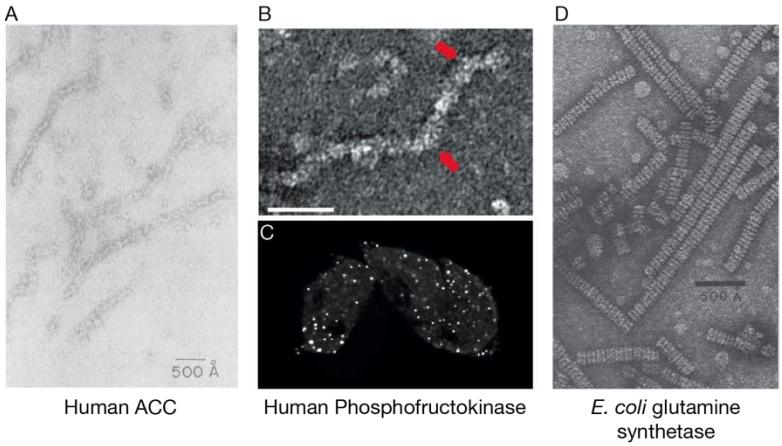

Figure 1.16: Example of filament forming enzymes
A: First negative stain EM image of human ACC filaments isolated from liver tissue [127]. B: Negative stain image of a phosphofructokinase (PFK) filament. C: Fluorescence microscopy image of GFP-tagged human (PFK) [182]. D: Helical filaments of glutamine synthetase from *E. coli* [179]. Images used with permission from the publishers.

Enzyme filaments are defined as reversible and functional linear assemblies of a single-type of enzyme [183]. Reversibility of assemblies allows to differentiate between irreversible protein aggregates due misfolding and functional oligomerisation. Yet, some proteins undergoing aggregation under heat shock, also tend to form filaments under nutrient stress conditions [184]. However, reversibility has not been tested for all observed enzymes displaying filament formation.

For a long time, detailed structural characterisation of filament forming proteins has been challenging due to limitations in the methods applied. Initial structural studies by X-ray crystallography typically favoured stable enzymes due to their ability to crystallise. Because of their dynamic and heterogeneous nature, this excluded many filament forming enzymes from initial crystallographic screens [183]. For this reason, filament-forming enzymes have typically been under-studied. Only since recently, thanks to advances in cryo-EM circumventing the requirement for protein crystals, it has been feasible to solve structures of large filament forming enzymes at atomic resolutions [123, 185, 186]. However, the challenges of sample flexibility and heterogeneity remain, making structural work still challenging.

Over the last decade, several enzymes have been shown to associate into filaments in cells of both micro-organisms and higher eukaryotes [183, 185, 187-193]. To observe the proteins of interest inside cells, they are typically fluorescently labelled, either by fusions with fluorescent proteins or with antibodies for immunofluorescence or electron microscopy (EM). Proteins are appear as distinct puncta or foci inside the cells as observed by fluorescence microscopy [188, 194]. On electron microscopy micrographs, these filaments are very heterogeneous in appearance, and the proteins can appear as rings, filament arrays, single spots, rods or fibres [183, 192]. They are often formed under particular conditions involving *e.g.* nutrient depletion or other cellular stress, but they also appear under normal physiological conditions [183, 188, 194]. For *in vitro* studies on the mechanism of filament formation and the effect of filament formation on enzyme regulation, they are typically recombinantly expressed and purified, then screened for oligomerisation by biophysical methods, such as light scattering [123, 182].

The effect of filamentation on enzyme activity is only known in detail for a few enzymes. ACC can be present as isolated homodimers, active and inactive filaments [123]. For ACC, the enzyme is fully active or inactive depending on its filament form and filamentation is suggested to stabilize the active or inactive states of the enzyme [123, 183]. Another example of active filaments is in the type II restriction enzyme SgrAI from *Streptomyces griseus.* It protects the bacteria from phage infections by cleaving phage DNA. Presence of phage DNA leads to rapid formation of

enzyme oligomers, enhancing the enzymes' ability to sequester DNA, reducing its DNA sequence specificity [195] and increasing its cleavage speed [196].

For these examples filament formation affects enzyme activity, but this is not the only advantage of filament formation. Based on their helical twist, certain plant nitrilases can accommodate different substrates. These findings changed the perception of substrate recognition for these enzymes [197]. The same conserved active site can catalyse different substrates based on the quaternary structure. These findings would not have been possible by sequence analysis alone.

So far it is not well understood which general advantages filament formation offers to enzyme function or regulation compared to other forms of regulation. The filament form can adjust the enzyme activity [123, 196], affect its stability, act as a buffer controlling the levels of active enzyme, direct product synthesis to a specific cellular region or change substrate specificity [197]. Assemblies are well ordered helical repeats of the proteins or layers of aligned filaments [192]. These could function as large local assemblies of specific enzymes, behaving like a gel-matrix [198] and provide high local concentrations of enzymes whilst allowing substrates and products to freely diffuse through the assemblies [199, 200].

Several enzymes have the ability to form filaments. For some this ability has been known for years but detailed insights into these mechanisms have only been obtained recently. The link between filament structure and its effect on enzyme or cell function, still remains largely unknown. However, with new developments in imaging methods, such as cryo-electron tomography, facilitate advanced studies to investigate the advantages of enzyme filamentation [201, 202].

1.7 Aim of this thesis

Multienzymes are large biosynthetic assemblies catalysing multi-step chemical reactions, involving unstable intermediates. All enzymatic domains and the substrate carrier protein are integrated into giant polypeptides. These form obligate dimers for the modPKS and ACC multienzymes. Individual observations, in cells or based on crystal contacts for modPKS, and in cell extracts or *in vitro* for ACC; further indicate the formation of giant sheath or filament supramolecular structures. The requirement of dimer formation for function has already been extensively rationalised at the level of individual domains, but the functional role and structural basis of higher-order, supramolecular assemblies remains enigmatic. In this thesis, I address this question with targeted approaches for modPKS and ACC. For modPKS, I aim to establish a biochemical system suited for *in vivo* and *in vitro* analysis, to allow correlating in vivo assembly to product formation, while being accessible to structural analysis *in vitro*. For ACC, earlier work in our lab already demonstrated the general regulatory role of filament formation on enzymatic activity. However, the mechanism controlling formation of active filaments in response to the allosteric activator citrate remains unknown and is revealed in this work. In detail, the aims of this thesis are:

1. Establish a system for *in vivo* analysis of higher-order assemblies, localisation and product formation of modPKS. The approach is to establish genetic manipulations, to set-up an assay to detect polyketide synthesis and to fluorescently label and image the assembly line proteins of the mupirocin PKS of *Pseudomonas fluorescens in vivo.* This part is discussed in chapter 2.
2. To analyse the structural organisation of the same modPKS *in vitro*. I approach this by cloning, recombinant expression and purification of the assembly line proteins of the mupirocin PKS, followed by biophysical and structural analysis. This part is discussed in chapter 3.
3. Reveal the mechanistic basis of citrate-mediated ACC filament formation and activation. This is addressed by structural analysis of ACC filaments by cryo-EM, establishing a reduced protein fragment for detailed studies of citrate binding by X-ray crystallography biophysical methods as well as functional analysis of ACC variants, which is discussed in chapter 4.

2 *In vivo* organisation of the mupirocin *trans*-AT PKS in *Pseudomonas fluorescens*

Microbial war: Fighting competitors from the pole position

Florian X. Delbart, Marek Basler, Timm Maier
Biozentrum, University of Basel

2.1 Declaration of project contribution

All work done described in the part on the *in vivo* organisation of *trans*-AT PKS was done by me. This consists of a literature search to find suitable *trans*-AT PKS and organisms for *in vivo* study. All microbial manipulations, optimisations of the protocol for genetic modifications, cloning of constructs and generation of new bacterial strains. Design and set-up of bacterial killing assays, preparing the cells for microscopy and performing all imaging experiments. All data analysis and making figures.

2.2 Mupirocin biosynthetic gene cluster of *Pseudomonas fluorescens*

Pseudomonas spp. are Gram-negative, aerobic bacteria commonly living in soil, water and on plant roots [203]. *Pseudomonas fluorescens* are typically found on plant roots, in a habitat called the rhizosphere [204]. Here, they protect plants from pathogens by producing a variety of antimicrobial and antifungal compounds [205]. Mupirocin (also called Pseudomonic acid) is a polyketide antibiotic produced by the soil bacterium *P. fluorescens* NCIMB10586 [206] and inhibits isoleucyl-tRNA synthetase [207]. It is in clinical use for the topical treatment of bacterial infections, commonly caused by methicillin-resistant *Staphylococcus aureus* (MRSA), under the trade name Bactroban® [208-210]. Due to poor pharmacokinetic properties and limited oral bioavailability, mupirocin is restricted to topical use [207]. However, mupirocin and antibiotic resistance in general is increasing [211-213]. Understanding of mupirocin biosynthesis would facilitate engineering of new variants with improved pharmacological profile, such as increased oral bioavailability [212].

The mupirocin gene cluster was first identified by transposon mutagenesis in the chromosome of *P. fluorescens* NCIMB10586 and later sequenced [51, 206]. Additional *Pseudomonas* strains carrying mupirocin producing genes have subsequently been identified [214]. Mupirocin synthesis occurs in two parts, first the polyketide moiety is formed by the assembly line proteins (Figure 2.1). Next, tailoring enzymes further modify it to produce the final product [215, 216]. It contains two large type I modPKS: MmpA and MmpD, consisting of two and four modules, respectively, and an iteratively acting module MmpB (Figure 2.1). Because the cluster is a *trans*-AT PKS, the AT domains are encoded by a separate gene *mmpC* [51]. Additional genes in the cluster are involved in its regulation by quorum sensing [217], self-resistance and tailoring enzymes further modifying the product [51, 216].

The bacillaene *trans*-AT PKS forms a large complex at the membrane, however, it is unknown if other *trans*-AT PKS also form such membrane associated complexes [103]. The megacomplex formation is hypothesised to be mediated by the LINKS-motif [104]. Based on sequence identities, the mupirocin PKS cluster is closely related to the bacillaene cluster, and it is already well characterised with all of its assembly line proteins described [51, 206]. Reporter experiments suggested that the cluster is expressed in late exponential to early stationary growth phase [217]. Additionally, genetic tools for manipulation of *Pseudomonas* strains are available [218]. The mupirocin PKS cluster contains eight KS in total and has been hypothesised to also form

megacomplexes mediated by the LINKS motif [104]. Until present, the assembly line proteins of the mupirocin PKS cluster have not been imaged in the native host. To gain more insights into the higher order organisation of bacterial *trans*-AT PKS, the mupirocin PKS assembly line of *P. fluorescens* was chosen for *in vivo* studies. To observe PKS assemblies in their native producer cells, the assembly line proteins were fluorescently labelled and visualised by microscopy.

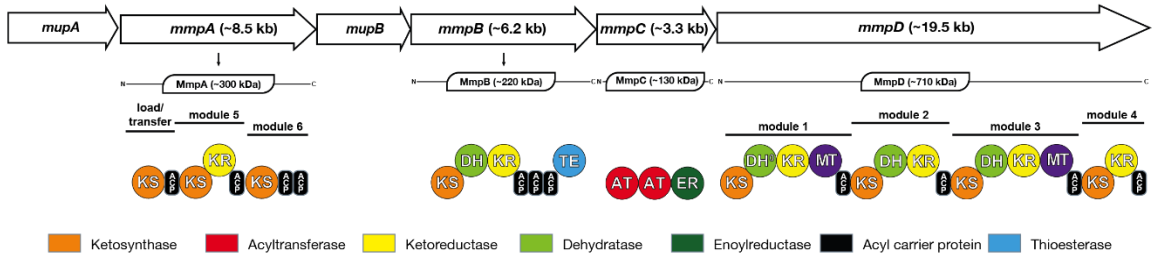

Figure 2.1: Scheme of the assembly line proteins of the mupirocin *trans*-AT PKS cluster
The gene, protein and assembly line protein organisation are shown. Each enzymatic domain of the assembly line proteins is shown in its distinctive colour and the module order is displayed on top.

2.3 Materials and methods

2.3.1 Bacterial growth conditions

Cells were grown in the following conditions, unless stated otherwise.
P. fluorescens NCIMB10586 from an overnight inoculation in LB or on LB-agar plate was grown overnight at 30 °C in LB supplemented with 20 µg/ml irgasan or on an LB-agar plate containing 20 µg/ml irgasan. *E. coli* cells were grown in LB, SOB or SOC medium or on LB-agar plates supplied with appropriate antibiotics at 37 °. *Bacillus subtilis* 168 was grown in LB or on LB-agar at 37 °C. Liquid cultures were incubated whilst shaking at 200 rpm.

2.3.2 Plasmid DNA isolation by miniprep and midiprep

Plasmid DNA was isolated from 4 ml or 300 ml overnight cultures by miniprep or midiprep, respectively, using a kit and following the manufacturer's instructions (Merck, Promega or Macherey-Nagel).

2.3.3 Transformations

Competent *E. coli* DH5α, SM10, JKE201 or NEB10β cells were transformed by heat shock or electroporation. For heat shocks, 50 µl chemically competent cells were incubated with 100 ng of DNA on ice for 30 min. Afterwards, cells were heat shocked at 42 °C for 45 s. For electroporation, 100 ng of DNA was added to 50 µl electrocompetent cells to a pre-chilled 1 mm gap cuvette and electroporated (15 kV, 5 µs). After transformations, 0.5 ml SOB or SOC medium

was added and cells were recovered for 1–3 h before being plated on selective LB-agar plates and incubated overnight at 37 °C.

2.3.4 Genomic DNA isolation

Two 5 ml overnight cultures of *P. fluorescens* NCIMB10586 grown in LB-irgasan (20 µg/ml) were used for genomic DNA (gDNA) extraction using the GenElute Bacterial Genomic DNA kit (Merck) and following the manufacturer's instructions. Briefly, cells were resuspended and lysed, incubated with RNase A and proteinase K. Cell debris was pelleted and the gDNA mixture was added to a spin column. Bound gDNA was washed twice and eluted. DNA concentration was measured by absorbance at λ=260 nm by Nanodrop and stored at -20 °C until use.

2.3.5 PCR amplification and purification of DNA sequences

The DNA regions of interest were PCR amplified from gDNA or custom synthesized DNA (Twist Bioscience) templates. Primers were designed *in silico* using Snapgene or Geneious Prime. For restriction cloning primers had 5' overhangs containing the restriction sites of interest extended with up to four nucleotides. For overlap extension PCR (OE-PCR), Gibson Assembly or In-Fusion cloning, primers contained 5' overhangs corresponding to the fragment to be annealed with. For OE-PCR, equimolar amounts of purified DNA products were used as templates.

Based on the purity, PCR products of the desired size were purified using either a PCR product purification kit or a gel purification kit (Promega or Qiagen). PCR conditions and cycling times mentioned in Table 1 and Table 2 were typically used, with variations for the annealing temperatures and elongation times. For colony PCR, 2 µl of colony resuspended in 15 µl MQ H_2O was used as a template. Typically, PCR conditions had to be extensively screened to obtain the desired product and several optimized primer pairs had to be designed for each construct.

PCR mastermix		
225 µl	**Concentration**	**Volume**
HF, GC or Herculase buffer	5x	45 µl
GC enhancer (NEB)	5x	45 µl
dNTPs (Promega)	10 mM	4.5 µl
F primer	10 µM	11.25 µl
R primer	10 µM	11.25 µl
Template	~100 ng/µl	1 µl
Q5, Phusion or Herculase polymerase	2 U/µl	2.25 µl
MQ H_2O	Adjust to final volume of 225 µl	

Table 1: PCR mastermix used for generation of *P. fluorescens* constructs
The mix is used to make 9x25 µl reactions for initial screening. Conditions yielding the desired product are repeated.

Step	T° (°C)	time	
Initial denaturation	98	10 min	
denaturation	98	30 s	x30
annealing	Variable, 50-70	30 s	
elongation	72	Variable	
final elongation	72	10 min	
storage	10	∞	

Table 2: PCR thermocycler settings
At the annealing step a variable temperature gradient was used to screen for optimal annealing temperature. As an elongation time 15 s/kB was used.

2.3.6 Construction of conjugative plasmids

For genome editing, the insert (*e.g.* a fluorescent protein coding gene) is flanked by 500 -1500 bp homology arms, matching the region of interest in the genome. The insert is then cloned into the multiple cloning site (MCS) of the pEXG2 [219] or pEX18Gm [220] conjugative suicide plasmids for *Pseudomonas* [218] by restriction cloning, Gibson assembly or In-Fusion cloning [221]. The plasmids carry a gentamicin resistance gene, the R6K ori that makes them unable to replicate into *P. fluorescens* and an oriT that allows conjugative transfer from donor *E. coli* to recipient *P. fluorescens*. Additionally, they contain the sacB suicide gene that encodes levansucrase.

For construction of an inducible plasmid, the transcriptional activator of the mupirocin cluster *mupR* was PCR amplified and cloned in the MCS of the pPSV35 plasmid using In-Fusion cloning [221]. For restriction cloning, purified PCR products or plasmids were incubated with the corresponding restriction enzymes according to the manufacturer's protocol (New England Biolabs, NEB). Digested DNA was purified using PCR product purification or gel purification (*cfr.* 2.3.5). Ligations were done with T4 DNA ligase or Quick Ligase (NEB) and the protocol was followed. For Gibson assembly, the NEBuilder® HiFi DNA Assembly Master Mix (NEB) was used and the instructions were followed. For In-Fusion cloning the In-Fusion Snap Assembly master mix (Takara Bio) was used and the manufacturer's instructions were followed. Plasmids were transformed in SM10 or JKE201 donor *E. coli* cells, and inserts were confirmed by colony PCR and sequencing (Microsynth).

2.3.7 Conjugation and two-step allelic exchange

The genome of *P. fluorescens* NCIMB10586 was engineered by allelic exchange [218]. The pEXG2 [219] or pEX18Gm [220] suicide plasmids were transformed into conjugative SM10 or JKE201 *E. coli* cells and grown on LB-agar gentamicin (15 µg/ml) plates overnight at 37 °C. *P. fluorescens* was grown on LB-agar irgasan (20 µg/ml) plates and incubated overnight at 30 °C.

For conjugation, *E. coli* donor cells were scraped, mixed with *P. fluorescens* cells and spotted on an LB-agar plate. Conjugation occurred for 3-5 h at 25 - 30 °C. The cell mixture was recovered in 700 µl LB and streaked, or immediately streaked on LB-agar gentamicin (15 µg/ml) irgasan (20 µg/ml) plates, to select for the first crossover. The integrated plasmids in *P. fluorescens* are selected with gentamicin and irgasan kills donor *E. coli* cells. Plates were incubated at 25-30 °C for 18 – 24 h. To select for the second crossover, *P. fluorescens* was grown on LB-agar no salt, sucrose (5 to 10 %) irgasan (20 µg/ml) plates for 24-36 h at 15 – 30 °C. After the second crossover colonies were screened for loss of integrated plasmid *i.e.* loss of gentamicin resistance by selection on LB-agar gentamicin (15 µg/ml) irgasan (20 µg/ml) plates. Finally, mutants were identified by colony PCR followed by sequencing of the chromosomal region of interest.

There were considerable difficulties in obtaining the second crossover step and extensive troubleshooting was done to identify conditions producing a second crossover and the desired mutants. New constructs were cloned, different inserts were attempted and different chromosomal regions were tried. After selection of the first crossover, cells were grown at 15 °C, 20 °C, 25 °C or 30 °C in 4 ml M9 minimal medium, LB without antibiotics, LB-irgasan (20 µg/ml) and LB no salt sucrose (5 %, 8%, 10 % and 30 %). Colonies were extensively screened for loss of integrated suicide plasmid, for some conjugations up to 600 colonies/construct were screened for loss of gentamicin resistance by multiple rounds of streaking or replica plating.

2.3.8 Conjugation and protein expression from inducible plasmids

For conjugative transfer of the inducible plamid containing *mupR* into *P. fluorescens,* SM10 donor cells were grown overnight on LB-agar gentamicin (15 µg/ml) plates at 37 °C. *P. fluorescens* was grown on LB-agar irgasan (20 µg/ml) plates and incubated overnight at 30 °C. For conjugation, *E. coli* donor cells were scraped, mixed with *P. fluorescens* cells and spotted on an LB-agar plate. Conjugation occurred for 3-5 h at 25-30 °C. The cell mixture was recovered in 700 µl LB, streaked on LB-agar gentamicin (15 µg/ml) irgasan (20 µg/ml) plates and incubated overnight at 30 °C, for two rounds. The presence of the plasmid was confirmed by colony PCR or isolated by plasmid miniprep followed by sequencing (Microsynth).

2.3.9 Widefield fluorescence microscopy

Overnight *P. fluorescens* cultures grown in liquid or on plate were resuspended to an OD_{600} of ~10 in 50-100 µl LB to obtain a dense cell population for imaging. 1 µl of cells were spotted on a thin LB-PBS (1:2 ratio) 1 % agar pad and covered with a coverslip. Samples were kept at room temperature (RT) during microscope setup. For induction of plasmid-based expression of the

gene of interest, *P. fluorescens* cultures grown for 16 h were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h before imaging. Cells were imaged using a Nikon Ti-E inverted microscope equipped with Perfect Focus System and Plan Apo 100x Oil Ph3 DM (NA 1.4) objective lens and Zeiss™ Immersol™ 518F/30 °C objective oil, SPECTRA X light engine (Lumencor), ET-GFP (Chroma #49002) and ET-mCherry (Chroma #49008) filter set for fluorescence excitation and filtration and detected with a scientific complementary metal-oxide-semiconductor (sCMOS) camera pco.edge 4.2 with pixel size of 65 nm (PCO). Imaging of the GFP wavelength exposure time was varied from 100 to 200 ms and for the mCherry wavelength 200 to 250 ms was used, with 20 % power output for the SPECTRA X light engine for all excitation wavelengths. The microscope was controlled with VisiView software (Visitron Systems). Imaging was performed at 30 °C and 95 % relative humidity in an Okolab cage incubator controlled by a T-unit (Okolab). Images were analysed in FIJI [222] and channel alignments were done with a customised plugin based on Stackreg 3D [223, 224]. Quantifications were done in Fiji, where the total amount of cells was counted with the analyse particles function after adjusting the image threshold to remove background and finding the edges of the cells. Fluorescent spots were counted with the find maxima process. Plots were made in Prism 9 (Graphpad).

2.3.10 Structured-illumination microscopy (SIM)

Overnight *P. fluorescens* cultures grown in liquid or on plate were resuspended to an OD_{600} of ~10 in 50-100 µl LB to obtain a dense cell population for imaging. 1 µl of cells were spotted on a thin LB-PBS (1:2 ratio) 1 % agar pad and covered with a coverslip. Samples were kept at RT during microscope setup. Imaging was done with a DeltaVision OMX Flex (Cytiva) equipped with UltimateFocus Hardware Autofocus module using Focus Assist and four sCMOS cameras. The 60×/1.42 NA Plan Apo N objective (Olympus) was used with 1516 custom blended laser liquid immersion oil (catalog #20130, Cargille labs). The 488 nm and 568 nm excitation lasers were used with 5.0 % and 10.0 % transmission, respectively with exposure times ranging from 70 to 200 ms. 2D SIM reconstructions were done in DeltaVision OMX softWoRx.

2.3.11 Bacterial killing assay

To assess mupirocin dependent killing, a plate-based bacterial killing assay was set-up [216]. *P. fluorescens* wild-type (WT) or mutant was grown in liquid culture at 30 °C in LB, shaking 200 rpm for 16 h. Cells were diluted to OD_{600} of 0.5 in LB and 10 µl was spotted on the centre of an LB-agar plate, dried at room temperature for 30 min, followed by overnight incubation at 30 °C for 18-24 h. For the prey *B. subtilis* 168 strain, an overnight liquid culture was diluted to OD_{600} of 0.5 in LB and 40 µl/ml of cells were addedd to molten LB-agar that contained 0.25 mg/ml of triphenyl

tetrazolium chloride (TTC). Each plate containing a central *P. fluorescens* spot was overlaid with 15 ml of LB-agar containing *B. subtilis* and TTC, and incubated for 16 h at 30 °C [225]. Afterwards, the zones of inhibitions were measured. Quantification was done by correcting the diameter of growth inhibition with the size of central *P. fluorescens* strain, followed by normalising to WT inhibition. An ANOVA was done to test for differences between mutants in Prism 9 (GraphPad).

2.4 Results

2.4.1 Construction of mupirocin knockout strains

In-frame and full deletions of the *mupR* and *mupN* genes of the mupirocin cluster were made. *mupR* encodes the transcriptional activator of the mupirocin cluster and *mupN* encodes the phosphopantetheinyl transferase [51, 217]. Both knockouts would be unable to produce mupirocin, yet by different means. In the Δ*mupR* strains, the mupirocin PKS cluster would not be expressed. In the Δ*mupN* strain, the cluster would still be expressed, but the carrier proteins would lack the phosphopantetheine moiety required to tether the growing polyketide.

For the in-frame KO, the first and last 10 codons of the gene were maintained. For Δ*mupR* and Δ*mupR* in frame mutants, clones 6, 7 and 1, 2, 7 contained the deletion, respectively (Figure 2.2). For Δ*mupN*, 4 out of 32 colonies screened maintained the conjugative plasmid. Eight of these colonies were screened by colony PCR followed by sequencing, and confirmed the deletion of the *mupN* gene.

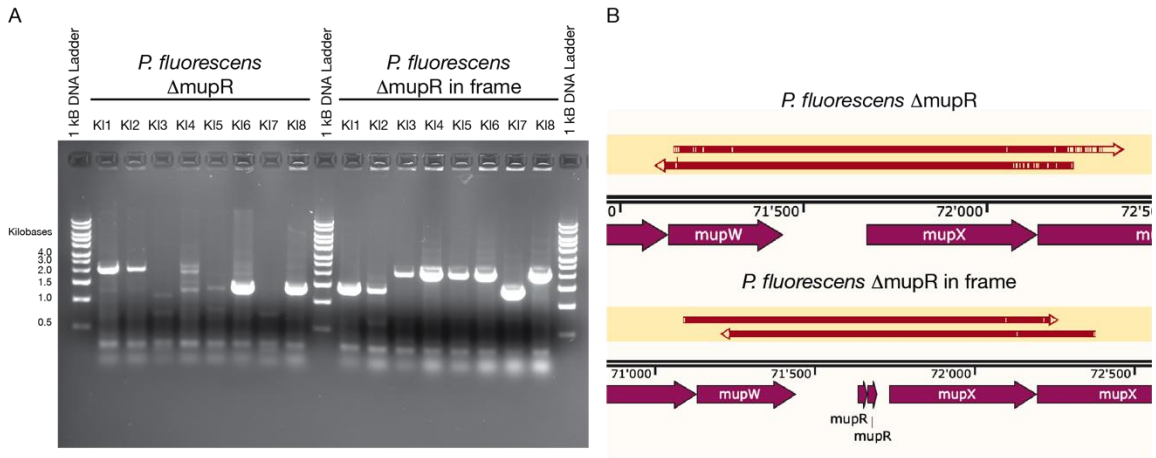

Figure 2.2: PCR screening and sequencing of Δ*mupR* constructs of *P. fluorescens*
Panel A shows an agarose gel of a colony PCR to screen for the gene KO with primers spanning the chromosomal region of KO. B: Sequencing result for the Δ*mupR* strains.

2.4.2 Bacterial killing assays of knockout strains

To assess mupirocin dependent killing, of mutants, a plate-based bacterial killing assay was done. A central *P. fluorescens* culture is overlaid with prey *B. subtilis* cells in LB-agar containing TCC. The redox indicator TTC is reduced into the red coloured formazan by bacteria undergoing cellular respiration, and is used as an indicator for living bacteria [226]. Zones containing bacterial growth appear red, while zones of growth inhibition are coloured light brown-yellow.

A significant decrease in size of the zones of inhibition is observed for the Δ*mupR* and Δ*mupN* strains compared to WT *P. fluorescens* NCIMB10586 (Figure 2.3). These observations are consistent with the expected outcome of these mutations. No mupirocin is produced as a consequence of no transcription of the biosynthetic gene cluster or inability of forming the polyketide moiety due to absence of the phosphopantetheine arm on the ACPs.

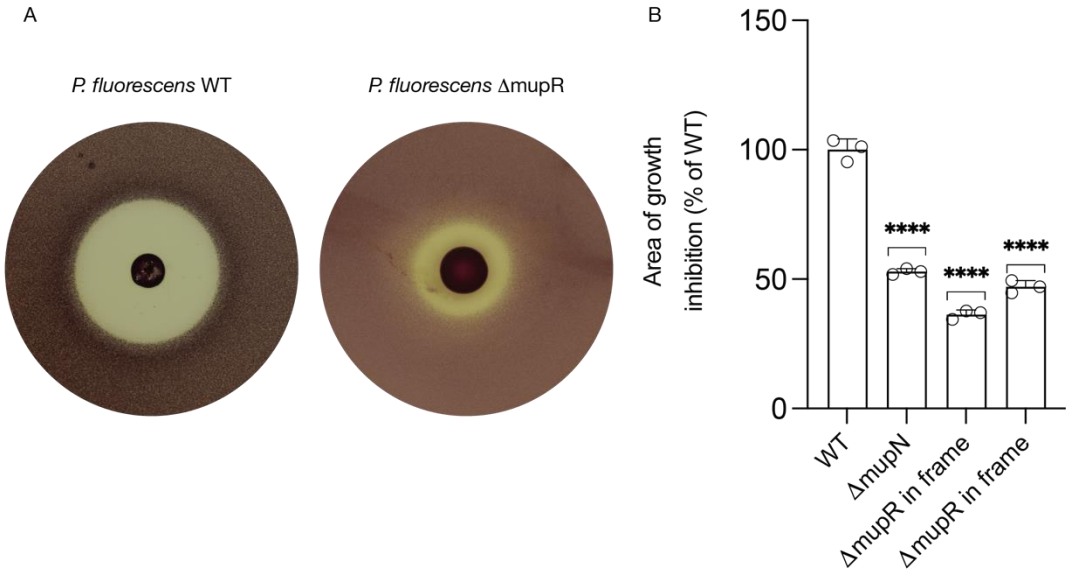

Figure 2.3: Bacterial killing assay assessing mupirocin dependent killing
A: Plates of *P. fluorescens* WT and *mupR* KO showing the zone of inhibited bacterial growth surrounding the central *P. fluorescens* spot. B: Quantification of growth inhibition of *P. fluorescens* strains. A significant decrease in inhibition is observed between WT and mutants, due to absence of mupirocin production. Error bars represent SD, n = 3, **** $p<0.0001$ at $\alpha = 0.05$ by ANOVA.

2.4.3 Construction and troubleshooting of fluorescent protein fusions to PKS assembly line

To fluorescently label the mupirocin assembly line proteins, fluorescent protein coding genes were fused in frame to the proteins of interest. Chromosomal fusions were chosen over recombinant plasmid-based expression or expression from a neutral site, to maintain protein stoichiometry and transcription from the same operon. Fluorescent proteins initially chosen were mNeonGreen (mNG) and mScarlet-I (mSc), based on their high quantum yields and fast maturation times *in vivo* [227, 228].

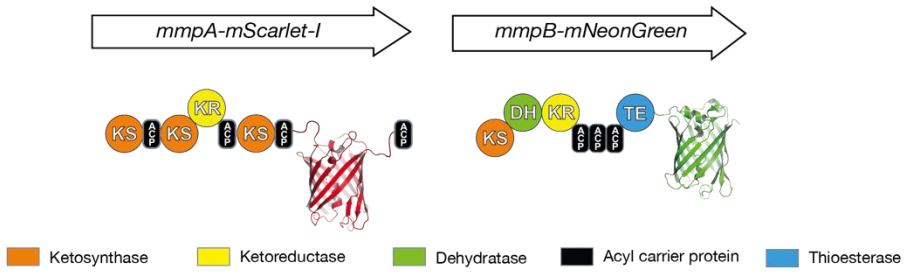

Figure 2.4: Fluorescent protein fusions for MmpA and MmpB assembly line proteins
mmpA is fused to mScarlet-I in the linker between both tandem ACPs. mNeonGreen is C-terminally fused to the TE domain of *mmpB*.

Initially, C-terminal fusion of the mNG coding gene to the *mmpB* assembly line and mSc insertion in the inter-ACP linker of *mmpA* were attempted (Figure 2.4). These regions were chosen to prevent potentially disrupting assembly line formation due to insertion of fluorescent proteins. The inter-ACP linker of mmpA was chosen for mSc insertion, because an NMR structure of the tandem ACP is available therefore its exact boundaries are known [229]. Manipulations of the number or type of tandem ACPs did reduce mupirocin production, but not completely abolish it indicating that mutations at this region are tolerated [230]. The last module of the assembly line is MmpB and after its TE domain, no interactions with assembly line proteins are expected.

After conjugation and sucrose selection for the second crossover, colonies were screened for loss of integrated plasmid by growth on LB-agar gentamicin (15 µg/ml) plates. After successful second crossover, colonies lose the integrated plasmid and thereby lose their gentamicin resistance. For *mmpA* mutants all screened colonies remained gentamicin resistant, for *mmpB* mutants, colonies 10, 19 and 58 (not shown) lost their gentamicin resistance (Figure 2.5).

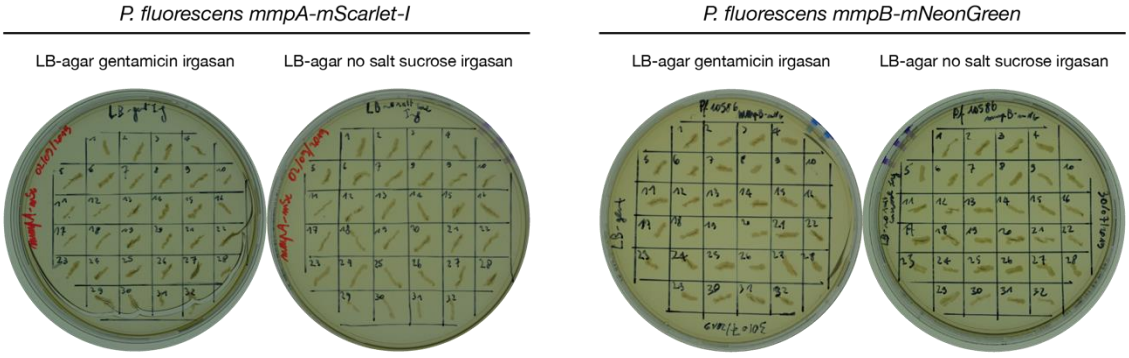

Figure 2.5: Screening for loss of integrated plasmid in *mmpA* and *mmpB* mutants
Single *P. fluorescens* colonies after sucrose selection were streaked on LB-agar gentamicin (15 µg/ml) irgasan (20 µg/ml) plates and LB-agar no salt, sucrose (10 %) irgasan (20 µg/ml) plates. Colonies that lost their integrated plasmid should not grow on LB-agar gentamicin plates.

After extensive screening of both *P. fluorescens mmpA* and *mmpB* mutants, in total about 200 colonies/construct were screened. To find out if the fluorescent protein coding gene was successfully inserted, these were screened by PCR (Figure 2.6). Screened colonies all reverted to WT after the second crossover.

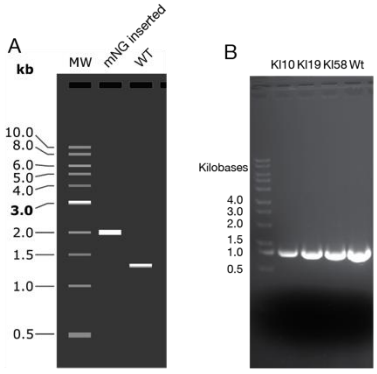

Figure 2.6: Simulated agarose gel and screening PCR for inserted mNG gene
A: Simulated agarose gel for mNG insertion in the region of interest or WT. When the mNG gene is inserted a product of ~2 kb is expected. B: Colony PCR of *P. fluorescens* colonies that lost the integrated mmpB-mNG plasmid and WT control. Screened colonies do not contain the mNG gene inserted. WT: wild-type control DNA that was amplified with primers spanning the genomic region of interest.

2.4.4 Conjugation in mupirocin non-producer strains

While I was able to generate KOs of the desired genes, generation of knock-ins using the same approach was unsuccessful. Transferring the plasmids by conjugation into *P. fluorescens* and generating the first crossovers was working, as indicated by colonies growing on under conditions selecting for the presence of the plasmid. Issues were at the second crossover step, based on screening of hundreds of colonies the majority maintained the integrated plasmid and were able to grow on selective plates despite carrying the sacB suicide gene (*cfr. infra* 2.4.6). In the rare cases when cells lost the conjugative plasmid, the genotype reverted to WT (Figure 2.6).
One hypothesis was that genetic manipulations at the chosen regions disrupted the mupirocin PKS assembly line, resulting in accumulation of a toxic polyketide intermediate (*i.e.* polyketide derailing product). For this reason, the insertions were tried in non-mupirocin producer strains (Δ*mupN* and Δ*mupR*). After sucrose selection, colonies were screened for second crossover by growth on selective LB-agar gentamicin plates (Figure 2.7). Only one picked colony lost gentamicin resistance and PCR screening revealed WT. In total 96 colonies/construct were screened.

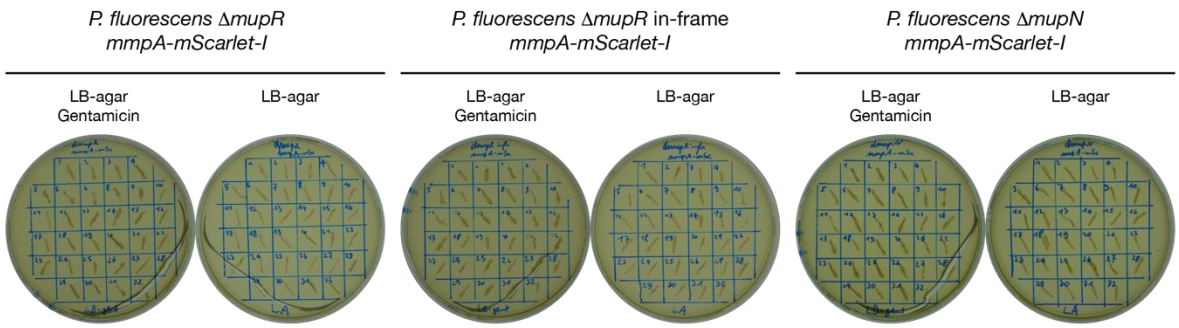

Figure 2.7: Screening for loss of integrated plasmid in non-mupirocin producer strains
After sucrose selection, colonies were streaked on LB-agar gentamicin (15 µg/ml) and LB-agar plates (control) to assess loss of gentamicin resistance *i.e*. second crossover.

2.4.5 PCR screening of *P. fluorescens* colonies outgrowing counter selection

To find out if the *P. fluorescens* colonies outgrowing the counter selection contained any insert in the region of interest they were screened by colony PCR using primers amplifying the chromosomal region of the insert. PCR products matching the expected size for the inserted fluorescent protein coding gene as well as WT were observed from the same colonies. Sequencing of these PCR products revealed a perfect match for the fluorescent protein coding gene insert (~2 kb product) and a perfect match for no insert (~1.3 kb product), *i.e.* WT (Figure 2.8). The PCR product at ~1.9 kb gave ambiguous sequencing results. Based on these findings, the assumption was made that colonies grown contain a mixed population of *P. fluorescens* containing the inserted fluorescent protein coding gene or WT. These were streak diluted for three rounds and again screened for loss of inserted plasmid and by PCR. However, all screened colonies revealed the same genotype. An additional PCR screen with primers specific for the conjugative plasmid revealed its presence in the screened colonies.

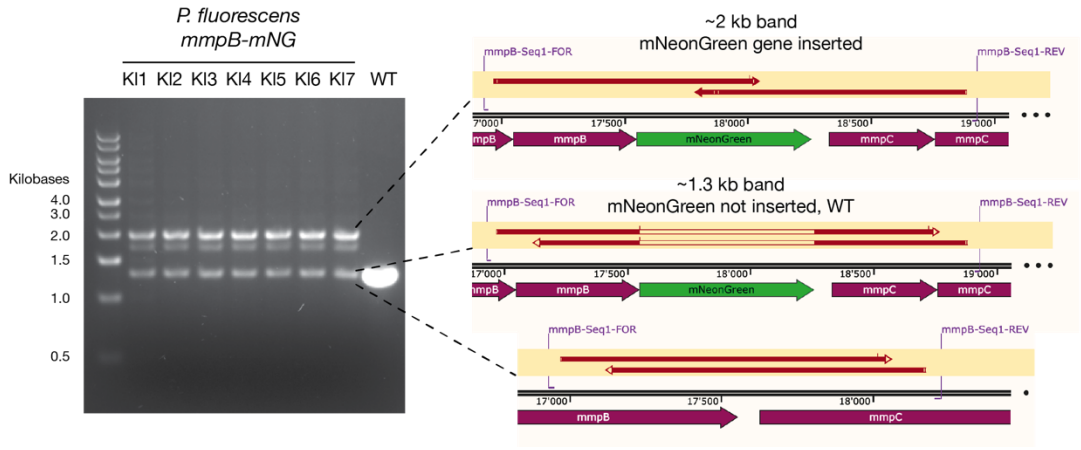

Figure 2.8: Colony PCR screen for insert in sucrose resistant colonies
The region of mNG insert was PCR amplified from colonies outgrowing sucrose selection. Three bands were visible on agarose gel at ~2 kb, ~1.9 kb and ~1.3 kb. Sequencing of the ~2 kb band revealed the insert inserted in the chromosome, while the ~1.3 kb band did not contain the insert, i.e. WT genotype. The ~1.9 kb band did not produce a clear sequencing result. WT: wild-type control DNA that was amplified with primers spanning the genomic region of interest.

2.4.6 Sequencing of *sacB* suicide gene and its promotor reveals mutations

In principle, there is a 50 % success rate in obtaining the desired genotype after the second crossover. Depending on the locus of the second crossover, WT genotype is restored or the mutant genotype is obtained [218]. In rare cases, the *sacB* suicide gene is inactivated due to mutations, as a consequence sucrose and antibiotic resistant mutants are obtained. *SacB* encodes levansucrase, which is exported to the periplasm where it polymerises sucrose into levan. In low salt conditions and at lower temperatures this causes bacteria to die from osmotic pressure [231].

Mutations in the promotor region or in the coding sequence could prevent expression, periplasmic export or activity of SacB. To identify mutations, the *sacB* gene and its promotor region were sequenced. Sequencing revealed mutations all over the coding sequence of the *sacB* gene. These mutations were not present in the initial plasmid that was created. A new suicide plasmid, pEX18Gm [220], also containing an oriT and a *sacB* suicide gene was obtained for repeated cloning experiments [218]. Prior to any cloning or conjugation experiments, the *sacB* suicide gene was sequenced, and no mutations in its coding sequence were detected.

2.4.7 Construction of fluorescent protein fusions or *xylE* gene fusions in pEX18Gm plasmid

The same fluorescent protein coding gene constructs to make the *mmpA* mSc and *mmpB* mNG mutants (*cfr*. 2.4.3) were cloned by restriction-ligation into the pEX18Gm plasmid. Additionally, as a positive control, the *xylE* gene (catechol 2,3-dioxygenase) was cloned as a C-terminal fusion to *mupA*. *mupA* is the first ORF of the cluster (Figure 2.1), and this fusion has already been reported in the literature [217]. Cloning of the *mupA-xylE* fusion was done analogous to the publication. After conjugation and screening for loss of conjugative plasmid, one clone of the *mmpA* mutant and 7 clones for the *mupA-xylE* mutants lost their antibiotic resistance and thus second crossover occurred (Figure 2.9). PCR screening followed by sequencing revealed that the *mmpA* mutant reverted to WT, while the *mupA-xylE* fusion was successfully inserted. Additional *P. fluorescens* colonies for *mmpA-mSc* and *mmpB-mNG* mutants were screened for loss of conjugative plasmid. However, the majority of screened colonies retained the plasmid and none contained the desired genotype.

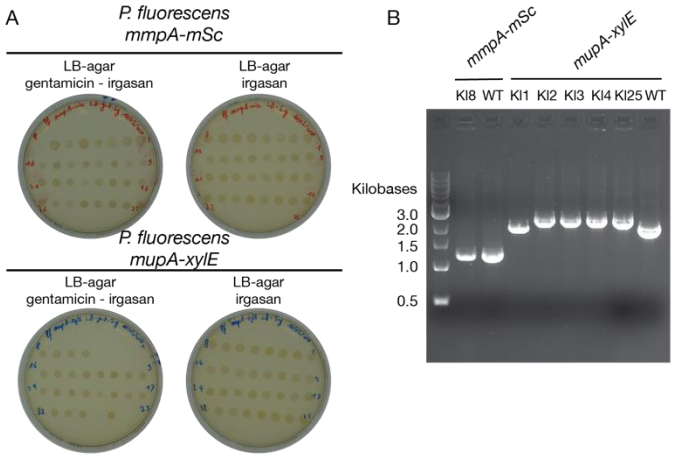

Figure 2.9: Screening for loss of conjugative plasmid
A: Screening for loss of antibiotic resistance of *mmpA-mSc* and *mupA-xylE* strains. Colonies that lost the integrated plasmid should not grow on gentamicin, all *P. fluorescens* colonies should grow on irgasan. B: screening PCR for insert. *mmpA-mSc* reverted to WT after the second crossover, for *mupA-xylE* four out of five screened colonies contain the insert, as indicated by the ~2.6 kb PCR product on the gel. WT: wild-type control DNA that was amplified with primers spanning the genomic region of interest.

2.4.8 Generation of constructs with extended homology arms

The first crossover occurs through one of the two homology arms, and instead of losing the plasmid by second crossover, cells are selected to escape sucrose killing by mutating the *sacB* gene [218]. One of the hypotheses is that crossovers occur preferentially through one of the two homology arms. The second crossover happens at the same locus and restores WT [218]. In an attempt to facilitate integration and second crossover through both homology arms, extended

homologous regions of ~1.5 kb length were cloned flanking the fluorescent protein coding genes of interest or the *xylE* gene. The reasoning behind the *xylE* insertion is that this gene was successfully inserted at a different locus (*cfr*.2.4.7) [217]. As a control, this gene insertion was also tested in the *mmpA* inter-ACP linker and at the C-terminus of *mmpB*. Unfortunately, again after extensive screening, no insert was obtained.

2.4.9 Repeat of cloning procedure and FLAG-tag insertion

Based on previous success with generating gene KOs we rationalised that since removing genes is possible, inserting shorter sequences at the regions of interest could also be possible. A FLAG-tag insertion in the inter-ACP linker of *mmpA* and at the C-terminus of *mmpB* was attempted with the goal to potentially enable immunofluorescence microscopy and protein pull-downs from cells. In parallel, a final try to insert fluorescent protein coding genes in the same regions was attempted. All cloning experiments were started with new materials and cloned from gDNA. Conjugation was done with an intermediate overnight growth step in a liquid culture containing LB-no salt-sucrose (10 %) irgasan (20 µg/ml) after selection for the first crossover, before plating on selective plates. All colonies were screened for loss of integrated plasmid by replica plating. For *mmpB-mNG*, the majority of screened colonies lost their gentamicin resistance. PCR screening followed by sequencing confirmed the presence of the insert in 2/7 screened colonies. For the FLAG-tag insertion in the inter-ACP linker of *mmpA*, all colonies outgrew sucrose killing and the experiment was stopped.

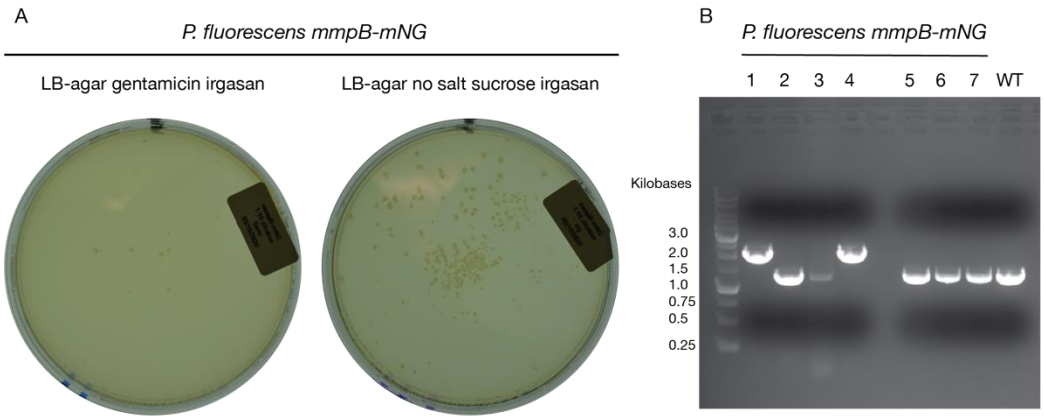

Figure 2.10: Replica plating and PCR screen for mNG insertion in *P. fluorescens*
A: Replica plating to screen all *P. fluorescens* colonies for loss of integrated plasmid. B: Colonies having lost their gentamicin resistance were screened for insert by PCR. Sequencing revealed the presence of mNG C-terminally of the mmpB gene. WT: wild-type control DNA that was amplified with primers spanning the genomic region of interest.

2.4.10 Killing assay with tagged *P. fluorescens* strains

To assess whether the *P. fiuorescens* strains containing fluorescent protein coding gene fusions retain their ability to produce mupirocin, a bacterial killing assay was performed. No significant difference in prey bacterial killing is observed between WT and fluorescently labelled strains (Figure 2.11). The fusions to fluorescent protein coding genes did not significantly alter mupirocin production and indicate that the mutants retain their ability to produce mupirocin.

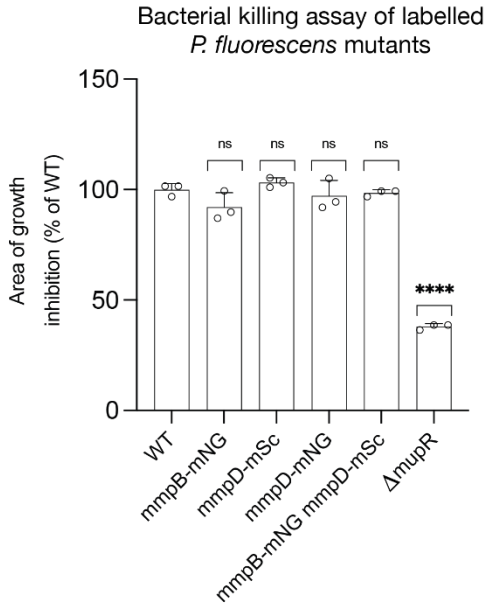

Figure 2.11: Labelled *P. fluorescens* strains retain mupirocin dependent killing
P. fluorescens strains containing fluorescent protein fusions retain their ability to produce mupirocin as tested in a bacterial killing assay. WT, positive control, Δ*mupR* negative control. Error bars represent SD, n = 3, ns: non-significant, **** p<0.0001 at α = 0.05 by ANOVA.

2.4.11 Microscopy of *P. fluorescens* mmpB-mNG reveals a single spot at the cell pole

Mutants carrying the *mmpB-mNG* fusion were imaged by 2D-SIM. Initially, the optimal timepoint for imaging was screened by growing bacteria to different densities. The *mmpB-mNG* fusion is expressed under the native promotor of the cluster. Fluorescent spots appeared 20-24 h after growth on LB-agar plates or in liquid cultures. This is consistent with expression of the mupirocin PKS cluster, starting after cells reached stationary phase and maximal cluster expression after 30 h [217]. Expression was heterogeneous, and below detection limit for a subset of cells. Yet, when fluorescence was detected, typically a single fluorescent spot was observed located at the cell pole, as visualised by structured illumination microscopy (SIM) (Figure 2.12). A single spot indicates that the tagged proteins are predominantly localized together at this region. In cells that appear to be dividing, the fluorescent spot is at the opposite side of the septum.

Figure 2.12: SIM images of *P. fluorescens* with *mmpB-mNG* fusion
2D-SIM images taken after ~20 h growth in liquid cultures reveals a distinct fluorescent spot for MmpB-mNG at the cell poles. The green channel was used for imaging. A 2 µm scalebar is indicated, and all images are shown at the same scale.

2.4.12 Construction of MmpB-mCherry2, MmpD-mNG and -mSc fusions

To test whether other assembly line proteins localize to the same spot, additional fluorescent protein fusions were made using the same approach that worked for constructing the *mmpB-mNG* fusion. Because of extensive issues in generating fusions of the *mmpA* assembly line gene, tagging *mmpD* was attempted. *mmpD* contains the first module of the assembly line and starts with a KS (Figure 2.1). At their N-terminus, KSs contain helices that extend away from their catalytic core. We hypothesized that N-terminal fusions with fluorescent protein coding genes at this region would likely not disrupt assembly line interactions [96]. Therefore, this site was chosen to insert mNG or mSc fluorescent protein coding genes to label MmpD. After conjugation, successful gene insertion was confirmed by PCR followed by sequencing and strains were imaged.

2.4.13 MmpD localizes to the cell pole as shown by fluorescence microscopy

P. fluorescens cells with mNG or mSc N-terminally fused to MmpD were grown for 20-24 h on plates and directly used for imaging (Figure 2.13). Three main phenotypes are observed. Similar to the MmpB-mNG fusions, a single, distinct fluorescent spot was observed at the cell pole. Some cells displayed diffuse fluorescence in their cytoplasm without a single fluorescent spot being visible. By imaging time series, we were able to observe cell an increase in fluorescence over the course of imaging. This may correspond to protein actively being translated, but not yet localized towards the cell pole. However, long-term imaging was complicated by rapid bleaching of the fluorophore over multiple exposures. Other cells did not show any detectable fluorescent signal, again revealing heterogeneity in expression of the fluorescently labelled protein for both mNG and

mSc labelled strains. Overall, imaging in the red channel gave better signal, likely as a consequence of increased autofluorescence in the green channel [232].

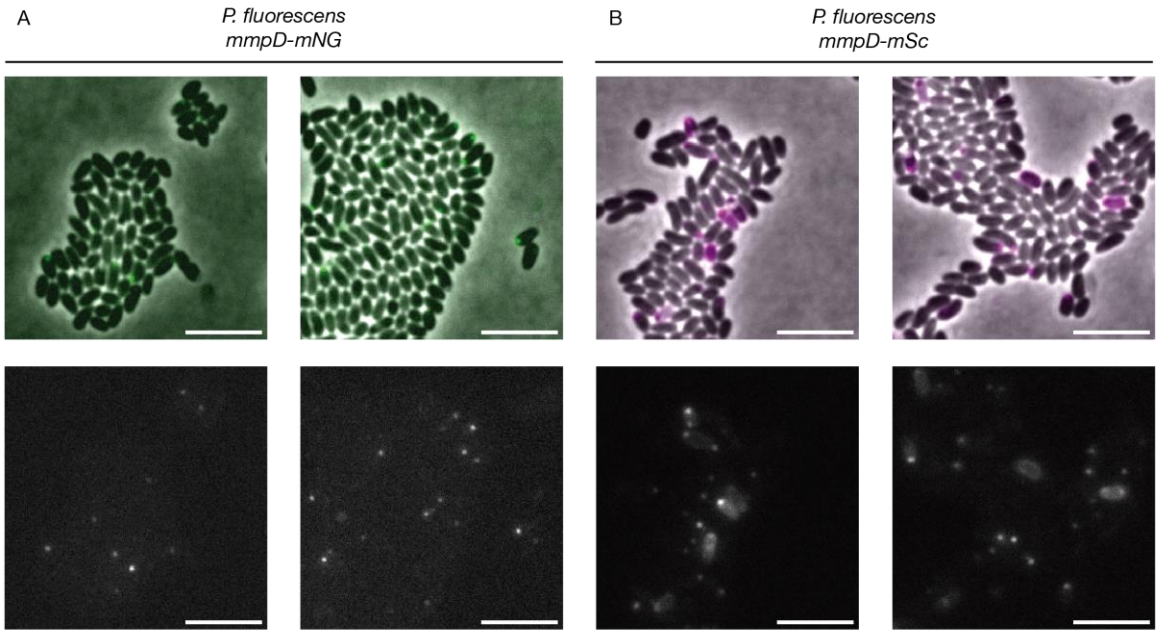

Figure 2.13: Fluorescence microscopy of *P. fluorescens* with labelled MmpD
Widefield live-cell microscopy of *P. fluorescens* with MmpD labelled with mNG (A) or mSc (B). Top images show the composite image of the brightfield and fluorescence channel. Bottom images only show the fluorescent channel (green or red channel for mNG or mSc, respectively). Fluorescent spots are visible at the cell poles, yet expression is heterogeneous between cells. 5 µm scalebar.

2.4.14 Imaging of MmpB and MmpD double labelled *P. fluorescens* strain

To assess if the assembly line proteins of the mupirocin PKS are localised together in *P. fluorescens,* both MmpB and MmpD were fluorescently tagged. Imaging of double labelled stains revealed distinct fluorescent spots in the cells for both imaging channels (Figure 2.14). Expression of labelled proteins was heterogeneous across the imaged cell population, as noticed before for single labelled strains. Cells were imaged between 20 to 24 h growth and across different cell densities, yet heterogeneity of expression remained. In cells grown for longer than 40 h, no fluorescence signal was detected. During imaging signal for either MmpB-mNG or MmpD-mSc was detected. However, in a subset of cells, fluorescence signal for both labelled proteins was detected. In these cells, the labelled proteins localised together at the pole (Figure 2.14). The presence of fluorescent foci of both labelled proteins at the cell pole suggests an interaction and supports the formation of large enzymatic assembly lines. To observe the dynamics of expression, imaging sequences were done. However, due to extensive bleaching caused by imaging with multiple wavelengths and high excitation power with low protein expression, imaging series were not possible. Detected fluorescent signals were already bleached withing the first frames.

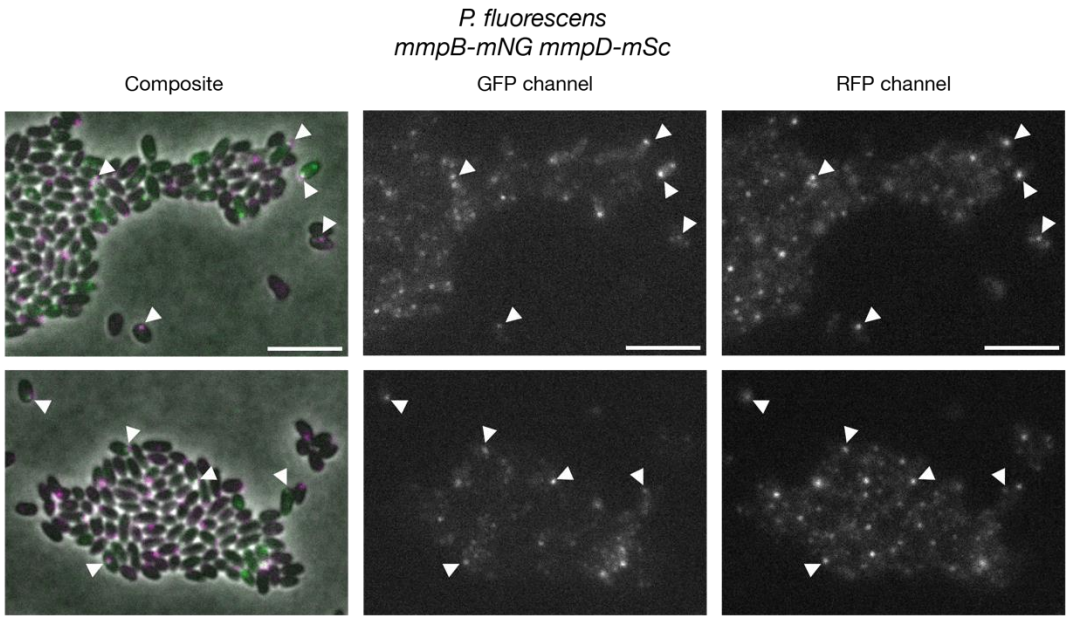

Figure 2.14: Localisation of labelled MmpB and MmpD PKS assembly line proteins
P. fluorescens with both mmpB and mmpD labelled shows bright fluorescent spots. Arrows indicate cells where both mmpB and mmpD localise together. Composite, GFP and RFP channels images are shown. Scalebar of 5 µm, both field of views shown at the same scale.

2.4.15 Overexpression of *mupR* transcriptional activator increases cluster expression

In an attempt to increase expression levels of the mupirocin assembly line proteins, the mupR transcriptional activator of the cluster was overexpressed from an inducible plasmid. While a slight increase in expression was observed in cells overexpressing *mupR*, heterogeneity remained (Figure 2.15). Quantification confirmed that cultures overexpressing *mupR* had increased fluorescence compared to uninduced cells (9.1 % fluorescent cells (n=4274) vs. 3.6 % (n=4490)). Overexpression of *mupR* did not affect localisation of PKS proteins, similar phenotypes are observed as for non-overexpression with assembly line proteins predominantly located at the cell pole.

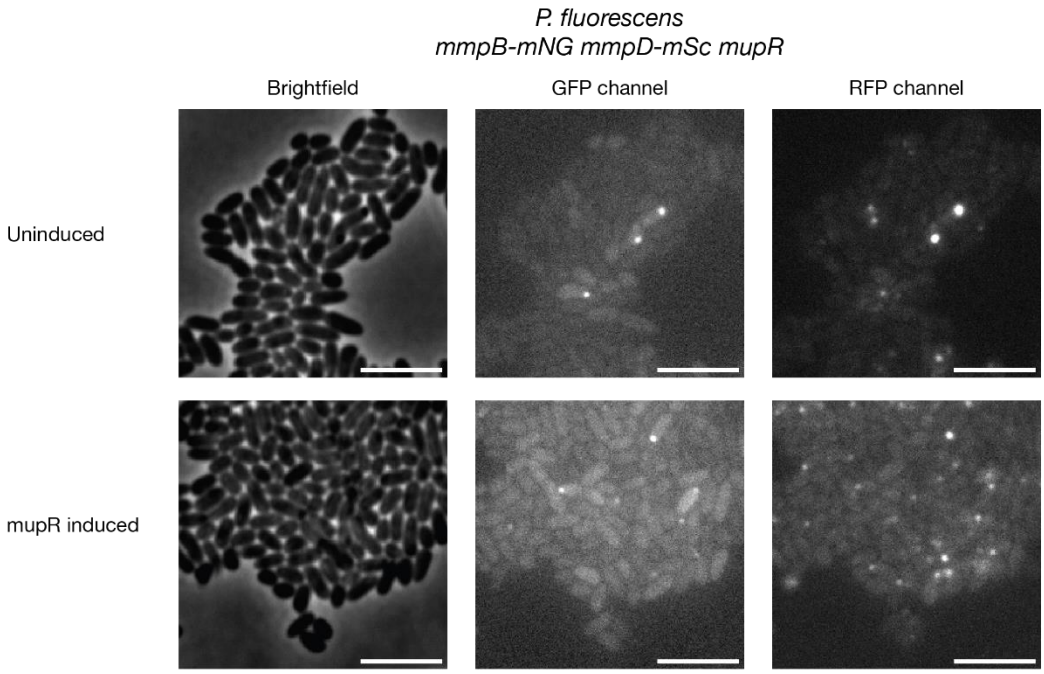

Figure 2.15: Overexpression of *mupR* transcriptional activator in labelled *P. fluorescens*
Imaging of *P. fluorescens* cells containing a *mupR* overexpressing plasmid. Uninduced and induced conditions are shown. Brightfield, GFP and RFP channels are shown. 5 µm scalebar.

2.4.16 Construction of fluorescently labelled MmpC *trans* acting AT protein

To assess if the *trans*-acting AT protein of the mupirocin cluster, MmpC (Figure 2.1) also localizes with the fluorescently-tagged assembly line proteins MmpB and MmpD, it was N-terminally fused to mNG or mSc. Cloning and conjugations were done as described above for *mmpB-mNG*. Antibiotic screening revealed loss of plasmid and the colonies were extensively screened for the insert by PCR (Figure 2.16). However, none of the screened colonies contained the inserted fluorescent protein coding gene, with all strains reverting to WT after the second crossover. Generation of a plasmid to express an N-terminal fusion of MmpC with fluorescent proteins also failed. None of the screened bacterial colonies contained the full-length *mmpC* gene fused to the fluorescent protein coding genes. Due to time constraints, these cloning attempts were stopped.

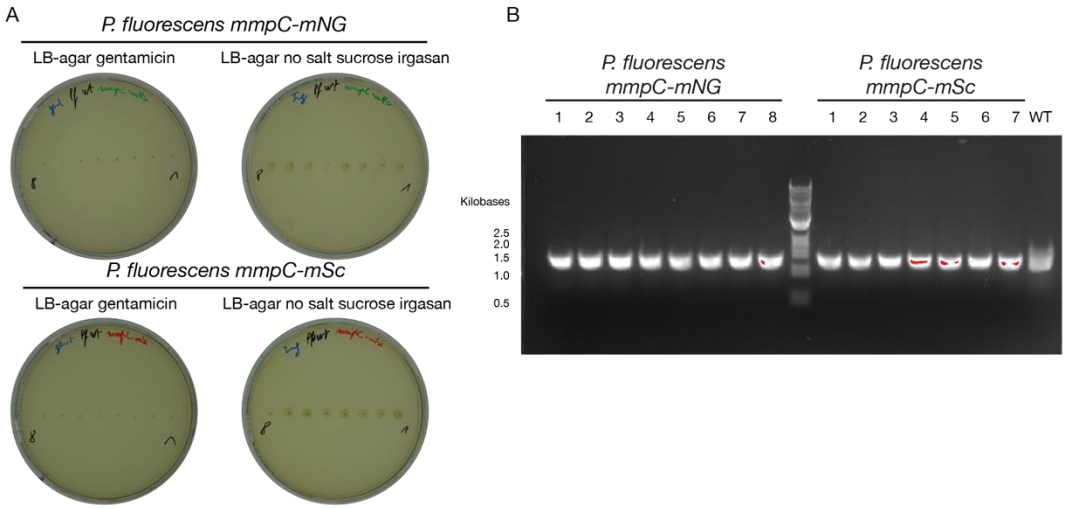

Figure 2.16: Screening for inserted gene fusion to *mmpC* after second crossover
A: Screening after sucrose selection for loss of integrated plasmid. B: PCR screening revealed that all mutant reverted to WT. WT: wild-type control DNA that was amplified with primers spanning the genomic region of interest.

2.5 Discussion

modPKS are large, dimeric multienzymes that are organised in gigantic molecular assembly lines, commonly reaching MDa sizes [18, 101]. Structural evidence suggests that *trans*-AT modPKS are able to form large complexes, however there is limited information about their organisation in native producer cells. Here, we studied the organisation of the mupirocin PKS in its native host *P. fluorescens* and used it as a model for *trans*-AT PKS. After considerable optimisation of cloning methods, we fluorescently labelled and imaged two of the assembly line proteins of the mupirocin PKS. Fluorescence microscopy revealed that the proteins form a single fluorescent focus, typically located at the cell pole. Imaging of double labelled strains revealed both proteins located at the same spot, although heterogeneity of expression is present in the population. Using a bacterial assay for product synthesis, we show that the labelled assembly lines remain functional and produce the antibiotic mupirocin. An attempt to assess if the *trans*-acting enzymes localise with the assembly line proteins failed due to cloning issues.

Fluorescence signal was detected when cells were grown to late-exponential phase, consistent with expression of other secondary metabolite clusters in this phase [27, 217]. Localisation of different assembly line proteins at the same region is consistent with their function, where the product is handed over to each assembly line protein. From these data, combined with structural findings [104], we hypothesise that as the assembly line proteins are synthesised by the ribosome and as they fold, they associate through their LINKS-motif. Due to the large size of and the

presence of multiple ribosomes on the transcript, during synthesis local protein concentration is high and would support these interactions. Localisation of different assembly line proteins agrees with their ability to interact through N- and C-terminal docking domains, enabling them to form an assembly line.

The association with a dedicated membrane transporter has been hypothesised [103, 105, 107], however no dedicated transporter is present in the mupirocin cluster [51]. In *Streptomyces coelicolor*, the type II PKS producing actinorhodin contains a membrane associated KS as shown by cell fractionation experiments followed by western-blot identification [233]. The type I PKS producing mycolactone of *Mycobacterium ulcerans* is also membrane associated, as shown by immunofluorescence [108]. Here, presence at the membrane is consistent with its role in synthesising mycolic acids as components of the mycobacterial cell wall. Microscopy reveals that the labelled mycolactone PKS is predominantly present at the cell pole, yet the exact reason why remains unclear. So far, it is not known if proteins of the mupirocin cluster are tethered to the membrane and through which mechanism they localise at the pole. Based on its molecular structure, mupirocin is able to cross membranes circumventing the requirement of a dedicated transporter. To potentially identify interaction partners responsible for tethering the assembly line at the pole, pull-downs of the assembly line proteins followed by protein identification through mass spectrometry can be performed. These findings show that mupirocin is synthesised at the pole of *P. fluorescens,* but it remains unknown how this large, MDa-sized assembly line is arranged in this region.

Future experiments would consist of combining mutations inferred from structural studies of enzymatic domains with imaging to study the structural determinants of protein interaction and localisations. Combining these mutational and imaging experiments with a functional assay would allow to study these enzymes *in vivo* while assessing the effect of specific regions and disruptions on product formation. However, throughout this study there were considerable difficulties in generating mutants. To study PKS oligomerisation *in vivo*, the ideal system should be feasible to genetically manipulate and have a characterised PKS with known product. Further experiments remain to be done to investigate the determinants of protein complex formation and its effect on product synthesis. *Trans*-AT PKS offer great potential for re-engineering to create novel polyketides with therapeutic potential. Ultimately, better knowledge of their organisation and function will advance engineering strategies.

3 Structural and functional studies of the mupirocin polyketide synthase assembly line

This research is part of the master thesis of Mirko Markovic

Mirko Markovic (Msc)[1,2], Florian X. Delbart[1], Timm Maier[1]

June 2021 – March 2022

[1]Biozentrum, University of Basel

[2]Swiss Nanoscience Institute

3.1 Declaration of project contribution

This work was part of an 8 months master thesis project for the master student Mirko Markovic from the Swiss nanoscience institute. I designed the project, taught the background and methods to the experimental part and was involved in day-to-day experimental supervision. Under my supervision, we optimized existing and set-up new protocols for gene cloning, expression and purification of the proteins of interest. I assisted in optimizing protein samples for biophysical studies and grids for single-particle cryo-EM data collection. Together, we performed biophysical and structural measurements, and analysed the data. Additionally, I assisted in lab book documentation, writing corrections and preparation for the master thesis exam.

3.2 Introduction

Polyketides are synthesised by PKS through the sequential condensation and modification of small acyl-building blocks into large, structurally complex polyketides [29]. Many polyketides are in clinical use as antibiotics or anticancer drugs [38]. modPKS are multienzymes that work in an assembly-line manner, where the growing polyketide is handed over to the next module [40]. Two evolutionary distinct classes of PKS exist: *cis*-AT and *trans*-AT PKS [52]. In *cis*-AT PKS the AT domain forms an integral part of the multienzyme, in *trans*-AT PKS it's encoded as a free standing enzyme [52]. The bacillaene *trans*-AT PKS forms a megacomplex *in vivo.* Docking domains are responsible for linking the assembly line proteins in N- to C-terminal direction. However, no additional domains have been identified to mediate megacomplex formation. A three-helix bundle flanking the KS of *trans*-AT PKS termed LINKS-motif was suggested to zip the assembly lines forming a PKS megacomplex [104].

Research on higher order assemblies of modPKS is an extremely important field, due to its potential high impact on drug research, drug development and capability to synthesize chemically complex molecules from engineered assembly lines. At present, most functional and structural insights of *trans*-AT PKS architecture relied on truncated or dissected fragments [97, 108, 234-236]. At the supramodular level, information about the connections between the subsequent enzymatic domains, but also their higher order organisation in the context of an enzymatic assembly line is missing. Therefore, this project focuses on reconstituting all parts of the mupirocin *trans*-AT PKS assembly line from *P. fluorescens* for *in vitro* structural and mechanistic analysis.

To more extensively study these enzymes *in vitro*, expression constructs of all assembly line proteins need to be established. Next, optimal expression and purification conditions for the proteins of interest need to be established. The oligomerisation behaviour of these purified proteins will be studied, to assess whether conditions producing filaments can be identified. To obtain structural insights into the domain organisation between modules the assembly line proteins are characterised by cryo-EM. Ultimately, the proteins are to be combined to reconstitute an *in vitro* assembly line. This reconstituted system can subsequently be used as a platform to study critical protein regions determining substrate turnover and provide additional insights into the biochemistry of PKS.

3.3 Materials and methods

3.3.1 Buffer list

Name	Ingredients
Insect Cell Lysis Buffer	50 mM Tris-HCl, pH 8.0 400 mM NaCl 0.2 % (v/v) Nonidet P-40 1 mM TCEP 5 mM $MgCl_2$
Size exclusion chromatography buffer (SEC-buffer)	20 mM Tris-HCl, pH 8.0 150 mM NaCl 5 % Glycerol 2 mM TCEP
Ni-lysis buffer	50 mM Tris-HCl, pH 8.0 150 mM NaCl 5% Glycerol 5 mM β-mercaptoethanol 2 mM $MgCl_2$ 1 µM Pepstatin A 2 µM Bestatin 10 µM 1,10-Phenantrolin 1 µM Phosphoramidon 1 µM phenylmethylsulfonyl fluoride 2 mg DNaseI
Ni-buffer A	50 mM Tris-HCl, pH 8.0 150 mM NaCl 40 mM Imidazole 5% Glycerol 5 mM β-mercaptoethanol 2 mM $MgCl_2$ 1 µM Pepstatin A 2 µM Bestatin 10 µM 1,10-Phenantrolin 1 µM Phosphoramidon 1 µM phenylmethylsulfonyl fluoride
Ni-buffer B	50 mM Tris-HCl, pH 8.0 150 mM NaCl

	500 mM Imidazole 5% Glycerol 5 mM β-mercaptoethanol 2 mM $MgCl_2$ 1 µM Pepstatin A 2 µM Bestatin 10 µM 1,10-Phenantrolin 1 µM Phosphoramidon 1 µM phenylmethylsulfonyl fluoride
FLAG-lysis buffer	50 mM Tris-HCl pH 8.0 250 mM NaCl 5 mM $MgCl_2$ 1 µM Pepstatin A 2 µM Bestatin 10 µM 1,10-Phenantrolin 1 µM Phosphoramidon 1 µM phenylmethylsulfonyl fluoride 2 mg DNaseI
FLAG-elution buffer	50 mM Tris-HCl pH 8.0 250 mM NaCl 5 mM $MgCl_2$ 0.2-0.6 mg/ml FLAG peptide 1 µM Pepstatin A 2 µM Bestatin 10 µM 1,10-Phenantrolin 1 µM Phosphoramidon 1 µM phenylmethylsulfonyl fluoride
10x SDS running buffer	35 mM SDS 1.9 M Glycine 248 mM Tris Base
5x SDS loading Dye	250 mM Tris-HCl, pH 6.8 50 % Glycerol 0.5 mM DTT 10 % (w/v) SDS 0.25 % (w/v) Bromophenol Blue
TAE buffer	40 mM Tris Base 20 mM Glacial acetic acid 1 mM EDTA

6x DNA loading dye	0.25 % (v/v) Xylene cyanol FF 30 % (v/v) Glycerol 0.25 % (v/v) Bromophenol Blue
Dextrin-blue gel stain	80 mg/l Coomassie brilliant blue G-250 63 ml/l Ethanol (100 %) 80 g/l Phosphoric acid 5 g/l α-Cyclodextrin
TBS	200 mM Tris-HCl pH 7.6 1500 mM NaCl
TBS-T	200 mM Tris-HCl pH 7.6 1500 mM NaCl 0.1 % v:v Tween-20

Table 3: Composition of buffers used in this project

3.3.2 Cloning of expression constructs

Gateway cloning

All genes of the mupirocin PKS assembly line (*mmpA-D*, Figure 2.1) were PCR amplified from genomic DNA (gDNA) with specific primers containing attB-sites for Gateway cloning. Each expression construct corresponds to one assembly line protein, except for *mmpD*. Due to its large size of about 20 kb, the *mmpD* construct was divided in two, split between the module boundaries, with each fragment containing 2 modules.

For Gateway cloning, the DNA sequences contain attB sites, which enable recombination with compatible gateway sites. In theory, this method enables rapid and efficient exchange of DNA sequences between multiple vectors while maintaining the open reading frame [237].

PCR mastermix		
225 µl	**Concentration**	**Volume**
Q5, HF, Herculase or GC buffer	5x	45 µl
GC enhancer (NEB)	5x	45 µl
dNTPs (Promega)	10 mM	4.5 µl
F primer	10 µM	11.25 µl
R primer	10 µM	11.25 µl
Template	~100 ng/µl	1 µl
Q5, Phusion or Herculase polymerase	2 U/µl	2.25 µl
MQ H_2O	Adjust to final volume of 225 µl	

Table 4: PCR mix used for amplification of mupirocin PKS assembly line genes
The mix is used to make 9x25 µl reactions for initial screening. Conditions producing the desired product are repeated.

Step	T° (°C)	time	
Initial denaturation	98	10 min	
denaturation	98	45 s	x30
annealing	Variable, 50-70	30 s	
elongation	72	Variable	
final elongation	72	10 min	
storage	10	∞	

Table 5: PCR thermocycler settings used in amplification of mupirocin PKS genes
At the annealing step a variable temperature gradient was used to screen for optimal temperature. As an elongation time 15 s/kB was used.

When a single, clear PCR product was present on agarose gel it was purified by PCR product purification (Promega). For reactions showing multiple, unspecific bands the PCR products of desired size were purified by gel extraction (Promega). The manufacturer's protocols were followed. For conditions in which no product could be obtained, new primers with increased length were designed, to reduce mispriming.

For Gateway cloning (Invitrogen) the manufacturer's protocol was followed [237]. Entry clones were created into circular or linearized with BamHI (NEB) pDONR221 or pDONR/Zeo plasmids (Invitrogen). The BP reaction was incubated at temperatures from 18 to 37 °C before proteinase K digest, followed by electroporation into NEB10β cells and overnight growth on selective LB-agar plates. Colonies were screened by colony PCR for the insert using primers amplifying a specific region of the insert or primers amplifying the full insert. Positive colonies were sequenced (Microsynth) and subcloned into a baculovirus expression vector (pACEBac2, Geneva Biotech) modified with a Gateway cassette containing an N-terminal His_{10}-myc-FLAG tag.

Gibson assembly of codon optimised gBlocks

Due to considerable difficulties PCR amplifying the regions of interest or cloning the inserts in the plasmids, gBlocks were ordered (*cfr.* Appendix, Twist Bioscience). Because synthesis and delivery time is faster for gBlocks this approach was chosen over full custom gene synthesis, due to experimental time limitations. The genes of interest were divided into fragments between 2–2.9 kb containing approximately 40 bp overhangs to the next gBlock or to the expression plasmid backbone (Appendix). The coding sequence was codon optimised with the main purpose to reduce GC content and for optimal codon usage for expression in *Spodoptera frugiperda* 21 (*Sf*21) insect cells. To assemble the gBlocks into the expression plasmids Gibson assembly was tried [221]. Specific primers with overhangs were designed for PCR amplification of the gBlocks and to insert them into the expression vector by Gibson assembly [221]. The PCR products were purified (*cfr. supra*) and the NEBuilder® HiFi DNA Assembly Master Mix (NEB) was used following the manufacturer's protocol. Briefly, samples were mixed together in equimolar amounts and the enzyme mix was added. They were incubated for 1 h at 50 °C before being dialysed and electroporated into NEB10β electrocompetent cells (NEB). Transformed colonies were grown overnight at 37 °C on selective LB-agar plates. Colonies were sent for sequencing (Microsynth).

TA-overhang cloning

The pGEM®-T Easy Vector System (Promega) was tried to insert the PCR amplified genes in a plasmid [238]. Hits would subsequently be used as a template for Gateway cloning. Purified PCR product was incubated with Taq polymerase at 70 °C for 30 min, according to the manufacturer's protocol. Next, the reaction was ligated into the pGEM®-T Easy vector. Ligated plasmids were electroporated into NEB10β electrocompetent cells (NEB) and plated on LB-agar Ampicillin (100 µg/ml) plates containing X-gal and Isopropyl β-D-1-thiogalactopyranoside (IPTG). The vector enables identification of positive clones by blue-white selection. Unfortunately, no white colonies were observed indicating absence of insert.

3.3.3 Baculovirus generation and protein expression

The Bac-to-Bac Baculovirus expression system (Thermofischer) was used. The manufacturer's protocol was followed unless stated otherwise. Virus production was performed according to Fitzgerald *et al.* [239]. The only deviation from the protocol was that DH10EMBacY cells (Geneva Biotech) were used to generate the bacmid [240]. Isolated bacmid DNA was mixed and incubated with FuGENE® HD Transfection Reagent (Promega) before transfection. Insect cells were maintained in suspension at 27 °C shaking 120 rpm in SFM4 Insect cell culture media. 72 h after transfection, the supernatant was collected and used to infect 25 ml of *Sf*21 cells (0.7 M/ml) in suspension in a 1:25 v:v ratio. After 4 days, supernatant from this virus amplification was used to

infect a 100 ml suspension culture of *Sf*21 cells (0.7 M/ml) in a 1:50 v:v ratio. After 3 to 5 days, when cells displayed signs of infection (increased diameter, expression of YFP reporter gene, decreased cell viability), supernatant containing the amplified virus stock was harvested and an expression test was performed. For protein expression, 4 l of *Sf*21 cells at 1 M/ml were infected in a 1:100 v:v ratio with P2 baculovirus stock supplemented with Penicillin (1000 units) and Streptomycin (100 µg/ml). Expressions were harvested when cells displayed signs of infection and viability was decreasing.

3.3.4 Expression test and western-blot

One ml of baculovirus infected insect cells were pelleted (12 °C, 800 g, 5 min) and resuspended in 100 µl insect cell lysis buffer (50 mM Tris-HCl pH 8, 400 mM NaCl, 0.2 % (v/v) Nonidet-P40, 1 mM TCEP, 5 mM MgCl2). The supernatant was resolved by SDS-PAGE on (4-15 % Mini-PROTEAN® TGX™, Bio-rad) and transferred onto 0.2 µM pore size nitrocellulose membranes via the Trans-Blot Turbo Transfer System (Bio-rad). The membrane was blocked with Blocking Buffer (BB) for 1 hour at room temperature, then washed 3x with 10 ml of TBS-T. Afterwards, the membrane was incubated overnight shaking at 4 °C with 10ml of (1:1 TBS-T and BB + 1:1000 Anti-FLAG® M2-Antibody from mouse). After three washing steps with TBS-T the membrane was incubated with 10 ml secondary antibody mix (1:1 TBS-T and BB + 1:10000 IRDye® 800CW Goat anti-Mouse IgG Secondary Antibody). Finally, the membrane was washed twice with TBS-T followed by one wash with TBS and then dried in the dark until imaging.

3.3.5 Protein purification

All purifications were performed at 4 °C and the samples were kept on ice between each step. For FLAG-purifications, between 15 to 25 g of insect cell expression pellet was thawed in FLAG-lysis buffer. Cells were lysed by sonication (50 % amplitude, 2x 5 min, 1 s pulse, 1 s off, on ice). The lysate was cleared by ultracentrifugation at 160000 g and the supernatant was filtered with a 0.5 µm syringe filter. The filtered lysate was then incubated with 3 ml of FLAG-beads for 1 h. Beads were washed with 100 ml of FLAG-lysis buffer using a gravity flow column. The Äkta pure 25 system was flushed with FLAG-lysis buffer and a FLAG-column was packed with. The protocol consisted of 10 column volumes (CV) wash with FLAG-lysis buffer. The protein was eluted by three cycles of FLAG-elution buffer that was and incubated onto the column for 30 min followed by 3 CV of elution with FLAG-elution buffer in up-flow. Eluted fractions were pooled and concentrated using 100 kDa Molecular Weight Cutoff (MWCO) centrifugal filter concentrators (Merck-Millipore) at 3220x g, 12 °C in 15 min intervals. The protein was concentrated to a final concentration of 1 mg/ml and flash-frozen in liquid nitrogen before storage at -80 °C.

For Ni-affinity purifications, 15 g of cells were lysed by sonication (50 % amplitude, 2x 5 min, 1 s pulse, 1 s off, on ice) in Ni-lysis buffer and the lysate was cleared by ultracentrifugation at 160'000x g. Cleared lysate was filtered (0.8 µm) and loaded on a 5 ml Ni-NTA column (Ni Sepharose 6 Fast Flow, Cytiva), that was washed with 15 CV of Ni-buffer A and eluted with a gradient of Ni-buffer B over 10 CV. Optionally, a size-exclusion chromatography (SEC) purification was performed. Protein from FLAG- or Ni- elution was concentrated to 1.5-3 mg/ml and a volume of 250 µl. During concentration visible aggregates were formed, and the protein was spun down 16'000x g 12 °C, 10 min prior to injection. Protein was loaded on a Superose 6 increase 10/300 GL column. Fractions containing the protein were pooled and concentrated with 100 kDa MWCO concentrators (Merck-Millipore) at 3220x g, 12 °C in 15 min intervals to a concentration of 0.5-1 mg/ml. Aliquots of 50 µl were flash frozen in liquid nitrogen and stored at -80 °C until further use.

3.3.6 Sample preparation for cryo electron microscopy

For initial in a FEI Talos F200 transmission electron microscope (TEM) operated at an acceleration voltage of 200 kV, 4 µl of purified protein in SEC buffer at 0.1 to 0.5 mg/ml was added to Quantifoil® R 1.2/1.3 Cu 300 or UltrAufoil® 200 mesh grids that were glow discharged for 30 s at 50 mA (GloCube). Grids were prepared with a Vitrobot Mark IV (Thermofisher) at 15 °C, 95 % humidity, 10 s wait time. Blot time was varied from 2.5 – 3.5 s and blot force from 0 to 3.
For high resolution data collection, 4 µl of mmpD modules 3-4 in 20 mM Tris pH 8.0, 150 mM NaCl, at a concentration of 0.2 mg/ml was added Quantifoil® R 1.2/1.3 Cu 300 mesh that were glow discharged for 30 s at 50 mA (GloCube).
Grids were screened for ice thickness and particle distribution, a set of images on different grid positions was collected using a FEI Ceta 16M Pixel CMOS camera mounted in the Talos microscope. Final data were collected on a Glacios TEM (Thermofischer) operated at an acceleration voltage of 200 kV and equipped with a Gatan K3 24M Pixel direct Electron Detector with a pixel size of 0.878 Å. movies were acquired at 0.8–2.0 µm defocus with a total dose of approximately 55.6 $e^-/Å^2s$ exposure, fractionated over 39 frames, using beam-image shift to record four images per hole except for initial test images.

3.3.7 Single particle analysis of MmpD modules 3-4

A total of 4933 movies containing 39 frames were imported in into the program Cryo-EM Single Particle Ab-Initio Reconstruction and Classification (CryoSPARC2) [241]. All movies were aligned with patch motion correction and the contrast transfer function (CTF) was estimated with patch CTF. Micrographs were visually inspected and a total of 196 were excluded giving a final set of 4737 micrographs. Initial particles were picked with a blob picker with minimum and maximum

diameter of 80 and 120 Å. 2D classification revealed a heart-like structure resembling a KS. A selection of KS-like 2D classes were used as templates for the template picking.
Two strategies were used to try to obtain a high-resolution information of the PKS bimodule. To specifically focus on the KS, a small box size corresponding to 224 to 288 pixels was used in a template picker step. Multiple rounds of picking and 2D classification were done followed by non-uniform refinement (NU-refinement) and heterogeneous refinements. In an attempt to obtain density for the next domain of the construct, the KS was symmetry expanded in C2, followed by NU-refinement, local refinement and 3D classification. The other strategy consisted of using the KS as a template for picking with a large box size of a maximum of 512 pixels in an attempt to align additional domains in the 2D classification step. Several rounds of 2D classification and 3D *ab initio* reconstructions were performed searching for extra features around the KS.

3.3.8 Differential scanning fluorimetry and time control experiment

All experiments were performed using the Prometheus NT.48 nanoDSF (Nanotemper). Purified, full length, MmpA was loaded in high sensitivity capillaries at a concentration of 1.8 mg/ml in SEC. Thermal denaturation was monitored using the intrinsic fluorescence with excitation at λ: 280 nm and emission at λ: 330 nm and 350 nm. The temperature ramp went from 15 °C to 80 °C at 1 °C/min.

For the time control experiments, the protein sample was diluted to a concentration of 0.5 mg/ml and 50 µl aliquots were dialyzed against 100 ml of buffer for 1 hour at 4 °C. Samples were kept on ice during experimental set-up. High sensitivity capillaries were loaded and the protocol was set to start at 15 °C, incubate for 10 min, then raise by 5°C, incubate 10 min then return to 15 °C. This was repeated in incremental cycles until a temperature of 40 °C was reached, as this results in completely unfolded protein. Continuous measurement of the back-reflected light was monitored as a proxy for particle size. All measurements were done in triplicate. Raw data was analysed and plotted in Prism 9 (Graphpad).

3.3.9 Dynamic light scattering

All experiments were performed on a Zeta sizer Nano series (Malvern instruments). The protein sample was diluted to a concentration of 0.5 mg/ml and 50 µl aliquots were dialyzed against 100 ml of buffer for 1 hour at 4 °C. Samples were kept on ice during experiment set-up. Measurements were performed in a quartz cuvette, that was rinsed with water and ethanol between samples and dried by a nitrogen stream. Measurements were performed at 15 °C with 20 µl sample in a low volume quartz cuvette. Each measurement consisted 120 s of initial

equilibration time followed by 50 runs of 5 s. During data analysis, each buffer ingredient was specified and adjust for viscosity using a calibration table.

Buffer	Ingredients
B1	50 mM Na-Phosphate pH 7.5, **0 mM NaCl**, 5mM TCEP
B3	50 mM Na-Phosphate pH 7.5, **50 mM NaCl**, 5mM TCEP
B3	50 mM Na-Phosphate pH 7.5, **150 mM NaCl**, 5mM TCEP
B4	50 mM Na-Phosphate pH 7.5, **300 mM NaCl**, 5mM TCEP
B5	20 mM Tris-HCl pH 8.0, **0 mM NaCl**, 5mM TCEP
B6	20 mM Tris-HCl pH 8.0, **50 mM NaCl**, 5mM TCEP
B7	20 mM Tris-HCl pH 8.0, **150 mM NaCl**, 5mM TCEP
B8	20 mM Tris-HCl pH 8.0, **300 mM NaCl**, 5mM TCEP

Table 6: Buffer composition used to screen for MmpA filament formation
Tris and phosphate buffers were tested and to test the effect of ionic strength the NaCl concentration was varied from 0 to 300 mM.

3.4 Results

The mupirocin PKS assembly line genes are large in size, *mmpD* even reaching about 20 kb. Both the combination of large size and high GC content of the mupirocin PKS genes cluster (~62-65 %, [206]) complicates the process of cloning it into an expression plasmid. The assembly line proteins *mmpA*, *mmB*, *mmpC* and *mmpD* modules 3-4 were cloned into the insect cell expression plasmid using multiple cloning methods and optimizing reaction conditions. Due to time constraints, only three out of these four cloned genes, *mmpA*, *mmpC* and *mmpD* modules 3-4, were expressed in insect cells. The purification strategy was optimized to obtain sufficient amounts of pure protein for the experiments.

Purified proteins were then studied in isolation with biophysical methods to gain insights in their oligomerisation behaviour and to find conditions in which they would potentially form megacomplexes, as has been hypothesised for *trans*-AT PKS [104]. In a next step, an isolated protein was structurally studied by cryo-EM to obtain structural insights into its organisation.

3.4.1 Establishing expression constructs for the mupirocin PKS assembly line proteins

Due to the high GC-content in the mupirocin gene cluster, the size (up to ~10 kb) of the assembly line protein coding genes and long primer overhangs used for Gateway cloning, every major step in the cloning strategy had to be optimized. Initially, genes of interest were PCR amplified from gDNA. Multiple PCR conditions were tested, such as different polymerases (Q5, Phusion, Herculase), different buffers (Q5, GC, HF or Herculase buffers) as well as a temperature gradient to find the optimal annealing T° (Figure 3.1). After extensive screening, GC buffer combined with GC enhancer and Phusion polymerase gave the desired PCR products. These were purified and

subsequently used in the BP reaction of the Gateway cloning protocol, to generate the entry clones.

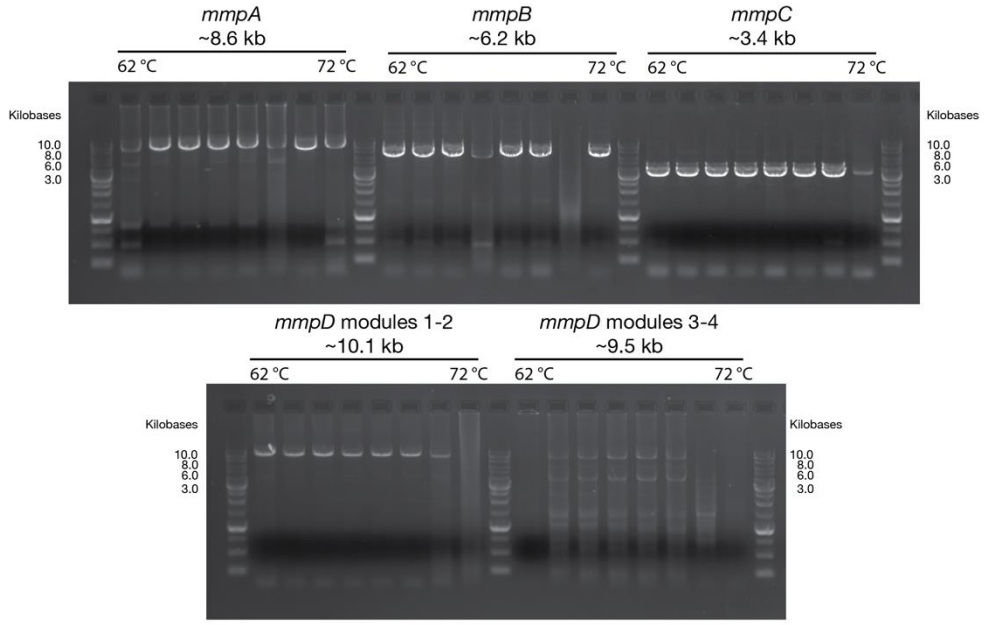

Figure 3.1: Agarose gel of the gradient PCR for the mupirocin PKS protein coding genes
PCR products of the mupirocin PKS assembly line proteins coding genes amplified with primers containing the attB-overhangs for Gateway cloning. Gene name and expected product size is indicated above the lanes.

A positive entry clone was obtained for *mmpC*, but for other constructs a high number of negative clones were obtained. Screening by colony PCR and sequencing revealed that some clones contained smaller fragments of the insert, suggesting potential recombination events, fragmentation of the gene or insertion of a small sized impurity. There were considerable difficulties in generating the entry clones for the remaining constructs. Extensive screening of BP reaction conditions was done by changing incubation conditions for the reaction. To facilitate the Gateway recombination reaction the plasmid was linearized and temperature and incubation time of the BP reaction were screened. Ultimately, entry clones were obtained for *mmpA* after incubation at 27 °C for 16 h using linearised pDONR221, *mmpB* and *mmpD* modules 3-4 were incubated at 30 °C with linearised and circularised pDONR221, respectively. These entry clones were subsequently used in an LR reaction into the expression plasmid. Positive colonies were confirmed by sequencing. For *mmpD* modules 1-2 no entry clones were obtained.

Due to considerable difficulties in establishing an expression construct, in parallel to optimisation of Gateway cloning, multiple cloning methods, such as TA-overhang cloning and Gibson Assembly of codon optimized gBlocks, were attempted. The pGEM®-T Easy TA-overhang cloning system uses blue-white screening as a method to identify positive clones. Colonies containing the

inserted gene in their MCS appear white when grown on LB-agar containing IPTG and X-gal, other colonies appear blue. Most colonies were blue, indicating they didn't contain the insert in their MCS, yet PCR screening of white colonies did not reveal the presence of the full-length insert.

The other cloning attempt consisted of Gibson assembly of codon optimized gBlocks. Each construct was divided into gBlocks of 2 to 2.9 kB in size containing overhangs with the next gBlock or expression plasmid. These were used in a 5 fragment Gibson assembly reaction using the expression plasmid backbone as the fifth fragment. The individual gBlocks could be successfully amplified, but there were considerable difficulties with inserting them in the plasmid backbone. In the meantime, after optimisation of the Gateway cloning protocol, expression constructs amplified from gDNA were ready and because of time constraints, these attempts were not continued.

3.4.2 Purification of mupirocin PKS proteins from insect cells

The mupirocin PKS assembly line proteins MmpA and MmpD modules 3-4 were expressed with an N-terminal His_{10}-myc-FLAG tag. Initially, a Ni-affinity purification was tested as this purification method provides a higher yield than FLAG-purifications, for the same resin volume used. SDS-PAGE and Coomassie stain analysis of eluted sample revealed the protein of interest present at a MW of 346 kDa for MmpD modules 3-4 (Figure 3.2). However, several intense protein bands are present in the 75 to 100 kDa range. These correspond to either impurities or degraded protein. To remove these impurities, a SEC was performed on a Superose 6 increase 10/300 GL column. The protein of interest has a size of 692 kDa in its homodimeric form. Therefore, we reasoned that these smaller protein contaminants could be separated in a SEC step. Concentrating the protein prior to injection revealed aggregation at a concentration of 0.587 mg/ml and concentrating was stopped. The SEC-profile reveals a small protein peak at the void volume of around 9 ml, followed by a broad peak (Figure 3.2). However, peaks were not intense and a large part of protein was lost, either due to proteolytic degradation or during the concentration step. SDS-PAGE showed that the protein of interest was present at 346 kDa, and a total of 0.35 mg of protein was obtained. However, this small amount was limiting for further biophysical and structural experiments.

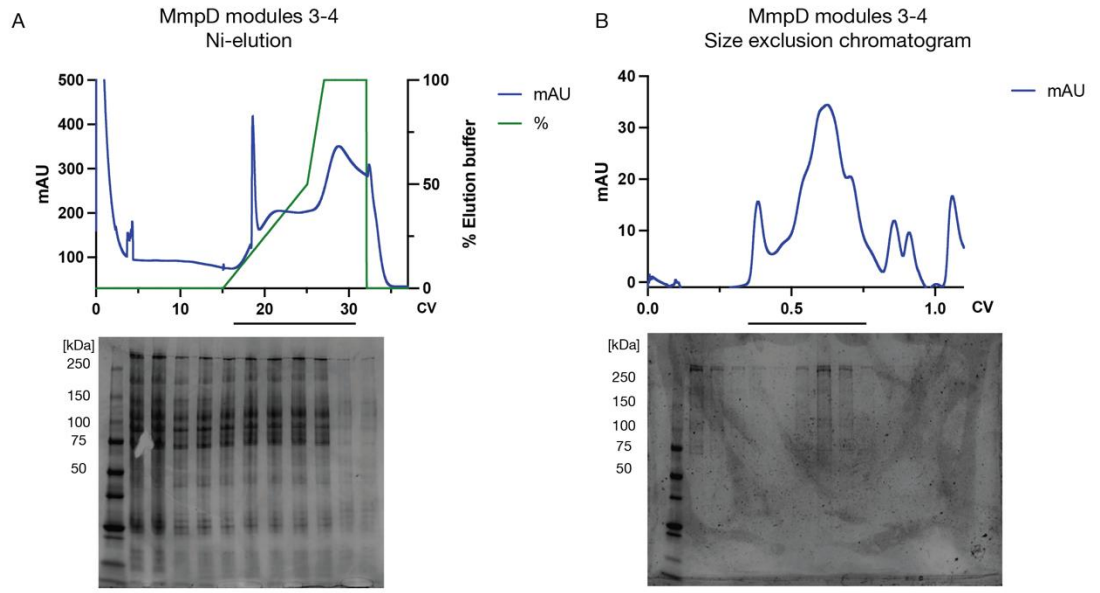

Figure 3.2: Initial purification of MmpD modules 3-4 by Ni-affinity chromatography
A: Elution from the Ni-column with an imidazole gradient. The UV signal at λ 280 nm is shown in blue, the green curve represents the gradient of Ni-buffer B. The Coomassie stained SDS-PAGE is shown below the chromatogram and loaded fractions are indicated with the black line. The protein of interest is present at 346 kDa. Several intense lower MW bands are present, indicative of aspecifically bound proteins or degraded protein. B: Size-exclusion chromatogram of MmpD modules 3-4 on a Superose 6 increase 10/300 GL column. The Coomassie stained SDS-PAGE is shown below the chromatogram and loaded fractions are indicated with the black line.

Due to low purification yields from Ni-purification as an alternative purification strategy, FLAG-purification was tested. Total yields after FLAG-elution were lower than after Ni-elution, and protein degradation was still detected. Due to previous issues concentrating the protein prior to SEC, we decided to omit SEC purification. The final purification strategy consisted of a single FLAG-affinity purification with three elutions (Figure 3.3). Total protein yields were 8 mg of MmpA, 1.5 mg of MmpC and 1 mg of MmpD modules 3-4.

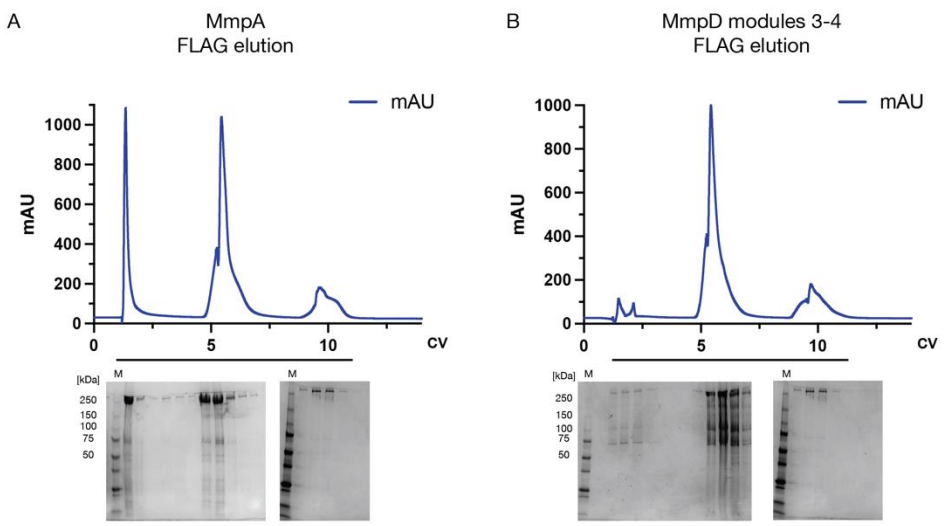

Figure 3.3: FLAG purification of MmpA and MmpD modules 3-4
FLAG elution for the assembly line proteins mmpA (A) and mmpD modules 3-4 (B). The UV signal at λ 280 nm is shown in blue. Three peaks represent the three consecutive elution steps and the corresponding Coomassie stained SDS-PAGE is shown below. Proteins of interest are present at 313 kDa for mmpA and 346 kDa for mmpD modules 3-4.

3.4.3 Thermal stability measurement of mmpA and mmpD modules 3-4

The thermal stability of the assembly line proteins MmpA and MmpD modules 3-4 was assessed by nanoDSF. Their melting temperatures were measured by thermal unfolding using the intrinsic protein fluorescence (Figure 3.4). Unfolding started at 30 - 31 °C, with a Tm of ~33 °C (Table 7). Here only a single unfolding event was detected, while for some multidomain proteins occasionally multiple thermal unfolding events can be observed (Figure 3.4). Interestingly, *P. fluorescens* is grown at 30 °C and at this temperature still produces mupirocin. However, based on thermal stability measurements, this is beyond the onset of unfolding and a fraction of the protein would be present in a partially unfolded state. Likely, inside the cell additional stabilising factors or interaction partners are present that increase the thermal stability of the proteins.

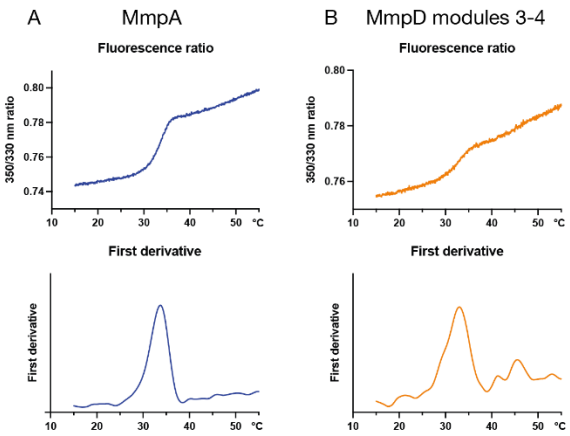

Figure 3.4: Thermal unfolding plots of mmpA and mmpD modules 3-4
The fluorescence ratio and first derivative of the curve is shown for mmpA (A) and mmpD modules 3-4 (B) in SEC buffer. Measurements done in triplicate and the average is plotted. Both proteins have a Tm of ~33 °C.

Sample ID	Tm	σ	Onset	σ
MmpA	33.54	0.01	30.14	0.12
MmpD modules 3-4	32.90	0.06	31.40	0.36

Table 7: Tm of MmpA and MmpD modules 3-4 assembly line proteins in SEC buffer
The Tm and onset of thermal unfolding are shown together with the σ. Measurements done in triplicate. Both proteins have a Tm of ~33 °C and onset of unfolding at 30-31 °C.

MmpA was screened for thermal stability in Tris and phosphate buffers with varying NaCl concentrations ranging from 0 to 300 mM to potentially stabilising conditions. Only small difference of 3 °C in thermal stability was measured between the conditions. To identify a trend between thermal stability and buffer conditions, the Tm was plotted as a function of buffer ionic strength (Figure 3.5). There was an inverse relationship between ionic strength and thermal stability. High NaCl concentrations decreased the thermal stability, likely to caused by salting-out of the protein [242].

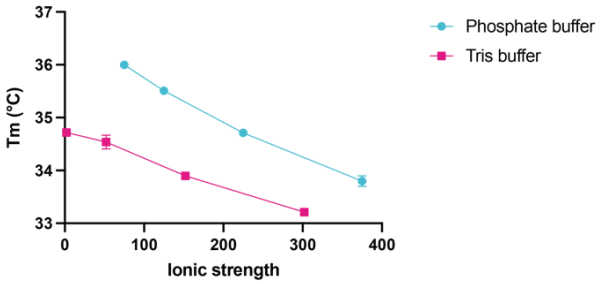

Figure 3.5: Melting temperature plotted as a function of ionic strength
As ionic strength of the buffer increases, protein Tm decreases. Likely a consequence of salting-out.

3.4.4 Screening for oligomerization by dynamic light scattering of mmpA

KSs of *trans*-AT PKS are hypothesised to form filaments mediated by their LINKS-motif [104]. In an attempt to maximise the chances of observing this filament formation, MmpA was chosen due to the presence of 3 KSs. A dynamic light scattering (DLS) screening experiment was done to measure particle sizes. In case filaments are formed, this would be observed as a stepwise increase in size of particles. The size volume distribution can subsequently be used to approximate the estimated MW of particles corresponding to this size. Protein concentration used was 0.5 mg/ml, as this corresponds to protein concentration where aggregation or oligomerisation was observed during protein purifications. Two different buffers, Tris and Phosphate, were tested and the NaCl concentration was varied from 0 to 300 mM. Data quality is indicated by the raw correlation data (autocorrelation function). At the starting time of the measurement, the autocorrelation function had a correlation coefficient above 0.8, indicating good data quality (Figure 3.6 A). Most particles had sizes approximating ~15 nm in every condition tested (Figure 3.6 B, C). Based on mass estimation, this size corresponds to homodimeric protein. Both Tris and phosphate buffers with 50 mM salt had particles with slightly increased sizes, yet still 98 to 99 % of particles were present in homodimeric form with a size ~15 nm. Increasing salt concentration led to mostly smaller particles being detected. To assess whether the slight increase in diameter of detected particles correspond to aggregated protein or potential oligomerisation, nanoDSF was performed and scattering of the back-reflected light and thermal unfolding was monitored.

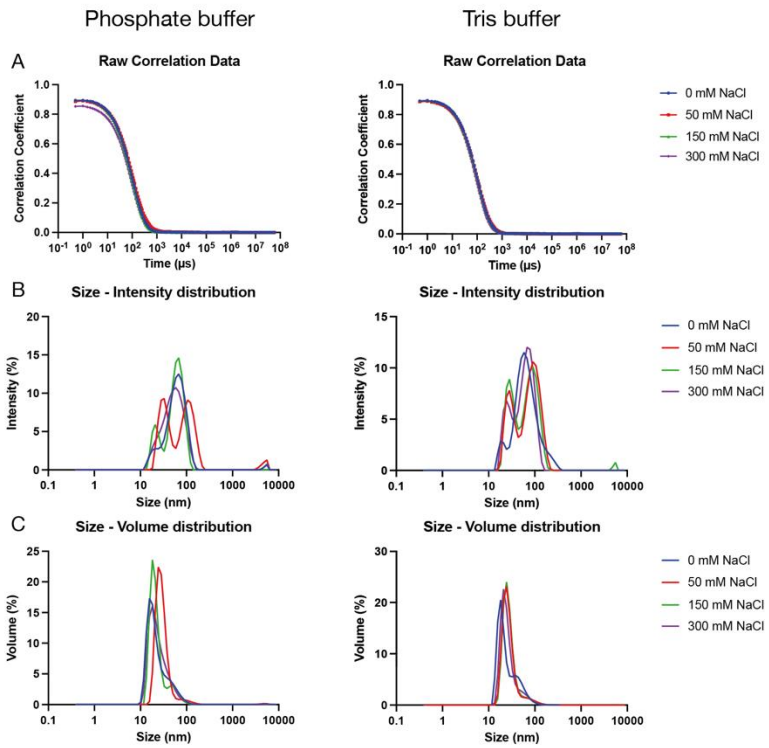

Figure 3.6: Dynamic light scattering measurement of MmpA
Filament formation of MmpA was screened in Phosphate or Tris-buffer with varying salt concentrations. A: Correlogram shown as a measure for data quality. B: size-intensity distribution plot, for 0 mM and 50 mM NaCl, there's an increase in scattering from larger particles. C: size-volume distribution plot showing the volume percent distribution of particles in solution. The majority of particles in the conditions tested have a size corresponding to ~15 nm.

3.4.5 Time series experiment of MmpA

To screen for filament formation, a time series experiment was done with incubation of the protein at different temperatures for 10 min intervals. Incubation temperatures were increased in 5 °C increments before returning to the set-point of 15 °C. This was done to assess the potential reversibility of filament formation. A final temperature of 40 °C was chosen as this is above the Tm of the protein and results in completely thermally unfolded protein. During incubation times, light scattering was measured as a proxy for particle size. An increase in scattering corresponds to an increase in particle size. The first derivative of the scattering curves gives information about conditions at which an increase in scattering and thus particle size is observed. Purified MmpA, incubated in phosphate or tris buffer at NaCl concentrations from 0 to 300 mM (Figure 3.7 A). An increase in scattering is observed for protein in 150 and 300 mM NaCl starting at 35 °C in both phosphate and tris buffers. In samples with lower salt concentrations, increased scattering is observed towards 40 °C. However, the fluorescence ratio shows that these temperatures the protein unfolds (Figure 3.7 B). Thus, the measured increase in scattering at these temperatures corresponds to aggregation as a consequence of thermal unfolding.

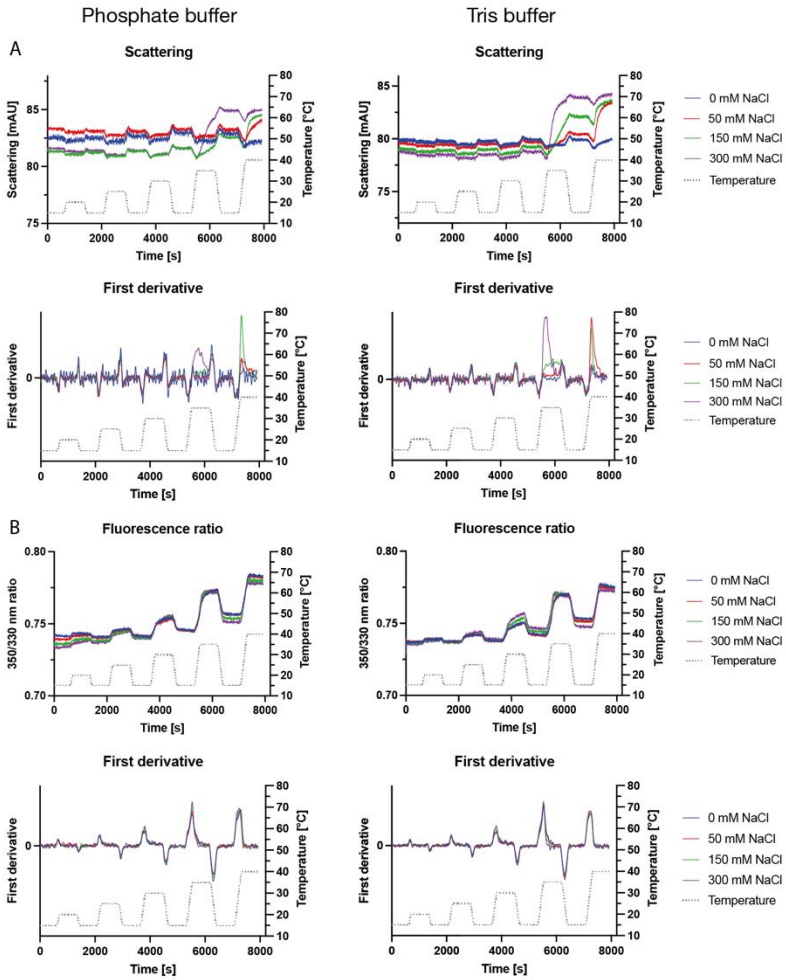

Figure 3.7: Time control experiment screening for filament formation of MmpA
A: Scattering curves for MmpA incubated at different temperatures in phosphate and tris buffer from 0 to 300 mM NaCl. Increase in particle size is indicated by increase in scattering. The first derivative of the curves is shown below. B: Ratio of fluorescence emission at λ 350/330 nm. Thermal unfolding is observed as an increase in the fluorescence ratio. Below, the first derivative is plotted. All measurements done in triplicate and the average is plotted.

3.4.6 Cryo-EM screening of mupirocin PKS proteins

Individual assembly line proteins of the mupirocin PKS were imaged by by cryo-EM to assess their organization. Grids were prepared for MmpA, MmpC and MmpD modules 3-4. In an initial session on a Talos TEM operated at 200 kV, grids and proteins are screened for optimal conditions consisting of thin ice, sufficient contrast and satisfactory particle distribution. From initial screening images, Quantifoil® Cu grids had better ice quality than the UltrAufoil® grids. Less defocus was required to obtain contrast due to slightly thinner ice as well. Micrographs showed particles that were clustered together, that can be interpreted as either oligomerization or protein aggregation. MmpD modules 3-4 was chosen for high-resolution dataset collection on a Glacios

TEM equipped with a K3 direct electron detector. The sample chosen for data collection was based on ice quality, particle distribution and available protein at the time.

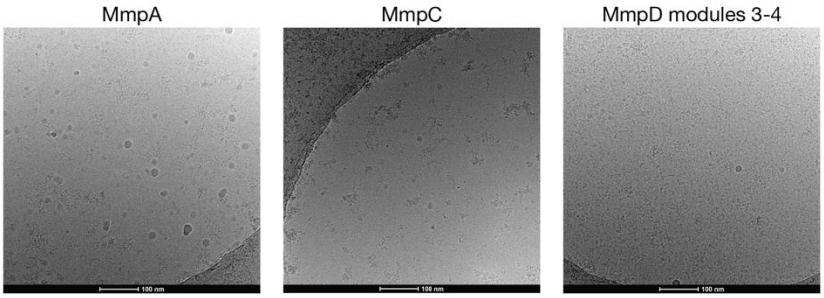

Figure 3.8: Cryo-EM grid screening for MmpA, MmpC and MmpD modules 3-4
Representative micrographs of MmpA, MmpC and MmpD module 3-4 screened on a Talos cryo-EM. Grids were screened for ice thickness, contrast and particle distribution.

3.4.7 Single Particle Analysis of the construct MmpD modules 3-4

Processing of the MmpD modules 3-4 dataset revealed the appearance of reference free 2D classes corresponding to the KS. However, other domains of the construct did not clearly appear during several rounds of classification (Figure 3.9 A). These KS classes were subsequently used as templates for picking and two approaches were done in an attempt to obtain structural information of MmpD modules 3-4. On the one hand, we focused on the KS domain that we recognised after initial 2D classification (Figure 3.9 B, C). This domain was used as a template to pick additional particles, and after multiple rounds of refinements a 3.44 Å resolution map of the KS was obtained based on the gold-standard method [243].

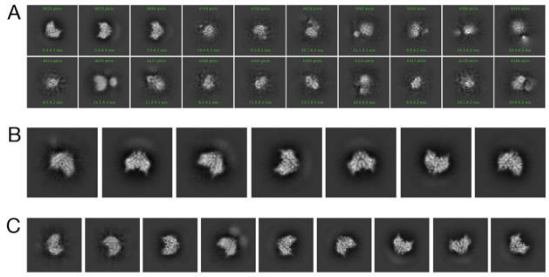

Figure 3.9: 2D classes of MmpD modules 3-4 focused on the KS
A: Subset of initial 2D classes obtained after blob picker. In the first three classes the shape of a KS is recognisable. B: Selected 2D classes of the KS for the template picker job. C: Selected 2D classes of the KS used in the subsequent refinements and *ab initio* 3D reconstruction.

This map allowed fitting of a crystal structure of a KS domain from the bacillaene PKS (PDB: 5ERB, ~55 % sequence identity). Despite the 3.44 Å resolution reconstruction, it was difficult to distinguish high-resolution features, such as sidechains, in the map. Likely a consequence of the presence of two KSs in the construct that align with each other during 2D classification. Both KSs

are very similar with sequence identities of 42 %, and the overall fold and secondary structure elements between KSs is conserved. Because the lack of extra features, the map obtained was an intermediate resolution of the KSs. The core of the KS fitted well into the map, and the overall shape of a KS is visible (Figure 3.10). However, the LINKS-helices were not visible, even if these regions are encoded in the DNA sequence and predicted to be present based on sequence alignments. Likely due to flexibility of these regions, suggesting that they only become ordered and visible after interaction with the LINKS-motif of the corresponding KS. Another reason can be structural differences between both KS that are aligned in this region. Sequences of the LINKS-motifs are less-well conserved than the KS core [104]. Variations in the sequence of this region can result in local structural differences between both KSs and thereby an unclear map.

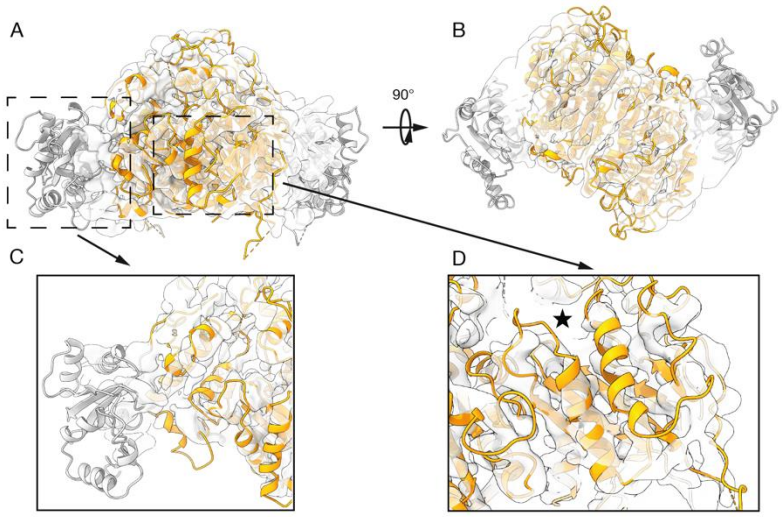

Figure 3.10: Model of the bacillaene PKS KS fitted into the KS map of MmpD modules 3-4
The KS core is coloured orange with the LD region in dark grey. The map is coloured light grey. A: Front view, B: Top view. C: Close-up on the LINKS-helices, which are not visible in the map. D: Close-up on the KS active site entrance (indicated with a star). There's no clear density for the helix on the right, which lines the KS active site entrance, assumed to be due to local flexibility.

Sorting both KS during 2D classification is challenging and requires high resolution or additional ordered features enabling us to distinguish between the KS. One of the strategies was to impose C2 symmetry to increase the number of picked particles, followed by 3D classification. This resulted in a map where the DH density appeared on top of the KS, however it is not clearly defined. The closest known DH structure of the rzxB PKS from *P. fluorescens* (PDB: 5IL6, ~33 % sequence identity) was fitted in the map (Figure 3.11). In this map, a weak signal for the KS LINKS domain is visible. However, high resolution features were still lacking, likely due to low amount of particles, flexibility or divergence between both KSs of the construct.

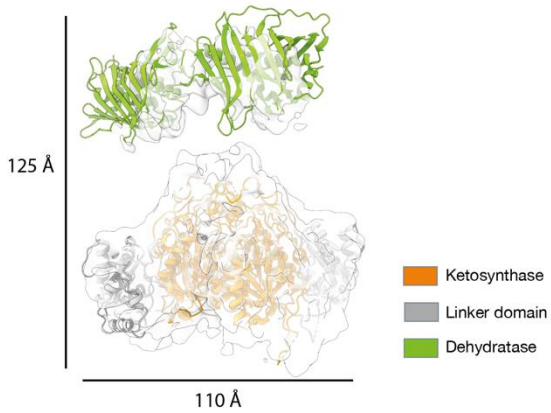

Figure 3.11: KS and DH domains fitted in the map of MmpD modules 3-4
The KS core is coloured orange with the LD region in dark grey, the DH is coloured green. The map is coloured light grey.

The other approach consisted of using the KS as a template for particle picking and a large window with a maximum of 512 pixels for particle picking in an attempt to extract additional domains during particle extraction. Different KS views were obtained in the 2D classification step, but there was a lack of or no clear additional density next to the KS visible. Several rounds of 2D classification and 3D *ab initio* reconstructions were performed searching for extra features around the KS. Further analysis revealed that some classes contained an additional density next to the KS, however due to extensive flexibility, no secondary structure features were visible (Figure 3.12).

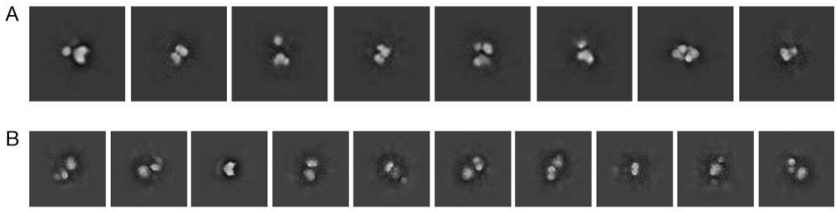

Figure 3.12: 2D classes of KS particles extracted with box size of 512 pixels
A: Selected 2D classes where additional densities are visible next to the KS. B: Selected 2D classes used in 3D *ab initio* reconstructions.

3.5 Discussion

ModPKS synthesise polyketides from small acyl-building blocks and function analogous to molecular assembly lines. The polyketide intermediate is handed over from one module to the next one, with each module performing a specific step of polyketide extension and modification until the final polyketide is released. Vectorial transfer between these assembly lines, required for product synthesis supports organisation of these enzymes into large clusters. In bacterial cells, megacomplex formation of the bacillaene *trans*-AT PKS has been shown [103]. The ability of other

trans-AT PKS, such as the mupirocin PKS, has been hypothesised based on the presence of the LINKS-motif flanking KSs [104]. Detailed information about the structural organisation of these large complexes in cells and *in vitro* is still missing.

Here, we attempted to reconstitute *in vitro* the assembly line proteins of the previously characterised mupirocin *trans*-AT PKS. First, expression constructs were established by cloning the PKS assembly line genes from gDNA into expression vectors. Gene size, high GC content and the repetitive nature of the sequences presented considerable challenges. This could be circumvented through custom gene synthesis of codon optimised constructs, reducing repetitive sequences and GC-content [234, 244, 245]. Additionally, the ORF could be separated into smaller DNA fragments to reduce its size, which are subsequently combined into the expression plasmid [246]. To clone these large, repetitive sequences, it was required to exceed traditional cloning protocols. Inspired by the field of synthetic biology, where large genomes are assembled from smaller DNA sequences [247, 248], we divided the genes of interest into smaller fragments and attempted to assemble them into the expression plasmid [221, 249]. Here, for the particularly difficult to clone constructs we tried to create the expression plasmids by Gibson assembly of codon optimised gBlocks. However, during synthesis time progress was already made cloning from gDNA.

Once expression constructs were established, proteins were expressed and purified. Again, this was a substantial challenge due to their large size, with full-length MmpD having a MW above 700 kDa in its monomeric form, but the active form is homodimeric. Therefore, the *mmpD* coding gene was split after module 2 into two fragments, containing modules 1-2 and modules 3-4, respectively. Purification yields had to be balanced with purity. Proteolytic degradation was observed on SDS-PAGE and aggregation when concentrating the protein above 0.587 mg/ml was observed. Different chromatography strategies were tested, but ultimately a single-step FLAG-affinity purification provided the best compromise between protein yield and purity. These findings are similar to previous attempts of expression and purification of full-length DEBS [81]. The authors reported considerable difficulties in generating the expression constructs, and ultimately were unable to express and purify full-length protein. Therefore, truncations were made at the boundaries between modules similarly to the approach tried here. In another approach, coexpression of chaperonins increased the yield and solubility of recombinantly purified DEBS [245].

Biophysical sample characterisation by nanoDSF revealed low thermal stability of the assembly line protein MmpA and MmpD modules 3-4, even in conditions corresponding to the cellular pH and NaCl concentrations. Protein oligomerisation was screened by varying the ionic strength and pH of the buffers. The assembly line protein MmpA was mainly present in its homodimeric state with no larger oligomers detected in the conditions tested. The range of ionic strengths tested could be too narrow or the protein concentrations used in these experiments could be too low to observe higher-order assemblies. For the bacillaene PKS, the assembly line proteins are present in 50 to 150 copies as shown by quantitative western blots [103]. Therefore, the requirement of high protein concentrations to mediate the LINKS-interactions in the cell is unlikely. Other reasons can be absence of potential binding partners or interactions that occur in the cell, but are missing *in vitro* due to studying the assembly line proteins in isolation.

Cryo-EM data revealed a disordered and flexible sample. Still, the KS domain was clearly recognised. To allow more conclusions from this dataset, both KSs need to be resolved separately. High resolution is required, because the core moiety of the KS is highly conserved and only at high resolution there will be visible structural differences enabling separate classification of both KSs. Next, because of their inherent flexible nature and large size, structural characterisation of PKS is challenging. The substrate needs to reach multiple catalytic sites which requires movement of the ACP and domain rearrangements [234, 235, 250]. This makes the task of aligning them to obtain high resolution structures even more challenging. Despite these challenges, cryo-EM reconstructions of PKS modules have been obtained [244] [234, 235]. One approach consists of adding stabilising Fab-fragments to the protein [251]. These have previously been used to solve cryo-EM structures of modPKS [234, 235, 244]. Still, even with Fab-fragments bound, the sample displayed flexibility and no clear densities were visible for every domain. Another approach to reduce flexibility is crosslinking. Yet here the drawback is that the protein is stably crosslinked in different conformational states and a mix of conformations may still be observed.

To set-up an assay to measure product turnover, the minimal components required for product synthesis must be combined. This consists of the PKS assembly line proteins, CoA-tethered polyketide extender units such as malonyl-CoA, the *trans*-acting AT protein required for loading the polyketide extender units onto the ACP (here MmpC) and NADPH. Product synthesis can be measured by identifying the released product by HPLC-MS, NADPH or precursor unit consumption. Substrate attachment to the ACP requires the presence of a phosphopantetheine (Ppant) cofactor. This can be identified with MS, or alternatively using fluorescent dyes that

specifically react with the Ppant moiety of the ACP (*i.e. holo*-ACP). When Ppant is present the protein becomes fluorescent [26, 252]. If the Ppant group is missing, it can be loaded using CoA and the promiscuous surfactin phosphopanetheine transferase (Sfp) and CoA [22]. Sfp has already been used to successfully load ACPs with the Ppant moiety, however its efficiency varies between different ACPs [253, 254].

Having an *in vitro* assembly line combined with detailed structural information on these enzymes can be used to generate mutants and to probe the effect of certain regions on product synthesis and turnover rates. This is an essential step to identify crucial determinants for product synthesis and regions involved in substrate specificity, substrate turnover and product yields. However, to the best of our knowledge, so far, no *trans*-AT modPKS assembly line has been reconstituted in an *in vitro* polyketide production assay. Here, we have achieved expression and purification of full-length or truncated *trans*-AT modPKS assembly line proteins from insect cells. Typically, only enzymatic domains or truncated constructs are recombinantly expressed in *E. coli* [81, 244, 246, 255]. Yet, only by studying full-length *trans*-AT modPKS, a complete picture of polyketide synthesis can be obtained. Clear understanding of these systems will enable us to re-engineer these enzymes to produce novel compounds with a potential impact on human health.

4 The mechanism of citrate-induced filamentation of human acetyl-CoA carboxylase 1

F. X. Delbart*[1], L. Anton*[1], M. Hunkeler[1,2], A. Hagmann[1], E. Stuttfeld[1], T. Maier[1]

* These authors contributed equally

1 Biozentrum, University of Basel, Switzerland

2 Current affiliation: Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA

Unpublished manuscript

Corresponding author:
Timm Maier:
Biozentrum, University of Basel
Klingelbergstrasse 50/70
CH-4056 Basel, Switzerland
Phone: +41 61 207 21 76
timm.maier@unibas.ch

Keywords: Acetyl-CoA carboxylase, citrate, filament, cryo-electron microscopy, x-ray crystallography

4.1 Abstract

Acetyl-CoA carboxylase (ACC) catalyses the first and rate-limiting step in fatty acid biosynthesis. Acetyl-CoA is carboxylated in two distinct and consecutive reactions resulting in malonyl-CoA, the committed substrate of fatty acid synthase. Eukaryotic ACC is a multienzyme, which contains both catalytic sites on a single polypeptide chain and assembles into a dimer. Over 50 years ago, it was observed that the feed-forward allosteric activator citrate leads to filament formation of ACC from higher eukaryotes (ACC-Cit). However, the mechanism of citrate-dependent assembly of ACC-Cit filaments remained an enigma. A 3.4 Å resolution cryo-EM reconstruction of ACC-Cit combined with a 2.0 Å crystal structure of the non-catalytic central domain (CD) enabled the identification of a positively charged pocket as the place of allosteric activation by citrate. Point mutants in this pocket disrupt citrate-dependent ACC activation and filament formation. Binding of citrate to the CD causes a cascade of domain rearrangements, creating a docking platform for the next ACC dimer to bind, resulting in ACC-Cit filament formation and full enzyme activation.

4.2 Introduction

Human acetyl-CoA carboxylase (ACC) catalyses the first and rate-limiting step in *de novo* fatty acid synthesis. ACC carboxylates acetyl-CoA to malonyl-CoA, which is used as a building block in fatty acid synthesis [129, 256]. The carboxylation of acetyl-CoA to malonyl-CoA occurs in two steps [119]. First, the biotin cofactor of the Biotin Carboxyl Carrier Protein (BCCP) is carboxylated by the Biotin Carboxylase (BC) domain. Next, the BCCP translocates to the Carboxyl Transferase (CT) domain, where the carboxyl group is transferred to acetyl-CoA producing malonyl-CoA. During the carboxylation reaction bicarbonate donates CO_2 and ATP is consumed [120, 121]. Besides these catalytic domains, eukaryotic ACCs also contain a central domain (CD), between the BC and CT domains (Figure 4.1 a). While the CD has no catalytic activity, it is serves as a hinge to position the BC and CT domains in the correct orientation for catalysis [122, 123]. The CD is further subdivided into four domains: CD_N (N-terminal), CD_L (linker), CD_{C1} (C-terminal 1), CD_{C2} (C-terminal 2).

Human ACC is expressed in two isoforms, ACC1 and ACC2 with a sequence identity of ~75 %. Both isoforms are functional as ~540 kDa homodimers, and ACC1 can polymerise into filaments in its most active form (Figure 4.1 b) [7, 127]. ACC1 is expressed in lipogenic tissues such as the liver and adipose [116]. In these tissues, malonyl-CoA produced by ACC1 is directly used for incorporation into fatty acids by fatty acid synthase (FAS). ACC2 is uniquely expressed in oxidative tissues. Initially, it was discovered in the rat heart and later also in skeletal muscles, where little to no *de novo* fatty synthesis occur [116-118]. In these tissues malonyl-CoA produced by ACC2 negatively regulates fatty acid oxidation by allosteric inhibition of mitochondrial carnitine palmitoyl transferase 1 (CPT1) [116].

ACC1 is shown to be overexpressed in liver, breast and prostate cancer tissues, both at mRNA level and protein level [157-160]. ACC1's dysregulation and the subsequent changes in lipid metabolism contribute to increased lipogenesis required for cancer cell growth. Knockdown of ACC1 by RNAi inhibits pancreatic cancer cell proliferation and induced cell death, making it a potential target for antineoplastic drugs [167]. Additionally, ACC1 is associated with type 2 diabetes, cardiovascular diseases, atherosclerosis and non-alcoholic fatty liver disease.

Because ACC catalyses the rate-limiting step in fatty acid synthesis, its activity is tightly regulated both transcriptionally and post-translationally. At the transcriptional level, ACC expression is regulated by Sterol Regulatory Element-Binding Protein 1 (SREBP1) and Carbohydrate

Regulatory Element-Binding Protein (ChREBP). This regulation is linked to the cellular nutritional state by mammalian target of rapamycin complex 1 (mTORC1). At the post-translational level, ACC is inactivated by phosphorylation by AMP-activated protein kinase (AMPK) and cAMP-dependent protein kinase (PKA). Residues Ser80, Ser1200 and Ser1216 are phosphorylated by AMPK and Ser80 and Ser1200 PKA [134, 135].The tumour suppressor BRCA1 binds to the phosphorylated Ser1263 of ACC through its C-terminal BRCT domain [142, 143], inhibiting dephosphorylation of Ser80 and locking the enzyme in the open inactive filament form (Figure 4.1 c) [123, 142, 143].

Almost 60 years ago it was shown that allosteric activator citrate significantly increased ACC activity [126]. Later, it was shown that citrate addition results in ACC filament formation [127]. Yet, it was only until in 2018, that the high-resolution structure of the activated ACC-citrate filament was solved, revealing insight into its inactive and active conformation [123]. While this reconstruction at 5.4 Å-resolution allowed unambiguous placement of crystal structures of all domains, identification of small molecules and their binding sites was not possible. Due to the overall moderate resolution and local flexibility of interface regions in the dimer, it was not possible to answer the crucial question of how citrate triggers ACC filament formation. Until now the mechanism with which citrate induces these massive conformational changes is unknown.

4.3 Results and discussion

4.3.1 Additional density in a positively charged pocket of ACC-citrate filaments

Previously, we have determined the cryo-EM structure of citrate-induced ACC filaments at a resolution of 5.4 Å using a single particle-approach (PDB: 62GD, [123]). Collecting additional data, as well as employing recent algorithms for the correction of aberrations introduced by the optical system, resulted in an overall resolution of 3.4 Å (Figure 4.1 d, e). The highest resolution is present in the highly ordered CT domain and the resolution decrease towards the edges of the map (Figure 4.6). The biotin moiety attached to a lysine in the BCCP is buried at the active site of the CT, which shows the potential of the map to visualize small molecules (Figure 4.7). 3D classification revealed that the angles between the dimers within the filament are not identical, which is in part responsible for the strongly decreased resolution of the CD (Figure 4.6, Figure 4.7). The CD lies at the interface between filament protomers and the difference in angles can be caused by the filament being inherently flexible and dynamic in this region (Figure 4.2 a, b). The varying angles could also represent static states of filaments that adopt distinct conformations.

The partial or full occupancy by citrate in filament protomers, can be an influencing factor as well. Citrate could directly bind somewhere in the CD domains, bind at the interface between two dimers in a filament or bind between two domains of one dimer to fix them in place. Citrate is a negatively charged molecule, suggesting a cluster of positively charged residues for binding.

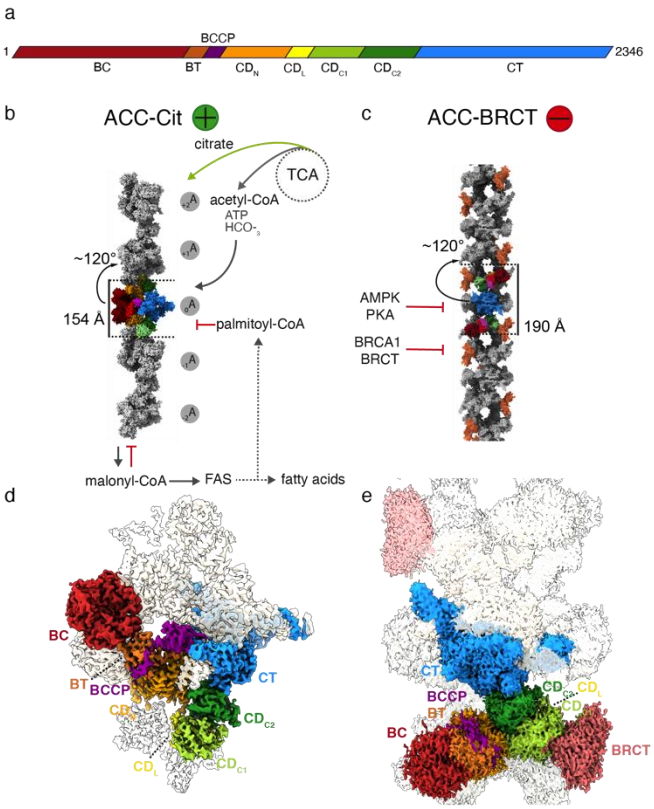

Figure 4.1: Filament formation of ACC upon activation and inhibition
a) Linear domain organization of ACC. b) ACC-Cit filament with one dimer in colour. Rise and twist are indicated. Regulation scheme of ACC by feedforward activators and feedback inhibitors. c) ACC-BRCT filament with one dimer in colour and all BRCT domains in colour with rise and twist indicated. Scheme on inhibition by phosphorylation and protein interaction partners. d) Cryo-EM map of ACC dimer upon allosteric activation by citrate at a resolution of 3.4 Å and contour level of 0.188. The map is colour coded according to the domains for one monomer, the other protomer is shown in grey. e) Cryo-EM map of ACC-BRCT filament at a resolution of 3.44 Å and contour level of 0.43. The map is colour coded according to the domains for one ACC monomer, the other protomers are shown in grey. The BRCT domains are shown in salmon.

Large conformational changes are necessary to go from an open inactive conformation to a closed active conformation. The CD has already been described as capable of undergoing large-scale hinge-like motions, where the CD_N can be at least 160° rotated from each other [122], to accommodate BC dimerization. In the cryo-EM reconstruction of ACC-Cit filaments, an extra density is present in a positively charged pocket of the CD (Figure 4.2 c). However, the resolution does not allow unambiguous placement of citrate. Therefore, we crystallised a truncated construct of the CD, consisting of CD_L-CD_{C1}-CD_{C2} in the presence of citrate (Figure 4.2 d).

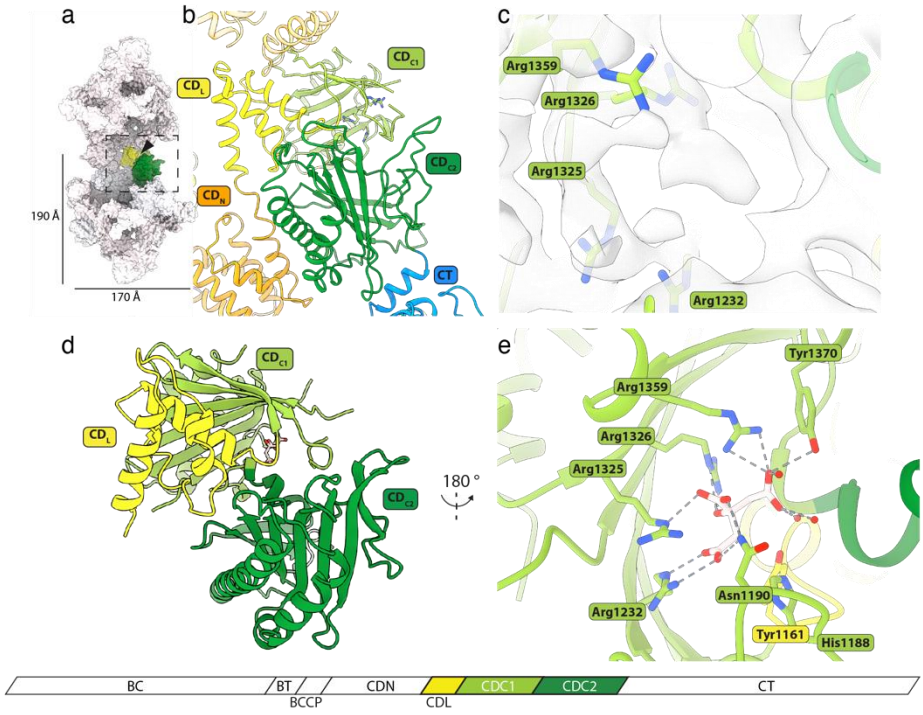

Figure 4.2: Citrate binds to a positively charged pocket in the CD$_{C1}$ subdomain of the CD
a) Surface view of two stacked ACC-Cit homodimers. The citrate binding site is indicated by an arrow in the colored CD surrounded by a box. b) Close up view on the CD with labelled domains. Arginines forming a positive pocket are shown as stick. c) Cryo-EM map of a positively charged pocket in the CD$_{C1}$ subdomain with additional density. Map shown at a contour level of 0.175. d) Crystal structure of CD$_L$-CD$_{C1}$-CD$_{C2}$ in its citrate-bound state. e) Citrate binding site of the X-ray structure with citrate bound to the same pocket as shown in c). Interacting residues are shown as stick. Domain architecture of ACC with the CD subdomains highlighted in colour.

In the 2.0 Å crystal structure, a density corresponding to citrate was present in a positively charged pocket, and citrate could unambiguously be built (Figure 4.8). Citrate binds to a conserved, surface exposed pocket and is coordinated by a cluster of arginines (Figure 4.2 e). Estimation of the citrate binding affinity for this domain by isothermal analysis resulted in an apparent Kd of 0.57 mM, which is within the range of cellular citrate concentrations (Figure 4.3 a, Figure 4.10) [144, 257].

4.3.2 Citrate affects a conserved pocket in the central domain

Six residues of the cluster are highly conserved across species, indicating evolutionary pressure to maintain this pocket. Next to Arg1232 and Arg1326, also Tyr1158, Tyr1161, Arg1325 and Pro1375 show high conservation (Figure 4.11). To assess the effect of citrate binding, Arg1232 and Arg1326 were mutated to alanines (R1232A/R1326A, Figure 4.3 b).

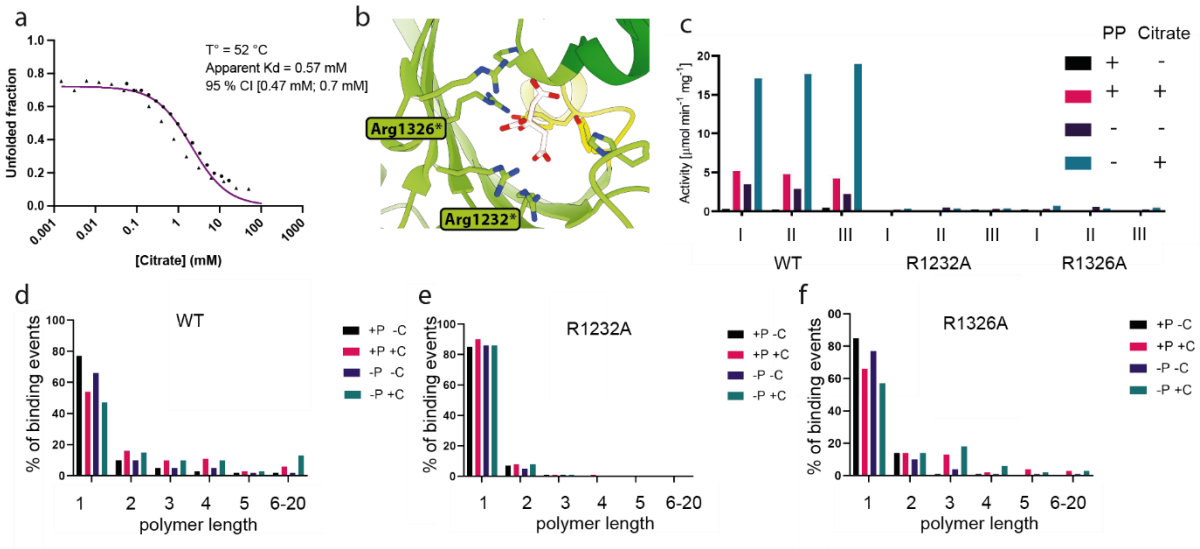

Figure 4.3: Point mutations impair citrate-dependent activation and filament formation
a) Isothermal analysis of nanoDSF thermal unfolding data to estimate the citrate dissociation constant for the CD_L-CD_{C1}-CD_{C2} domains. Dots and triangles represent two separate citrate concentration series. Temperature used for isothermal analysis, apparent Kd and 95 % confidence interval are indicated. b) Citrate binding site with conserved residues labeled and shown as sticks. Mutated residues are indicated in bold with an asterisk. c) ACC activity assay done in triplicate. Repetitions are labeled in roman numerals. d-f) Mass photometry results displayed as bar graphs. Polymer length refers to number of dimers. Color legend for graphs indicated at the bottom right. WT: Wild-type PP: phosphorylated.

To evaluate the impact of the mutations on citrate-induced activation, an assay measuring the incorporation of radiolabeled carboxyl groups into malonyl-CoA was used. In one experiment with three repeats, a reduction of activity for the mutants was observed (Figure 4.3 c). Dephosphorylated WT ACC in presence of citrate had the highest activities at 17.1-19.0 µmol min-1 mg-1. Progressively lower were the activities of phosphorylated ACC with citrate and dephosphorylated ACC without citrate, while phosphorylated without citrate had almost no activity. R1232A and R1326A had very low activity in all conditions, with between 0 and 0.7 µmol min-1 mg-1 (Figure 4.3 c). Qualitative assessment of the experiment is that mutants impair activity. Purified wild-type (WT) and mutant ACC run as a single band on an SDS-PAGE gel (Figure 4.12). Thermal shift assay of all proteins resulted in equivalent curve shapes of raw fluorescence. Melting points were calculated from the first derivative and were at 48°C for WT ACC, 47.5°C and 48.5°C for R1232A and 46 and 45 °C for R1326A. The two melting points refer to phosphorylated and dephosphorylated sample, respectively. Additionally, WT and mutants are all present as more than 70 % dimers in the inactivated state as can be seen by mass photometry Figure 4.14. The mutations do not affect expression and folding of the protein.

To investigate whether the mutations also affect filament formation we used mass photometry to assess the particle size distribution based on single molecule measurements of ACC mutants in

the presence of citrate and phosphorylation. The ACC dimer has a molecular mass of about 540 kDa and polymer length is indicated by calculated number of dimers per particle (Figure 4.3 d-f). WT ACC displays 77 % dimers in the phosphorylated condition without citrate compared to 47% dimers in the dephosphorylated condition with citrate. All conditions show progressively lower percentage of binding events as polymer length increases until polymer lengths of 6-20 dimers. Phosphorylated and dephosphorylated ACC in the presence of citrate show 6 and 13% binding events at high polymer lengths, respectively (Figure 4.3 d). R1232A shows 85-90 % of single dimers, 5-8% with two dimers and <1 % for other polymer lengths in all conditions (Figure 4.3 e). The mutation of R1232 to alanine appears to completely abrogate citrate-induced filament formation. For the R1326A mutation, 18 % of dephosphorylated citrate bound ACC are present as polymers of 3 dimers compared to the 10% in WT in the same condition. The mutant also assembles polymers of 6-20 dimers length, but only 3% of the total (Figure 4.3 f). This could hint at filament formation being impaired but not completely abrogated for R1326A as the mutant could still assemble shorter filaments (for all plots see Figure 4.14, Figure 4.15).
In summary, the results of the activity and filamentation assays demonstrate that mutant R1232A is indeed insensitive to activation by citrate. R1326A shows reduced sensitivity to citrate but is still able to assemble shorter polymers. Likely the construct can still accommodate citrate through interactions with the adjacent R1325 residue.

4.3.3 Loop repositioning in the CD leads to domain rearrangements

The key factor constituting ACC activity is dimerization of the BC domains allowing the BCCP to reach both catalytic sites [123]. In yeast ACC as well as human ACC, the CD domain acts as the hinge allowing for flexible movement of the BC domains [122, 123, 258]. To analyse the effect of citrate binding on structural rearrangements, the open ACC-BRCT dimer and closed ACC-Cit dimer are overlaid (Figure 4.4 a-c). Comparison of CD domains of ACC-Cit and ACC-BRCT show highly similar folds for CD subdomains, but their relative arrangement changes (Figure 4.4 b). The CD_N is an alpha-helical domain with 11 helices with a four-helix bundle and a helical hairpin. The CD_L is a linker domain containing a 4-helix bundle. The CD_{C1} and CD_{C2} adopt an alpha-beta fold with a six stranded β-sheet with two long helices on one side. CD_{C2} is C-terminally extended by a β-strand and a β-hairpin (Figure 4.13). The six stranded beta-sheet (β1-6) of CD_{C1} contains the positively charged pocket coordinating citrate. When overlaying the CD_{C1} domains of ACC-Cit and ACC-BRCT filament there is a 33° turn of the CD_{C2} relative to the CD_{C1} (Figure 4.4 a). This rotation occurs at the small helix after the loop connecting the domains. Binding of citrate interferes with the helix and repositions it inducing the rotation (Figure 4.4 a). CD_{C1} and CD_{C2} rotating against each other changes the position of the CD relative to CT. CD_N is now close to the

CT and BT domains and needs to rotate by 37° to close the dimer (Figure 4.4 b, c). The BC domains would be close enough to each other to dimerize. Repositioning of the whole CD subsequently exposes the interface needed for filament formation (Figure 4.4 d).

In yeast ACC (yACC), phosphorylation of Ser1157 (pSer1157) by Snf1, a homologue of AMPK, leads to a stark reduction in activity [152, 259, 260]. Structural analysis showed that pSer1157 is located in a positively charged pocket of the CD between CD_{C1} and CD_{C2} and is coordinated by Arg1173 and Arg1260 [122]. This site corresponds to the same pocket where citrate binds in human ACC (Figure 4.4 e). Closer inspection of the same positively charged cluster in the full length yACC X-ray structure reveals a large, unaccounted for electron density inbetween the arginines, which could potentially correspond to citrate (Figure 4.9). This cluster is highly conserved among eukaryotes, including human ACC (Figure 4.11). Residues Arg1232 and Arg1326 in human ACC are Arg1173 and Arg1260 in yeast, respectively.

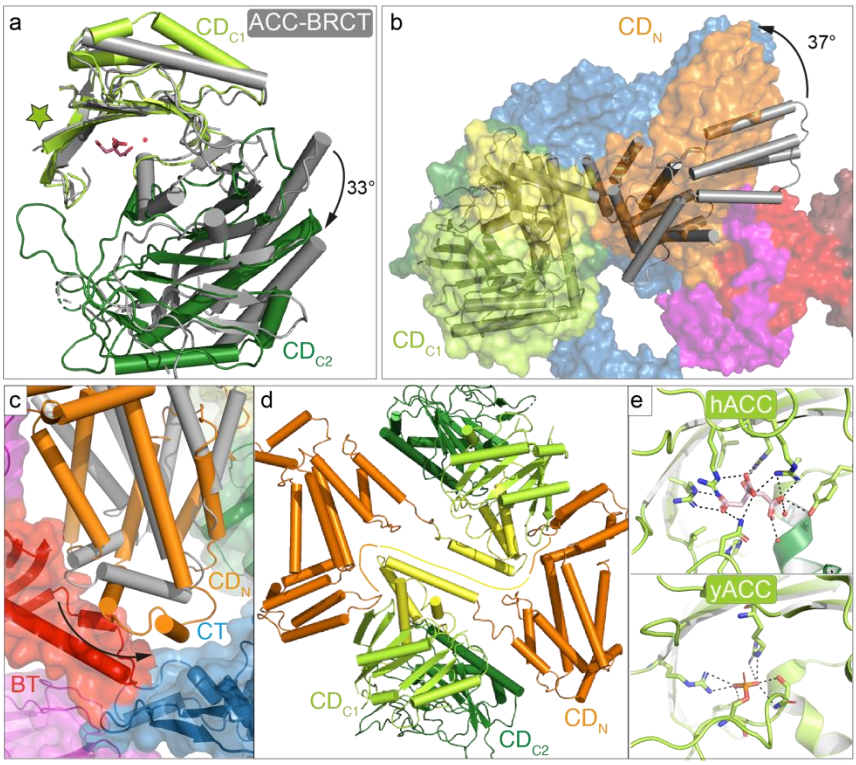

Figure 4.4: Domain rearrangements induced by citrate binding
a) ACC-Cit (color) and ACC-BRCT (grey) CDC1 and CDC2 aligned at CDC1. Shown in cartoon representation with cylindrical helices. Repositioned helix indicated with a red star. Rotational movement shown with arrow. b) Transparent ACC-Cit protomer with surface representation aligned at CDC1 with ACC-BRCT CD in grey cartoon style with cylindrical helices. Domain rotation indicated with arrow. c) Close up of CD_N with cylindrical helices and the interaction with CT and BT domains as transparent surface rendering. ACC-BRCT is in grey. d) Interface between two CD domains in filamentation competent state indicated with colored lines. e) The citrate binding site of human ACC (hACC) corresponds to the phosphoserine1157 binding site from yeast ACC (yACC, PDB: 5I6E). Interactions are shown as dotted lines with residues as sticks.

Citrate binding to the CD leads to conformational changes, creating a docking platform for another ACC dimer in a similar conformation to form filaments (Figure 4.5). It is also possible that citrate binding alone does not yet lead to a completely closed dimer but pushes the conformational change far enough so that another ACC dimer can bind and complete the dimerization.

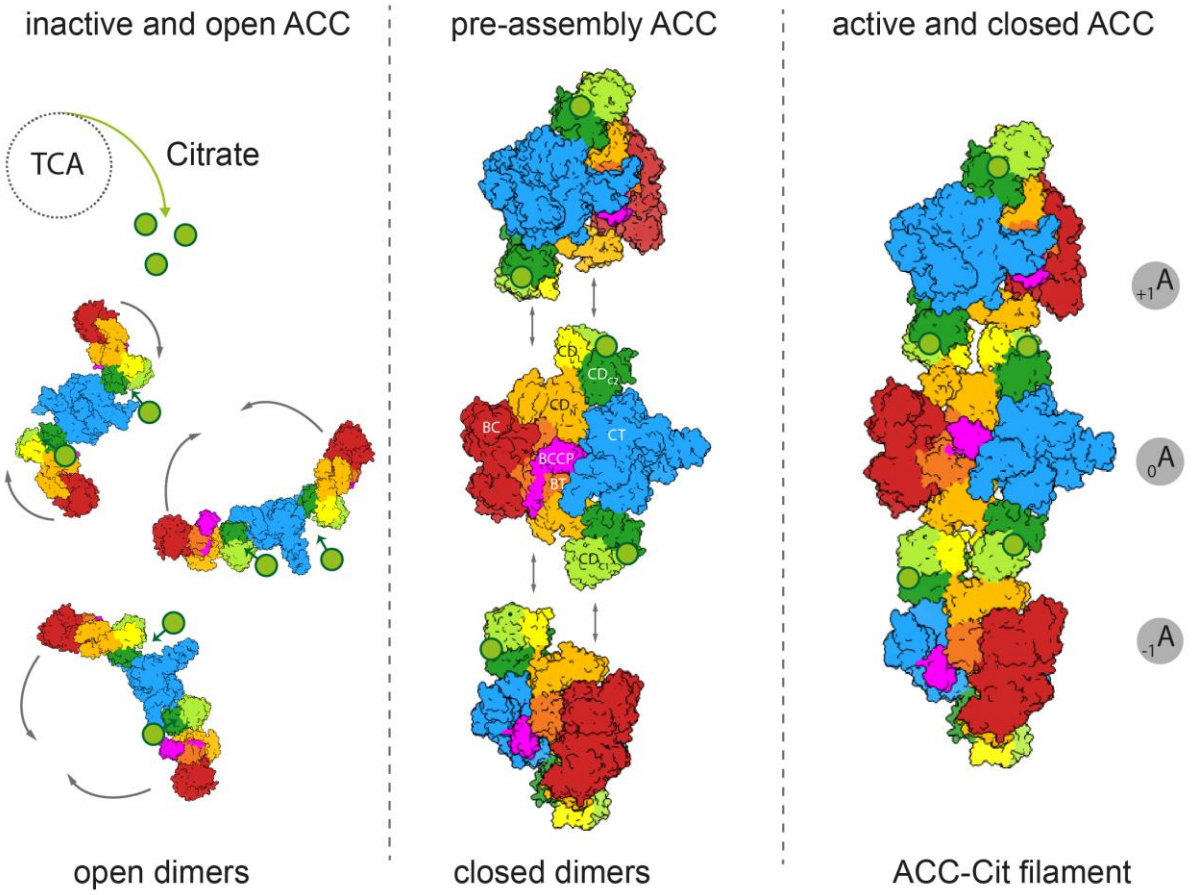

Figure 4.5: Mechanism of citrate-induced filament formation
Citrate allosterically activates ACC by binding to the conserved pocket in the CD. This causes large scale domain rearrangements making ACC dimers competent to stack into a filament. The filament locks ACC in a perpetually active and closed conformation.

Based on the higher-resolution cryo-EM data, mutations of the interface can be designed to disrupt filament formation revealing the contributions of filament formation to the increased activity. To assess the effect of citrate on ACC activation *in vivo,* site-directed mutagenesis of the binding site is possible. Monitoring cell survival and growth under conditions requiring endogenous fatty acid synthesis would reveal the importance of citrate activation in ACC regulation. Additionally, targeting the citrate binding site with inhibitors offers potential to design drugs preventing full enzyme activation. To gain deeper insights into the dynamics of the CD region, atomistic molecular dynamics simulations can reveal the conformational changes occurring in solution. Changes in the CD_{C1} and CD_{C2} domains in the presence and absence of

citrate can be simulated and provide insights into dynamics of the CD and be extrapolated to the complete ACC dimer.

4.4 Conclusion

Here, we reveal the citrate binding site and its mechanism of inducing ACC filament formation and full activation. Improved cryo-EM reconstructions from 5.4 Å to 3.4 Å enable the visualization of a conserved pocket in the non-catalytic CD region of human ACC filaments. A 2.0 Å crystal structure of the CD_L-CD_{C1}-CD_{C2} domain in the presence of citrate reveals its binding site in a positively charged cluster of the CD. Mutations of residues forming the cluster lead to abrogated activity and reduced filament formation in presence of citrate. In SceACC, this positive cluster can be tightly bound by a pSer on a loop, which leads to inhibition of SceACC activity. pSer prevents movement of the CD_{C1} and CD_{C2}, adopting a conformation very similar to CD_{C1} and CD_{C2} in the inactive ACC-BRCT filament. Citrate binding in the same pocket could change the position of the adjacent helix. The repositioning of the helix then causes a cascade of domain rearrangements. The resulting ACC dimers adopt a closed and active conformation, exposing the CD binding interfaces compatible with filament assembly (Figure 4.5)

After 50 years of research into the mechanism of citrate-induced filament formation, we present the first model for the assembly of ACC. This mechanistic model allows for the design of future studies with the goal of furthering our understanding of enzyme regulation through oligomerization and the potential of the allosteric site and binding interfaces as a drug target.

4.5 Author contributions

F.X.D solved the X-ray structure, processed the cryo-EM datasets and performed biochemical and biophysical assays. L.A. collected and processed cryo-EM data and performed biochemical and biophysical assays. M.H. cloned and purified ACC WT protein. A.H. collected cryo-EM data sets. E.S solved the X-ray crystal structure and built a model of human BT-CD. T.M. conceptualized the project. F.X.D, L.A. and T.M. wrote the manuscript and prepared figures.

4.6 Acknowledgements

We would like to thank the staff from the Swiss Light source beamline and Biozentrum BioEM facility for their excellent support. Cryo-EM data processing calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing center at the University of Basel. F.X.D,

L.A. and A.H. acknowledge funding from the Fellowships for Excellence program sponsored by the Werner-Siemens Foundation and the University of Basel.
We would like to thank the Biozentrum proteomics core facility for their support and help with measurement and analysis of phosphorylation states of ACC. We would also like to thank Katharina Häussermann from Refeyn and all of Refeyn staff for the demonstration and use of the instrument.
Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at UCSF.

4.7 Materials and Methods

4.7.1 Cloning and protein purification

The published ACC vector was used for site directed mutagenesis using primers containing the desired mutation followed by blunt end ligation. Primers used to clone R1232A mutation: forward (fw) TCACTCCACCTTGTCAGGCAATGGGCGGAATGGTCTC, reverse (rv) GAGACCATTCCGC CCATTGCCTGACAAGGTGGAGTGA; R1326A mutation: fw GTTGACCATGGGATCCGGGC ACTTACTTTCCTGGTTGCA; rv TGCAACCAGGAAAGTAAGTGCCCGGATCCCATGGTC AAC. Mutants were confirmed by sequencing.
Purification, *in vitro* biotinylation and dephosphorylation of full-length ACC was conducted as described [123]. Biotinylation was confirmed by using streptavidin antibody from abcam. Dephosphorylation was confirmed by mass spectrometry (Supplementary method 4.9).
A truncated construct of hACC consisting of CD_L-CD_{C1}-CD_{C2} (aa 1090-1590) was PCR amplified using hACC as a template with the following primers: fw GGGGACAAGTTTGTACAAAA AAGCAGGCTTAGAGAATCTGTATTTCCAGGGGTTGCCATCATATGAGCTTCGCC; rv GGGGA CCACTTTGTACAAGAAAGCTGGGTGTTAGTCTTTGGTCACATATGGAGTATTG and cloned into a pAceBAC2 expression vector (Geneva Biotech, Geneva, Switzerland) modified with a Gateway cassette with an N-terminal His_{10}-Myc-FLAG tag according to the manufacturer's instructions. Baculovirus generation and protein expression was performed in *Sf*21 cells using to the Bac-to-Bac expression system and the manufacturer's protocol was followed (Thermofischer Scientific). Cells were harvested by centrifugation (800 g, 20 min, 8 °C) 96 h post infection and stored at -80 °C. Cells were lysed by sonication in lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 40 mM imidazole, 20 mM Citrate, 5 % glycerol, 5 mM β-mercapto ethanol (β-ME), 2 mM MgCl2, 2 mg of DNaseI, protease inhibitors) and the lysate was cleared by ultracentrifugation.

Soluble protein was purified using Ni-NTA (Ni Sepharose 6 Fast Flow, Cytiva) and eluted in Ni-buffer B (50 mM Tris-HCl pH 8, 150 mM NaCl, 500 mM imidazole, 20 mM Citrate, 5 % glycerol, 5 mM β-ME, protease inhibitors), followed by tag cleavage with TEV-protease overnight at 4 °C in TEV buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 20 mM Citrate, 5 % glycerol, 5 mM β-ME). TEV-protease was removed by orthogonal Ni-NTA purification before size exclusion chromatography on a Superdex200 16/600 column in size exclusion buffer (SEC buffer, 50 mM Tris-HCl pH 8, 150 mM NaCl, 20 mM Citrate, 5 % glycerol, 5 mM DTT). Purified protein was concentrated to 10 mg/ml, flash frozen in liquid nitrogen and stored at -80 °C until further use. For affinity measurements the protein was purified analogously without citrate.

4.7.2 Sample preparation and cryo-EM data collection

ACC WT protein was dialyzed overnight against a buffer containing 50 mM Hepes/KOH pH 7.5, 10 mM K_3citrate, 0.1 mM EDTA, 5 mM β-ME. Protein was diluted in the buffer without β-ME to a concentration of 400 µg/ml and 17 µM DDM was added. C-Flat 1.3/1.2 200 mesh grids from Electron Microscopy Sciences were prepared using a FEI Vitrobot MarkIV (Thermofischer Scientific) with 4 µl of protein solution (4°C, 100% humidity, double blotting paper, 10 sec wait time, 3 sec blot time). Samples were imaged using an FEI Titan Krios equipped with a Gatan image filter (Quantum-LS GIF, 20 eV zero loss filtering) and a post-GIF K2 summit direct electron detector (Gatan). Movies were recorded at 300 kV with a pixel size of 1.05 Å and 50 frames at ~1 e^- $Å^{-2}$ per frame (resulting in a total dose of 50 e^- $Å^{-2}$). Sample preparation and collection parameters of the previous 13'671 movies (here called datasets 1-4) can be found in publication [123].

4.7.3 Cryo-EM Data processing

Collection parameters of the previous 13'671 movies (datasets 1-4) can be found in publication [123]. For the new data collection 3'746 movies were recorded using SerialEM (dataset 5) [261]. Processing for all datasets (DS) was done in cryoSPARCV2 and V3, the same strategy was used unless indicated otherwise [241]. Beam-induced motion was corrected using patch motion correction and patch CTF was used to correct for differing defocus values across one micrograph. Poor quality micrographs were discarded at this step. Particles were picked using a 2D template made from previous ACC-Cit reconstructions and employing the cryoSPARC template picker. The picked particles were classified into 2D classes and low-quality particles were sorted out, which resulted in a total of 467'027 useful particles (DS1: 146'013; DS2: 154'994; DS3: 43'698; DS4: 30'153; DS5: 92'169). Previous ACC-Cit reconstruction (EMD-4342) was as used as a starting model for refinement. DS1-4 were strongly affected by optical aberrations of the

microscope [123]. In recent publications, algorithms for the correction of the symmetrical and asymmetrical aberrations and magnification anisotropy were described [262, 263]. These algorithms were implemented in cryosparcV2.1 [264]. Through iterative local and global CTF refinements the resolutions of the 4 datasets were improved by 0.2-0.8 Å (Figure 4.6). DS1-5 were merged together and refined as one data set. 3D classifications showed heterogeneity in the angles of filaments (Figure 4.7). Number of particles in the separated classes were too low, to separately calculate another reconstruction. Using all particles non-uniform refinement and local masking of one dimer, an overall resolution of 3.4 Å was reached, based on the FSC 0.143 threshold criterion [243].

4.7.4 Cryo-EM model refinement

Phenix_real_space_refine (version 1.20) was used to refine the model against the improved map and the model was built in the new map using coot [265]. Citrate was manually place into the binding pocked and oriented based on the best fit into the density and its interactions with sidechains based on the crystal structure. This new model containing citrate was again refined in phenix.

4.7.5 Protein crystallization, X-ray data collection and structure determination

The CD_L-CD_{C1}-CD_{C2} domain construct in SEC buffer with of 20 mM citrate was crystallized at 19 °C using sitting-drop vapour diffusion in 2-drop MRC plates (Swissci) with a 1:1 protein:reservoir ratio (8 %v/v Tacsimate, 20 %w/v PEG 3350). Crystals grew to their maximal size in 10 days, were cryoprotected in mother liquor containing 25 % ethylene glycol and vitrified in liquid nitrogen. Diffraction data was collected at the PXIII (X06DA) beamline of the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) equipped with an Eiger 16M detector (Dectris) at a wavelength of 1 Å at 100 K. Raw diffraction images were processed with XDS [266] and scaled with aimless [267]. Phases were obtained with PhaserMR using a BT-CD model (PDB:5I87) [122] modified to contain only the CD_L-CD_{C1}-CD_{C2} subdomains [268]. Model building was done in coot [265] and refinements were done with Phenix [269]. Model validation was done with Molprobity [270].

4.7.6 Differential scanning fluorimetry with citrate concentration series

CD_L-CD_{C1}-CD_{C2} at a final concentration of 10 µM in SEC buffer without citrate was mixed with citrate (mono-Na citrate, diluted in SEC buffer) concentration series ranging from 100 mM to 1.53 µM. Thermal unfolding was monitored by differential scanning fluorimetry (DSF) based on internal Trp fluorescence using a Prometheus NT.48 machine (NanoTemper) controlled by the

PR.ThermControl software. High-sensitivity capillaries were used and samples were heated with a temperature gradient of 1 °C/min from 20 °C to 95 °C. All measurements were done in triplicate. To obtain the binding affinities from nanoDSF, the data was analysed according to the method of Bai *et al.* [271] using FoldAffinity [272]. Briefly, the unfolded protein fraction for different citrate concentrations was calculated at an isothermal slice by fitting the heat capacity change during unfolding across the different temperature ranges. Next, a 1:1 binding model was fit to the unfolded fractions for each citrate concentration and used to obtain the apparent Kd value (Figure 4.10).

4.7.7 Activity assay

The catalytic activity of ACC was measured by following the incorporation of radioactive ^{14}C into acid-stable non-volatile product. The assay was preformed according to protocol from [123]. Measurement was carried out in three replicates and catalytic activities were calculated using a standard curve derived from measurements of varying concentrations of $NaH^{14}CO_3$ in reaction buffer from [123].

4.7.8 Mass Photometry

In mass photometry the size of the molecule or complex is determined by the scattered light of a molecule at the moment of impact on a glass slide. The scattered light correlates linearly with mass and can therefore be used to weigh complexes [273]. Impact on the glass slide is called a binding event and they can be separated by their different masses. The limit of accuracy of the instrument is around 3 MDa, therefore from 6 dimers onwards binding events are still detected but cannot be properly separated. The highest recorded masses were around 10 MDa, corresponding to 20 dimers in one molecule so all binding events from 6 to 20 dimers are reported in one category. WT and mutant ACC were diluted to 1 µM in buffer containing 20 mM bicine pH 8, 150 mM NaCl, 5% Glycerol, 5 mM TCEP and 8 mM K_3citrate if indicated. Samples were incubated for 30 min-2h and the diluted 1:50 for measurement. Except dephosphorylated WT with citrate, which could only be diluted 1:10. This was to get comparable amounts of binding events per sample. To enable comparisons, the number of binding events should be similar between different samples. Too many binding events crowd the glass slide and limit accuracy of the detection, too few binding events will limit the amount of information you get out of the sample. The Refeyn One^{MP} instrument was used for measurements [273]. The presented data is from a single experimental set up (see Figure 4.14, Figure 4.15).

4.7.9 Data analysis and presentation

Structural figures were prepared using ChimeraX [274] and Pymol [275]. Graphs were made with GraphPad Prism version 9, GraphPad Software, La Jolla California USA, www.graphpad.com.

4.8 Supplementary information

	hACC CD_L-CD_{C1}-CD_{C2}
Wavelength (Å)	1
Resolution range (Å)	48.08 - 2.003 (2.075 - 2.003)
Space group	P 21 21 2
Unit cell	91.179 101.843 56.588 90 90 90
Total reflections	72109 (7030)
Unique reflections	36067 (3514)
Multiplicity	2.0 (2.0)
Completeness (%)	99.62 (98.35)
Mean I/sigma(I)	11.33 (1.06)
Wilson B-factor	39.02
R-merge	0.03178 (0.7844)
R-meas	0.04494 (1.109)
R-pim	0.03178 (0.7844)
CC1/2	1 (0.514)
CC*	1 (0.824)
Reflections used in refinement	36049 (3514)
Reflections used for R-free	1766 (159)
R-work	0.2419 (0.3658)
R-free	0.2613 (0.3733)
CC(work)	0.948 (0.713)
CC(free)	0.923 (0.620)
Number of non-hydrogen atoms	3565
macromolecules	3398
ligands	18
solvent	154
Protein residues	413
RMS(bonds)	0.003
RMS(angles)	0.57
Ramachandran favored (%)	98.02
Ramachandran allowed (%)	1.98

Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	1.77
Average B-factor	59.47
macromolecules	59.5
ligands	92.59
solvent	56.17
Number of TLS groups	7

Table 8: X-ray data collection and refinement statistics for the CD_L-CD_{C1}-CD_{C2} construct
Highest resolution shell is shown in parentheses.

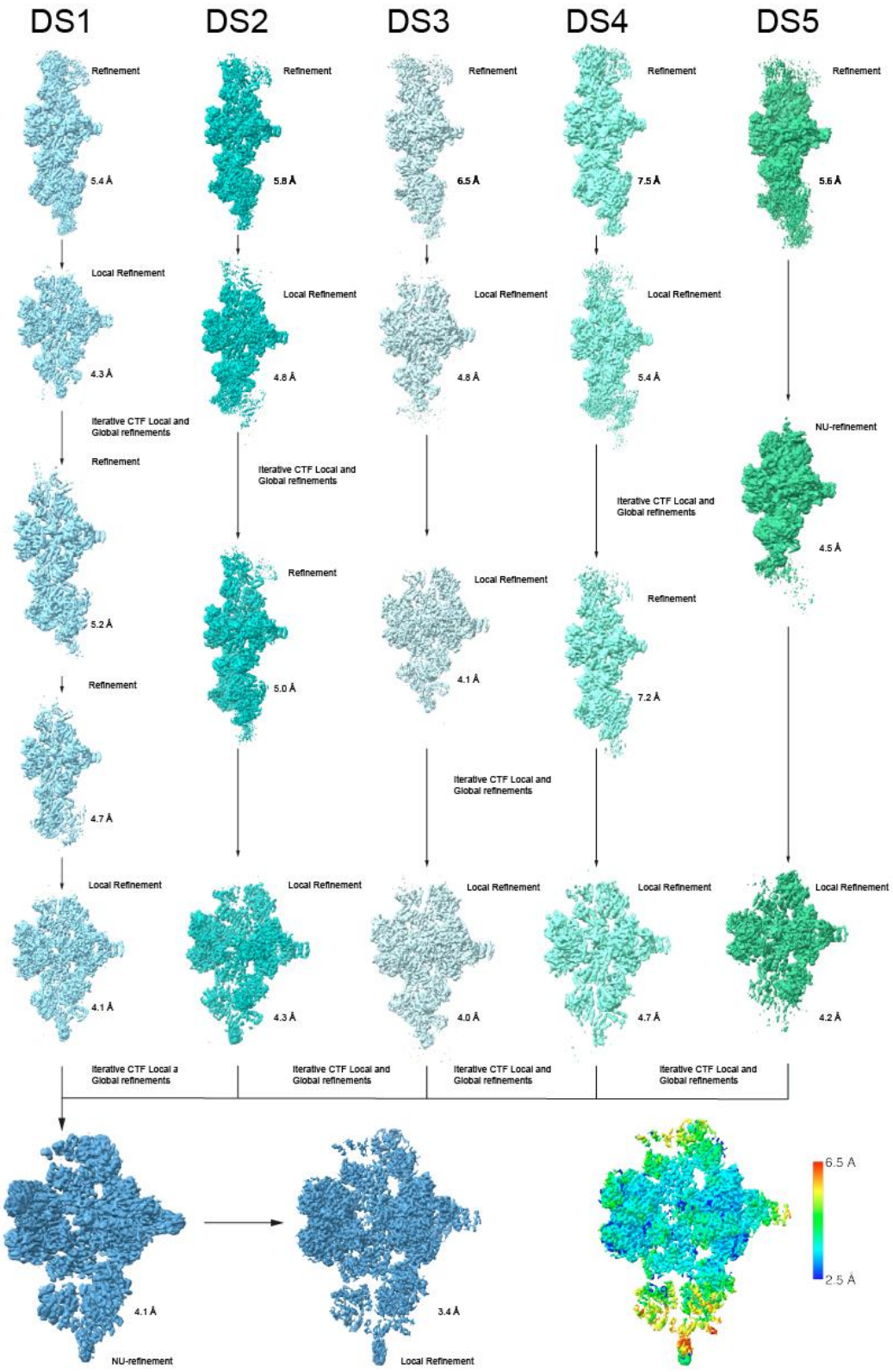

Figure 4.6: Processing scheme for ACC-Cit
Processing was done in cryoSPARCV2.1 and 3.3.

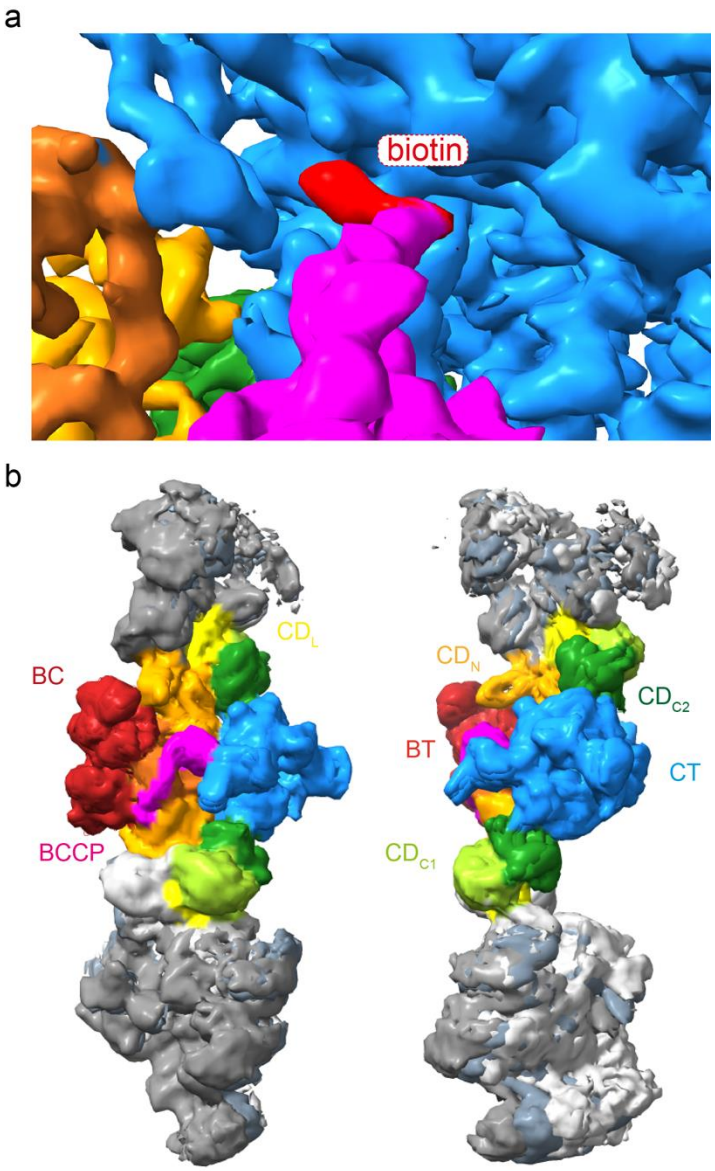

Figure 4.7: Visualisation of biotin and filament heterogeneity
Biotin moiety (red) visualized in the CT of the ACC-Cit filament. b) 3D-classification shows heterogeneity in classes and distinct angles between two ACC dimers with the filament.

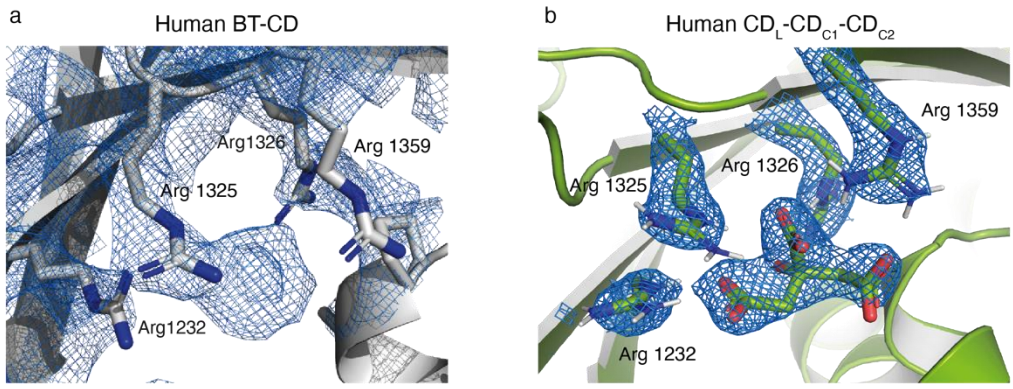

Figure 4.8: Electron density in the positively charged pocket of the CD
a) Human BT-CD crystal structure with extra density (PDB 5I87, [122]). b) electron density for citrate in human CD_L-CD_{C1}-CD_{C2} crystal structure. Arginines are shown as sticks.

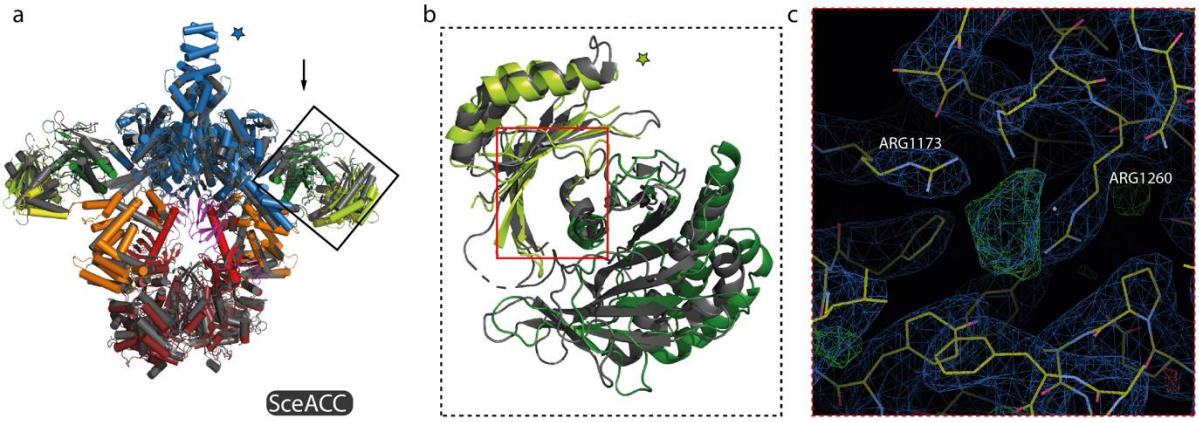

Figure 4.9: Extra density in conserved pocket in yACC crystal structure
a) Overlay of ACC-Cit dimer (color) and SceACC dimer (black) aligned at the CT domain (indicated with blue star) in cartoon representation with cylindrical helices. b) Cartoon representation of CD_{C1} and CD_{C2} domains aligned at CD_{C1} indicated by the green star. c) Image from coot from red highlighted region in b). Electron density map (blue) and model (yellow) of yACC crystal structure (PDB: 5CSK). Positive difference density shown in green. Contour level is 0.12 e^-/A^3.

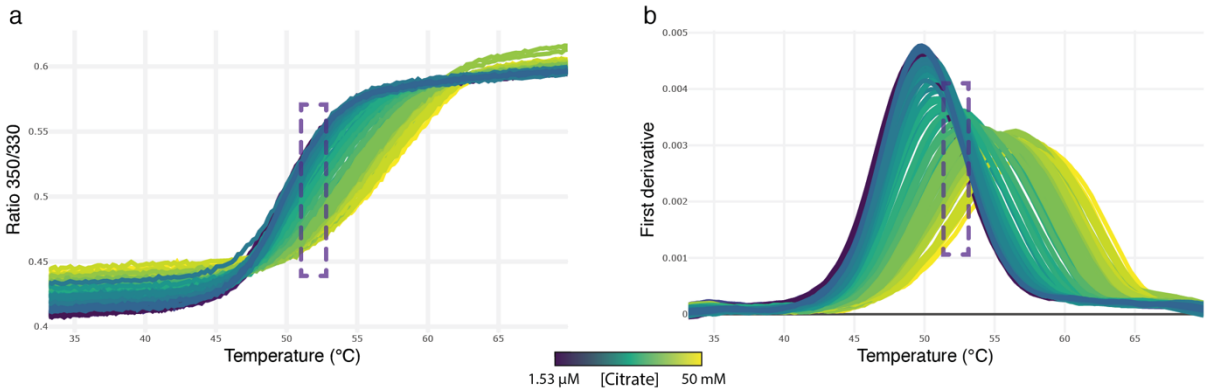

Figure 4.10: nanoDSF signal in the presence of varying citrate concentrations
a) Fluorescence emission ratio at 350/330 nm of the CD_L-CD_{C1}-CD_{C2} domain. b) First derivative of the fluorescence signal. Curves are coloured according to the citrate concentration. The region in the purple box was used for isothermal analysis. Isothermal analysis done with the FoldAffinity web server [272].

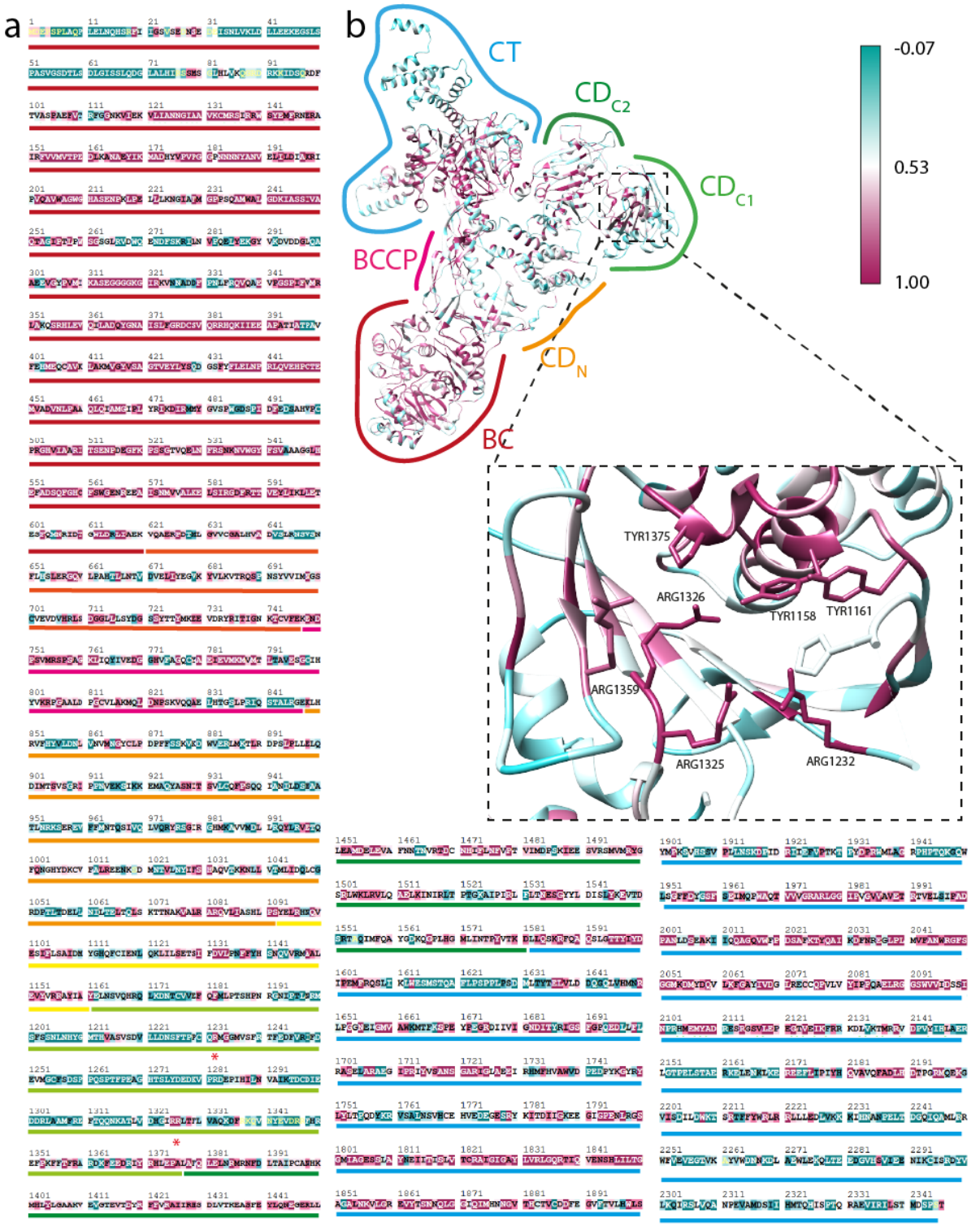

Figure 4.11: Sequence alignment shows highly conserved citrate binding pocket

a) sequence alignment with domains indicated in colour coded lines. Stars indicate mutation sites. b) conservation mapped onto ACC protomer with domains indicated. Bar shows level of conservation with 1 being highly conserved and inset of citrate binding pocket.

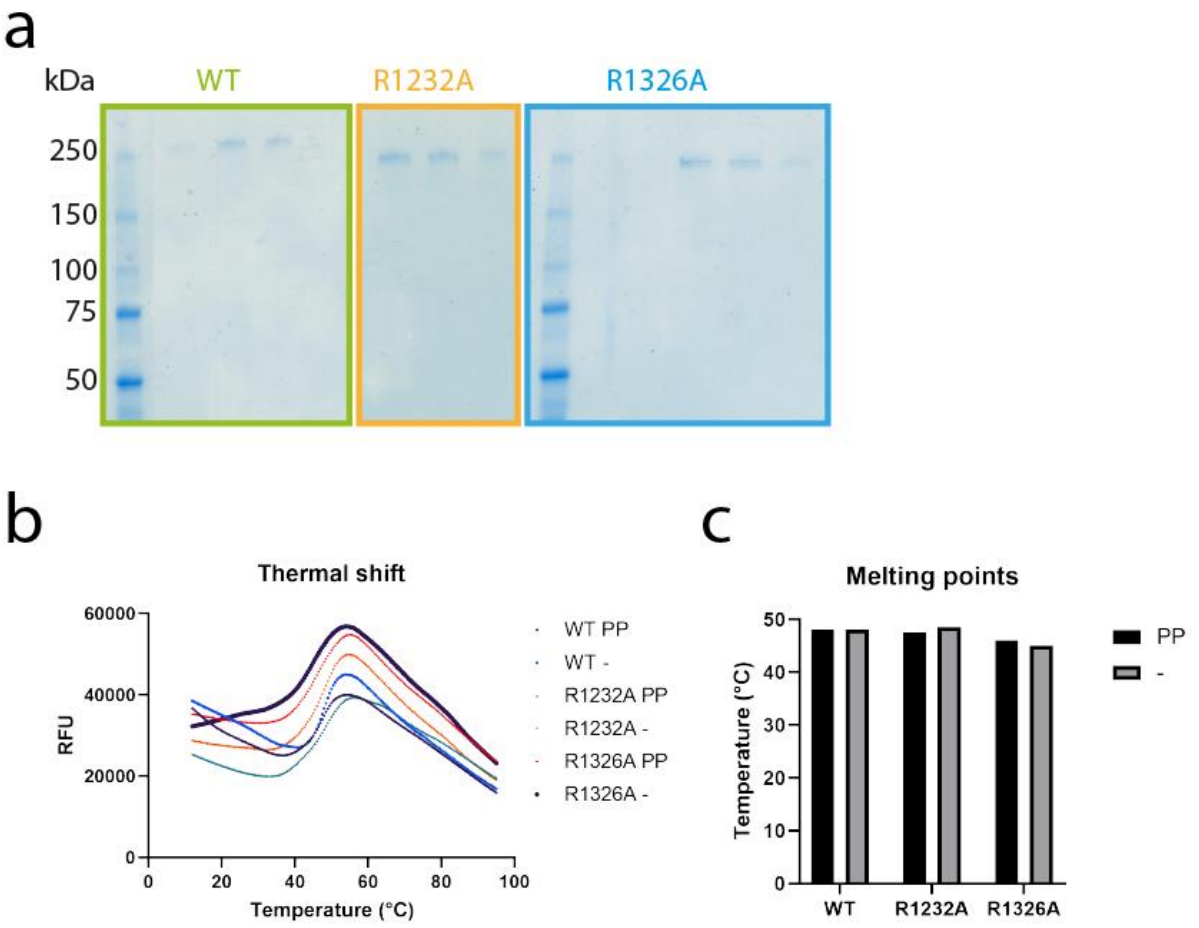

Figure 4.12: Expression and folding of ACC WT and point mutants
a) Coomassie stained SDS-PAGE of purified WT and mutants. c) Fluorescence readout of thermal shift assay plotted into temperature axis. RFU = raw fluorescence. d) Melting points determined by the first derivative of thermal shift assay plotted as bar graphs for visualization (PP=phosphorylated).

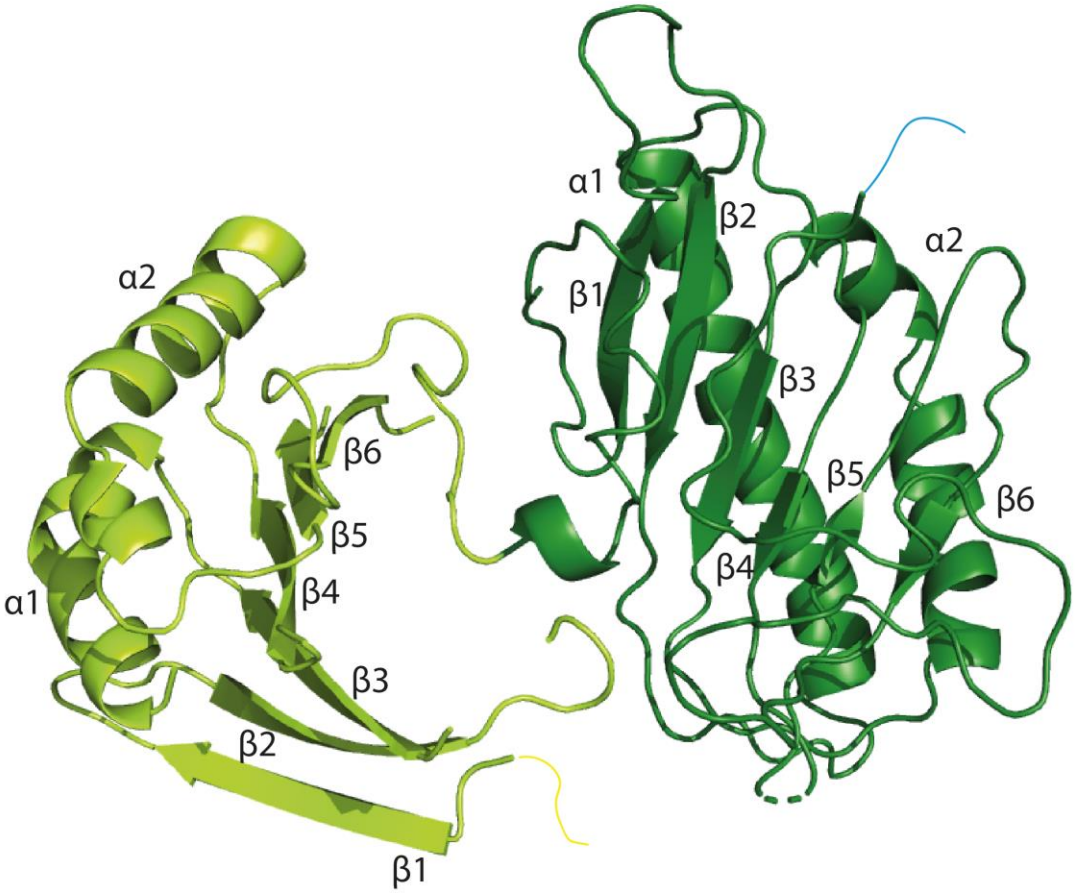

Figure 4.13: Human ACC-Cit CD_{C1} and CD_{C2} domains with labelled α-helices and β-sheets.

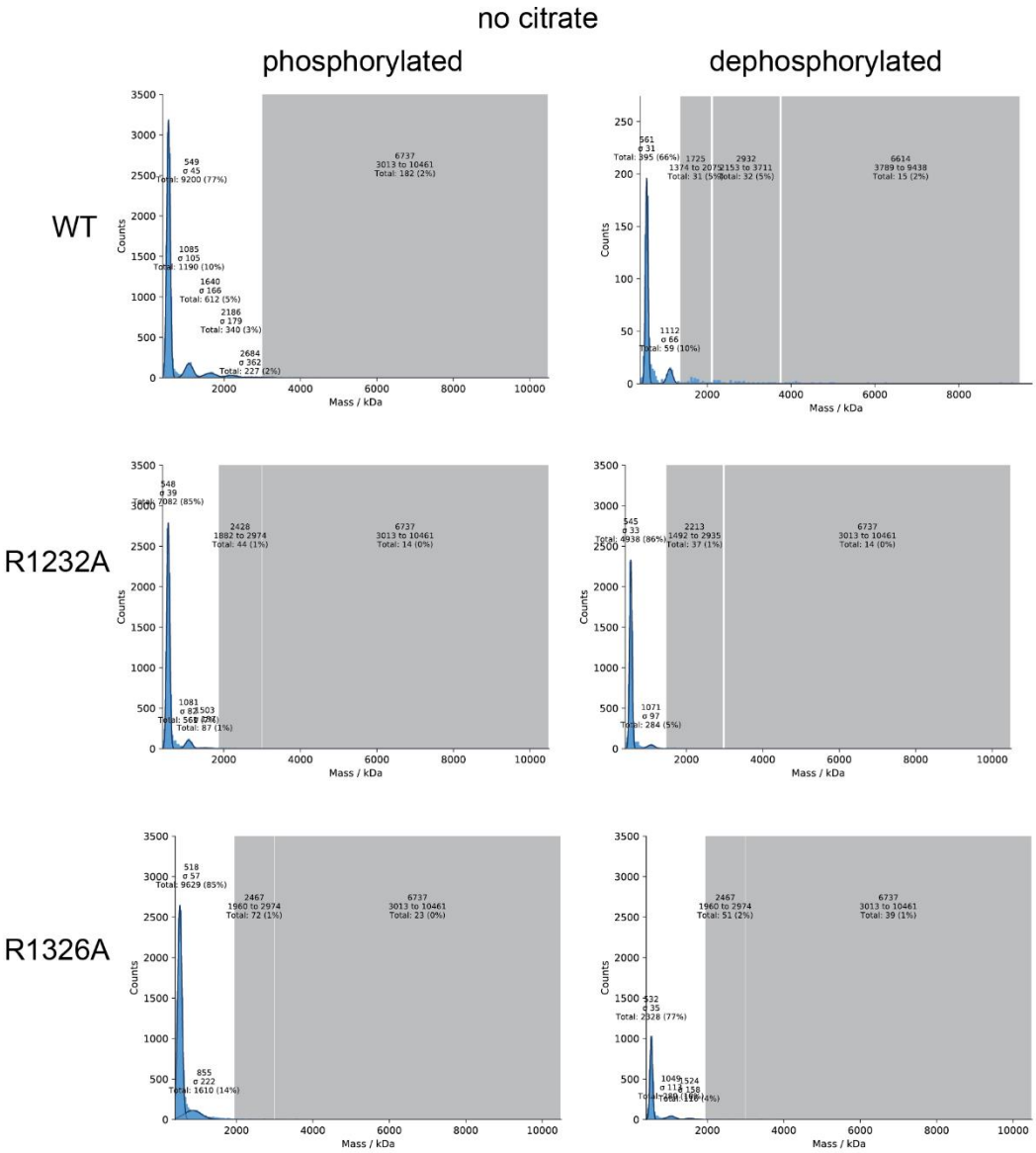

Figure 4.14: Mass photometry without citrate shows mostly low molecular weight species
Label: top: Mean peak value or integral of region (grey box); middle: Width of the fitted Peak (σ represents the standard deviation); bottom: Sum of the number of counts under the fitted peak/region (expressed as the number of counts and the proportion of counts– compared to the total number of events). X-axis: Molecular weight in kDa from start of the first peak to 10 MDa. Y-axis: Number of binding events show as counts.

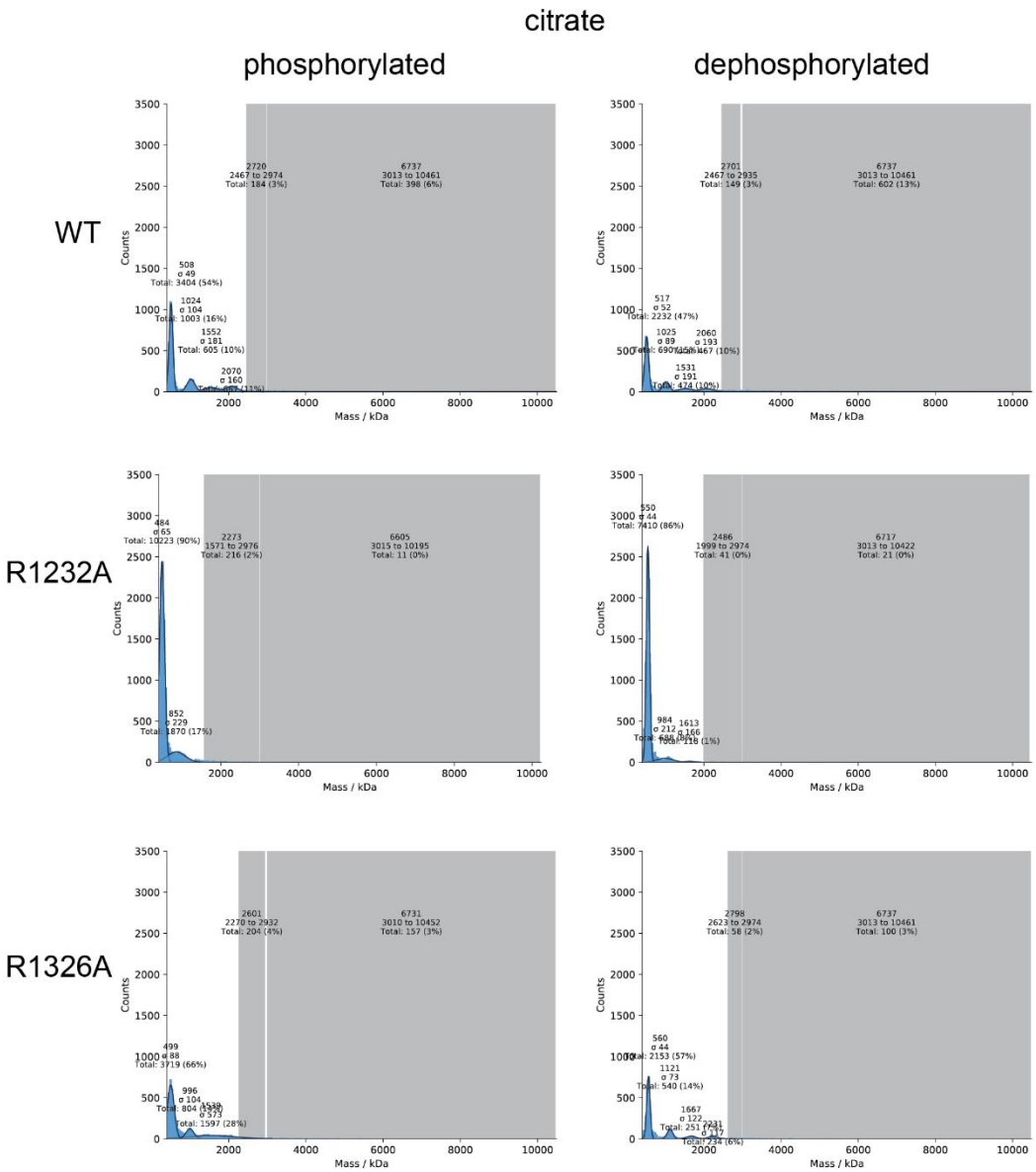

Figure 4.15: Mass photometry of WT with citrate shows high molecular weight species
Label: top: Mean peak value or integral of region (grey box); middle: Width of the fitted Peak (σ represents the standard deviation); bottom: Sum of the number of counts under the fitted peak/region (expressed as the number of counts and the proportion of counts– compared to the total number of events). X-axis: Molecular weight in kDa from start of the first peak to 10 MDa. Y-axis: Number of binding events show as counts.

4.9 Supplementary method

4.9.1 Assessment of phosphorylation by mass spectrometry

Levels of phosphorylation before and after *in vitro* dephosphorylation were assessed by mass spectrometry. Dephosphorylation was preformed independently three times for WT and both mutants using λ-phosphatase (*cfr.* 4.7.1 Cloning and protein purification).

Proteins were diluted to 50 ug total and were prepared for digestion with 1 ug of trypsin and digested overnight at 37°C. Solid phase extraction was done by using C18 columns. 2.5ug where removed from the sample as Input and was vacuum dried. Resuspension in 20 µl LC-MS buffer (0.1 % formic acid in H_2O) was followed by concentration measurement by nanodrop. 4 µl of sample were injected into Q Exactive HF orpitrap LC-MS/MS (QE-HF) system by Thermo Fisher. The remaining sample was subjected to phosphopeptide enrichment by the AssayMAP Bravo Platform (Agilent). Eluted sample was vacuum dried and peptides dissolved in 20 µl LC-MS buffer and 4ul injected into QE-HF. Analysis of results was done using Scaffold4 (version Scaffold 4.11.0, Proteome software)

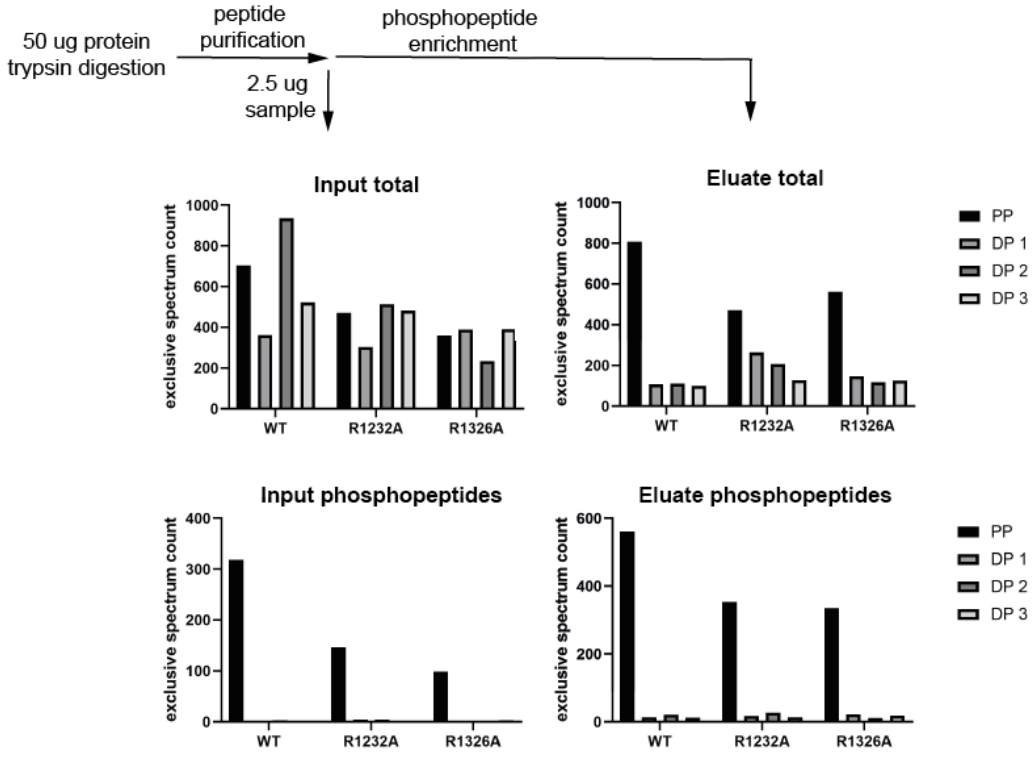

Figure 4.16: Mass spectrometry workflow
A strong reduction in phosphopeptides in dephosphorylated samples is detected.

5 Discussion

This thesis aimed to assess oligomerisation of multienzymes and its effect on catalysis. For *trans*-AT PKS, it has been hypothesised that assembly line proteins oligomerise into giant enzymatic complexes in the cell has been hypothesised. Filament formation and activation of human ACC by citrate has first been reported in the 1960s, yet the mechanism by which citrate binding is coupled to filament formation and increased ACC activity remained unknown.

5.1 Summary of findings

Aim 1 was to study PKS assembly and localisation *in vivo* in their native producers. The mupirocin *trans*-AT PKS from *P. fluorescens* was selected as a system and the approach consisted of labelling the assembly line proteins via chromosomal fusions to fluorescent protein coding genes. As successful labelling is a prerequisite for any further imaging, challenges in generating these fusions resulted in spending a substantial amount of time on modifying and optimising cloning protocols. Eventually, the N-terminus and C-terminus of the proteins MmpD and MmpB were successfully fused to fluorescent proteins. These proteins contain the first and last module of the mupirocin PKS assembly line. Fluorescence microscopy revealed that the labelled assembly line enzymes preferentially localise at a single region at the cell pole. Observing PKS expression and assembly with long-term imaging was not feasible due to extensive photobleaching combined with low fluorescence signal, likely a consequence of low expression levels. Additionally, expression of the biosynthetic gene cluster was heterogeneous within the population. Overexpression of the transcriptional activator led to boosted expression of the mupirocin PKS cluster, improving signal in imaging experiments. An increased number of cells showed fluorescence, without affecting localisation of labelled assembly line proteins. However, the issue of photobleaching remained, limiting imaging exposures. Despite the challenges of generating mutants by allelic exchange and long-term imaging, the labelled assembly line proteins MmpB and MmpD of the mupirocin *trans*-AT PKS were imaged in their native host *P. fluorescens* revealing their subcellular localisation at the cell pole.

Aim 2 consisted of the studying the oligomerisation of the mupirocin *trans*-AT PKS, by reconstituting the full assembly line *in vitro*. Because of their size, PKS proteins are typically studied as isolated domains or truncations, but we aimed for recombinant expression of full-length assembly line proteins of the mupirocin *trans*-AT PKS. In a first step, expression constructs were established despite considerable difficulties in cloning the genes of interest. Due to the high GC

content, large size and repetitive nature of the sequences, existing cloning protocols were extensively optimised to screen for conditions enabling generation of expression constructs. I describe approaches to optimise cloning of and generate expression vectors of these large assembly line proteins. Full-length the assembly line proteins were purified from insect cells. Their low stability and tendency to aggregate required optimisation of several chromatographic purification strategies. Ultimately, a single affinity purification step provided the best compromise between purity and yield. The aggregation tendency at concentrations above 0.5 mg/ml could potentially correspond to filament formation. However, screening for conditions producing higher order assemblies by light scattering revealed 98 to 99 % of homodimeric assemblies in the conditions tested. Cryo-EM analysis of a construct consisting of MmpD modules 3-4 did not reveal of higher-order assemblies. While a map of the KS domains was obtained at a resolution of 3.4 Å, no additional domains were visible in this map. Likely a consequence of extensive flexibility of the protein. Here, protocols to recombinantly express all assembly line proteins from the mupirocin trans-AT PKS were set-up. Ultimately, by reconstituting these proteins and manipulating the assembly line *in vitro*, important determinants for polyketide synthesis can be revealed.

Aim 3 was to understand the mechanism by which citrate binding to human ACC results in filamentation and full activation of the multienzyme. To identify the binding site from cryo-EM reconstructions, additional cryo-EM data were collected, and was merged with previously collected datasets. Using automated particle picking, the amount of picked particles was doubled. With new processing algorithms, that refine the CTF for each individual particle and improve microscope misalignments by aberration corrections, the resolution lead of ACC-citrate filaments was improved from 5.4 Å to 3.4 Å. Density in a positively charged pocket of the central domain was observed, however the local resolution did not allow unambiguous placement of citrate. Based on these structural findings, new truncated constructs of the central domain were designed, cloned, expressed and subsequently purified in the presence of citrate. A crystal structure at a resolution of 2.0 Å of the truncated central domain was obtained in the presence of citrate, enabling unambiguous placement of citrate. The binding affinity of citrate to the truncated central domain construct was estimated through DSF experiments. Findings on the truncated construct were combined with analysis of ACC variants with mutations in the citrate binding site to assess their effect on ACC activity and filament formation. Mutants exhibited reduced ability to form malonyl-CoA as measured in an ACC activity assay and reduced filament formation, as measured by mass photometry. Overall, these findings elucidate the mechanism of ACC activation by its allosteric activator citrate and the subsequent conformational changes exposing interfaces enabling active filament formation.

5.2 Multimodular organisation of PKS

ModPKS are large multienzyme complexes, consisting of multiple modules organised in an assembly line [40, 276]. The polyketide product is determined by the order and domain composition of the modules. Catalytic domains are connected by linkers and the multiple active sites are reached by the substrate tethered to the ACP. While the individual canonical PKS domains are already well-studied in isolation, understanding the mechanism of substrate transfer between domains combined with multimodular architectures of modPKS assembly lines remains a challenge. The presence docking domains combined with observation of a large megacomplex for the bacillaene *trans*-AT PKS in *B. subtilis,* suggests formation of giant enzymatic assembly lines in the cells [43, 104]. In *trans*-AT PKS, the three-helix LINKS-motif flanking KSs is capable of zipping the KS into a megacomplex [104]. Simultaneous research in our lab was performed on a truncated 438 kDa *trans*-AT PKS bimodule core from of *Brevibacillus brevis,* with the domain composition KS-DH-ACP-KS. A cryo-EM reconstruction revealed filament formation mediated by the LINKS motif of both KSs (Figure 5.1). To decipher the mechanism of substrate transfer and conformational coupling, a fungal iPKS was characterised during different states of catalysis using mechanism-based crosslinking of the ACP to catalytic domains [250].

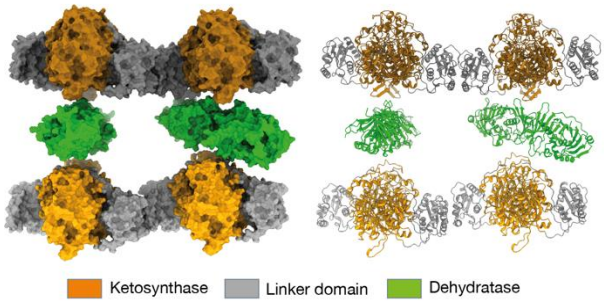

Figure 5.1: Filament formation of a *trans*-AT PKS bimodule mediated by the LINKS motif
Two bimodules of a trans-AT PKS are shown in surface (left) or cartoon (right) representation forming filaments through the LINKS-motif of their KSs (modified from Tittes *et al.*, in revision).

Concurrent research on PKS was aimed at elucidating the structure of modules, based on the divergence between the arch-shaped PikAIII structure [87] compared to mFAS [15], MAS [39] or DEBS [96, 97] (Figure 1.6). To assess if this arch-shaped conformation occurs during polyketide synthesis, a high affinity antigen-binding fragment (Fab) was generated for DEBS module 2 and 3. The Fab binds to and locks the module in its extended conformation, as observed by a crystal structure. An *in vitro* polyketide synthesis assay in presence and absence of the Fab did not reveal any significant difference in product turnover, indicating that the extended conformation is sufficient for product synthesis. This assay was repeated in a SAXS experiment, confirming that the module is mostly present in its extended conformation [236]. Additionally, by

combining SAXS, crosslinking Mass spectrometry (XL-MS) and integrative modelling, similar findings were made for DEBS module 2 confirming that in solution it is mostly present in its extended conformation [277]. The cryo-EM structure of the LovB module from the fungal fully reducing lovastatin PKS also confirmed that the structural organisation of PKS modules resembles the one of mFAS [244].

In 2021, two new structures of cis-AT modPKS were published [234, 235]. The structure of Lsd14 from the lasalocid A biosynthetic gene cluster of *Streptomyces lasalocidi* [234] and a hybrid structure of module 1 and 3 from DEBS [235]. Lsd14 consists of the KS-AT-KR-ACP domains, representing a small modifying region of a modPKS. The hybrid structure of DEBS contains the KS and AT fragments from module 3, the KR and ACP of module 1, fused to the TE of module 6, thus the final domain composition consists of KS3-AT3-KR1-ACP1-TE. Both modules have a similar domain organisation in the modifying part and to aid in structure determination a Fab-fragment was used [236]. Unexpectedly, the authors discovered an asymmetric architecture of the module. One of the ACPs is docked to the AT and the KR active site entrance is rotated towards the active site of the KS. The active site entrance of the other KR of the homodimer is rotated away from the KS and AT reaction chambers. This asymmetry contrasts previous models that revealed two-fold symmetry [87, 236, 244]. Binding of the ACP to one of the catalytic domains could induce conformational changes resulting in rotation of one of the KR domains. However, this asymmetric state is also observed in samples where no bound ACP is visible.

Overall, both of these structures revealed that modPKS are present in different conformational states, and suggest that flexibility is required for product synthesis. These findings also highlight the importance of cryo-EM to study large and flexible multienzyme complexes, combined with engineering constructs and stabilisation by Fab fragments [251]. Despite these structural insights into modPKS, the mechanism of directed substrate transfer between modules of assembly line PKS remains elusive.

Typically, modPKS are studied either as isolated domains or as truncated modules [97, 108, 234-236]. However, most of these proteins contain at least multiple modules on the same polypeptide. At present, there is limited insights into the connection between these modules required for substrate transfer. Information about the multimodular architecture of modPKS is also lacking. Findings from *in vivo* imaging combined with high resolution cryo-EM models, suggest large complex formation of *trans*-AT PKS [103] (Tittes *et al.*, in revision). At the N- and C-termini, docking-domains connect the assembly line proteins in the order and direction of product

synthesis [43], with lateral interactions mediated by the LINKS-motif stabilising the complex through avidity (Figure 5.2) [104]. These large complexes are held together by a multitude of weak interactions. This limits diffusion distances required for the ACPs to reach the active sites and possibly prevents crosstalk between subunits of different assembly lines [14]. At present, it remains unclear how the *trans*-acting enzymes are recruited. Such a complex provides a high local concentration of enzymatic domains and substrates, which could recruit and sequester *trans*-acting enzymes such as the essential AT domain from *trans*-AT PKS. Overall, formation of such a large complex is hypothesised to increase efficiency of product synthesis [104].

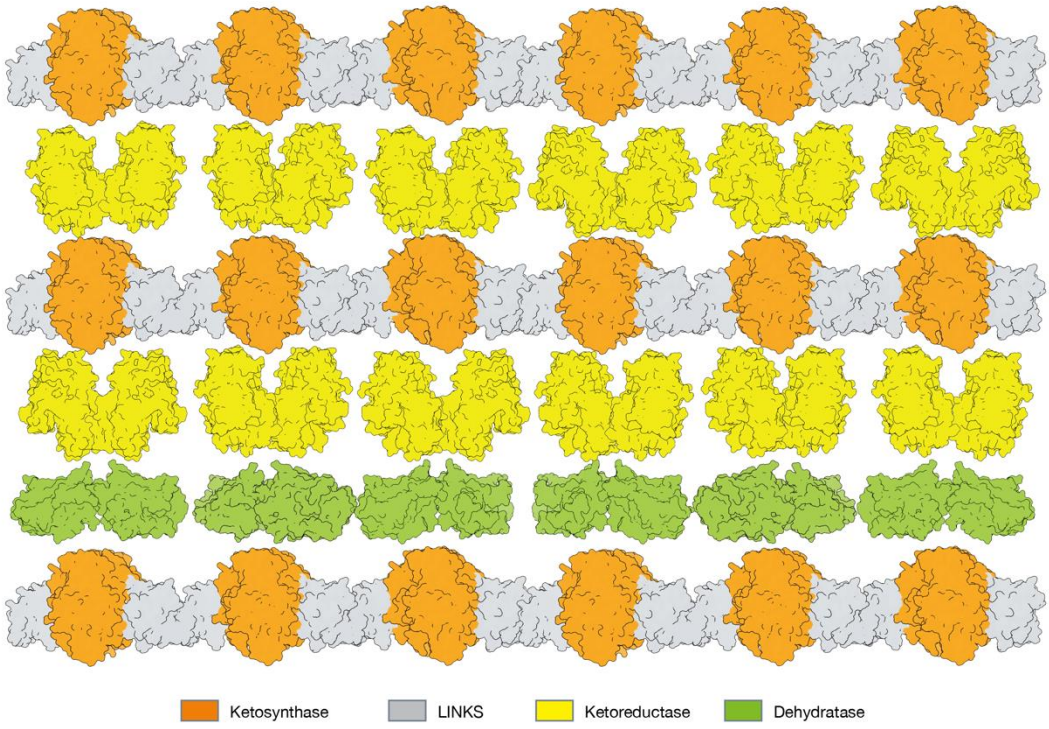

Figure 5.2: Suggested organisation of PKS megacomplex
Several modules of *trans*-AT PKS laterally stack through their LINKS-motif. Module architecture and domain organisation is based on available structures.

So far, to the best of my knowledge no additional insights into higher order organisation of PKS were obtained. This is likely a consequence of the challenges of gene size, flexibility and low stability of the proteins. However, new developments in cloning large DNA fragments from biosynthetic gene clusters [278], CRISPR/Cas engineering of modPKS genes in the chromosome [279] and new fluorescent proteins with improved properties [280] can expand upon our understanding of these enzymes in their native producer strains. Additionally, progresses in resolution and sample preparation for cryo-electron tomography (cryo-ET, see 5.4) may enable us to obtain detailed insights into the subcellular organisation of these enzymes, circumventing the requirement and challenges of studying PKS in isolation.

Engineering t*rans*-AT PKS has the potential to generate novel polyketides with therapeutic use. Several engineering attempts demonstrated that domain and module swaps are possible, but suffered from low yields or mixed products, likely as a consequence of lack of substrate specificity or suboptimal domain interactions [34, 67]. Ultimately, insights in the substrate transfer mechanism between PKS modules, obtained with high resolution structures in different catalytic conformations, will facilitate engineering attempts.

5.3 Inhibition of ACC filament formation as a therapeutic approach

Due to the role of human ACC in cancer and metabolic syndrome, two disease areas with a high burden on society, inhibition of ACC is a promising therapeutic strategy [156, 166]. A novel therapeutic approach could be targeting the citrate binding site to inhibit formation of active filaments. Small molecule ACC inhibitors are often unselective towards both ACC isoforms. Yet only ACC1 is known to form filaments in cells [127]. Targeting filament formation could be a specific way of inhibiting only the ACC1 isoform, and thereby limiting cellular lipid synthesis. Full inhibition of both ACC isoforms is not desirable due to accumulation of acetyl-CoA and subsequent acetylation of proteins [149, 172, 174]. By only inhibiting one of the two isoforms, acetyl-CoA accumulation can potentially be limited.

In solution, ACC1 is present in its homodimeric form and maintains basal activity in its dephosphorylated form [123]. After citrate binding it becomes fully activated and is present in its filament form [123]. In yeast ACC, a phosphorylated serine binds to an analogous pocket to the citrate binding site in human ACC, and inhibits its activity [122]. In yeast, binding of this loop prevents the conformational changes required for dimerization of the BC domains, keeping the enzyme inactive. Due to the conserved nature of this pocket, the same mechanism could be applied to human ACC, either through a phosphopeptide mimetic or a small molecule inhibitor targeting this site. However, more research is required to confirm this as a feasible approach of inhibiting ACC1 activity. Alternatively, the binding interfaces between the filaments can be targeted. Here, by destabilising the filament state of ACC1 its activity is reduced. Based on high-resolution structural insights into the interface between protomers, compounds which potentially disrupt binding can be screened.

Ultimately, the activity and effect of citrate activation on ACC must be tested *in vivo*. Introducing specific mutations preventing citrate binding and subsequent ACC activation will assess the

importance of this activation mechanism on fatty acid synthesis in cells. A typical western diet contains an excess of fatty acids, and as a consequence *de novo* fatty acid synthesis in most tissues is low. However, since ACC is upregulated in liver, breast, prostate and colon cancers, and its expression levels correlate with poor outcome [154, 157-161]. Therefore, assessing the effect of citrate activation in cancer cell lines and in animal models would reveal the importance of targeting this site as a therapeutic strategy.

5.4 Implications for structural biology of multienzymes

To obtain structural insights at atomic resolution, typically proteins of interest are overexpressed, purified and studied by NMR, crystallography or cryo-EM. These methods have proven highly successful in obtaining detailed insights into the function of multiple proteins, information about the cellular context of the protein of interest is missing. Moreover, analysis of selected parts of a protein complex in isolation can result in missing potential interaction partners. An attractive approach, which is rapidly gaining interest is cryo-electron tomography (cryo-ET) *in situ*.

In cryo-ET, the sample of interest is imaged with low dose and tilt-series are acquired [281]. These tilt series are combined into a 3D reconstruction. Using this method the structures of interest can be imaged *in situ* [201]. This method has the potential to visualise the assembly line organisation of modPKS at near-atomic resolution. Prior to imaging, samples need to be thinned to less than 0.5-1 µm, because thicker samples cause multiple scattering events of electrons used in microscopy [282]. Two approaches exist to thin samples while preserving the integrity of the structures of interest, cryo-sectioning [283] and milling with a focused ion beam [284]. To observe the PKS assembly line in bacterial cells by cryo-ET, cell thinning may be required depending on thickness [285].

Cells are densely packed with macromolecules [5]. The next challenge consists of identifying the molecules of interest within the tomograms. These can be identified after fluorescent labelling using cyro-correlative light electron microscopy [286] or matching them with a known template [202, 282]. In case of the mupirocin PKS of *P. fluorescens,* the assembly line proteins are predominantly located at the cell pole. This narrows down the regions to image by tomography. Recognising the molecules of interest may still be challenging due to lack of a template and the dense environment at the bacterial cell pole [192]. Cryo-ET has the potential to reveal the in cell architecture of modPKS at near atomic resolution, avoiding the challenges of associated with *in vitro* study of these enzymes.

One of the more recent developments is the AlphaFold2 program [287]. This solved one of the biggest problems in biology, and can predict a protein fold from its primary structure [3], based on evolutionary coupled residues and information of conserved peptide structure [287, 288]. The DeepMind team responsible for AlphaFold in collaboration with EMBL-EBI, quickly set-up a database containing protein structure predictions for every protein of model organisms [289]. Where traditionally, insights were obtained with high-resolution models by the end of a project, now a prediction is already available at the initial stages of a project. This provides researchers with initial models of their target proteins, that can subsequently be used for construct design, assay development, predicting ligand binding sites, fitting models into experimental EM maps, replacement models for phasing of X-ray crystallography datasets etc…. Some of the current limitations are the inability to accurately predict protein multimers. Recently, a preprint on multimer prediction by AlphaFold was released, yet its ability to accurately predict multimers varies between targets [290]. Prediction of single, homodimeric, enzymatic domains of PKS did not produce plausible models, with multiple clashes and polypeptides physically intersecting. Predictions of full-length assembly line proteins of the mupirocin cluster were too computationally intensive and did not yield any result. Likely due to lack of available co-evolutionary data between the sequences and structures to make an accurate structural prediction. On the contrary, for human ACC the predicted model nearly perfectly matched with the ACC-Citrate filament conformation (Figure 5.3).

While this structure prediction is almost indistinguishable from an experimentally determined model, AlphaFold was unable to predict the different conformational states of ACC. Many proteins are dynamic and can be present in multiple conformations. Other limitations are that flexible loops are modelled with decreased confidence compared to the folded core, inability to predict PTMs, interactions with cofactors and DNA or RNA protein interactions [288]. While limitations still remain, it is undeniable that AlphaFold certainly has a major impact and changes the way we approach structural biology.

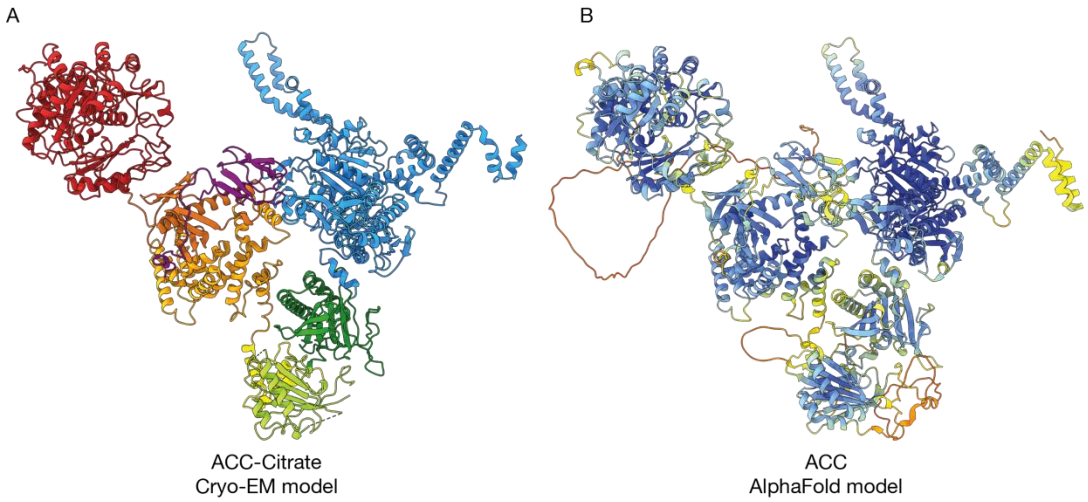

Figure 5.3: Comparison between cryo-EM and AlphaFold models of ACC
A: Cryo-EM model of ACC-Cit, domains are colour coded. B: AlphaFold prediction of ACC, colour code corresponds to model confidence ranging from blue: very high to orange: very low.

In recent years there has been tremendous progress in EM sample preparation, imaging and computational methods for data processing and structural analysis. One can imagine the field developing towards integrating information from multiple methods to gain a complete understanding of protein function. Predictions from AlphaFold models are used to aid construct design or to make hypotheses. To obtain high resolution models, single-particle cryo-EM analysis remains the method of choice to study large protein complexes in different conformational states. These cryo-EM or AlphaFold models could subsequently be used as templates in cryo-ET datasets to obtain insights into the organisation of these macromolecules *in situ*. With these advances in technology, we will be able to obtain a comprehensive understanding of fundamental biological processes performed by macromolecules in cells.

5.5 Outlook

Both multienzymes discussed in this thesis offer potential to be combined into a large “cellular factory” producing the desired compound. Linking ACC to a modPKS assembly line, could result in high local malonyl-CoA pools to be immediately used for polyketide synthesis instead of being directed towards fatty acid synthesis. In the past, several PKS engineering attempts have suffered from low product yields [67, 291-293]. Expanding or modifying the acyl-CoA precursor pool has already been able to increase production of the desired compound [66, 294, 295]. This is achieved through modification of expression levels of enzymes involved in precursor synthesis or consumption to alter the available precursor pool. By heterologous expression and manipulating

the activation of ACC, the malonyl-CoA precursor pool could possibly be controlled and expanded, and thereby polyketide synthesis as well.

Type I modPKS are attractive engineering targets as their module composition determines the product. Their versatility lies in the ability to combine modules to form the desired product. The advantage of such a system would be that all its components are encoded into DNA, requiring only ATP, acetyl-CoA and NADPH which are readily present in the cell. By using enzymatic synthesis of specific compounds, one would circumvent the requirement of chemical synthesis, which is especially challenging for polyketides [60-63]. One can imagine where ACC is combined with a PKS producing the compound of interest are present on a large plasmid or bacterial artificial chromosome (BAC) [278, 296, 297]. These can in principle be expressed in the cells of interest to generate large amounts of the desired polyketide. The strength of this system lies in the ability to make manipulations at DNA level to modify the desired product, without requiring optimisation of chemical synthesis.

Research in this area has not only attracted interest in producing therapeutically relevant compounds, but also for the synthesis of biofuels. [68, 69]. The majority of transportation vehicles remain petrol-fuel based and as a consequence of the production of atmospheric CO_2, the climate is changing [298]. Awareness is increasing and high-profile organisations are already moving towards the use of biofuels with the aim of being net-carbon neutral [299]. Microbial synthesis of biofuels has generated considerable interest as it forms a closed "CO_2 cycle", where CO_2 from the combustion of biofuels is incorporated into the synthesis of new fuels [68]. Specific substitutions introduced by PKS on the aliphatic chain of a synthetic fuel can result in high performance fuels. For a promising alternative, high fuel titers are required. Metabolic engineering by increasing polyketide precursors offer potential to improve production yields [300].

Alternatively, modifying bacteria to produce anticancer compounds of interest can be considered for targeted cancer therapy. The microtubule-targeting epothilones [246, 301] or the DNA crosslinking colibactin [62, 302] are polyketides which are already in clinical use or that have demonstrated cytotoxicity [303, 304]. Certain bacteria, such as *Salmonella Typhimurium* or *Listeria* spp. strains, colonise tumour cells [305]. Attenuated strains have been developed to selectively colonise tumours while being rapidly cleared from the host in other tissues [306]. This ability is already exploited to deliver specific protein effectors into the tumour cells with the aim of activating the immune system and clearing the tumour [307]. One could consider loading these strains with an engineered PKS assembly line producing high titers of polyketides with anti-cancer

activity. By colonising the tumour cells and producing the polyketide of interest in the tumour micro-environment, high local concentrations of the therapeutic compound could be achieved. This targeted therapy can result in decreased side-effects because systemic spread of the compound towards other tissues is minimised. One can expand on already established bacterial systems used in targeted tumour therapy, such as the delivery of protein effectors through the type III secretion system.

These tumour-colonising bacteria could be modified to contain a multitude of anticancer systems in a single organism, with the aim of fully eradicating the tumour cells. A first step would consist of engineering these bacteria for heterologous expression of the PKS producing the desired compound. In principle, this could be achieved through expression of the PKS cluster from a BAC [278]. Next, once in the tumour micro-environment the bacteria should have enough nutrients to express the PKS cluster and sufficient acyl-containing building blocks required for polyketide synthesis. Finally, the feasibility of this approach needs to be confirmed in animal models.

Advancements in our understanding of modPKS and ACC offers possibilities to re-engineer them in order to alter their products and improve yields beyond their natural metabolic requirements. By combining structural and biochemical knowledge with microbial engineering, considerable advances can be made in the efficient synthesis of novel compounds of important relevance for mankind.

6 Acknowledgements

Thank you to both Timm and Marek for enabling me to work on this shared project. Being an active member in both of your research groups was extremely valuable. It allowed me to be exposed to different environments and learning from both of your research approaches. Timm, I appreciated all your valuable advice through spontaneous conversations, your open door where I could just walk in to discuss problems and share successes, directly communicate to you about anything but also all the fun we had. Marek, through you I have learned so much on scientific thinking, how to approach problems and troubleshoot them. When I felt stuck you always had ideas, especially during our fights with *Pseudomonas* cloning. Thank you for having me as a part of your lab and including me in the lab events. Both of you spent a lot of time, shared useful and positive advice about science, research, experiences, career and life. In hindsight you were right, but I had to experience it myself before realizing it. Thank you to Richard and Jan-Willem for being part of my thesis committee and the positive discussions in our meetings.

To all the current and past group members, thank you for being part of the group. It was a pleasure to celebrate the ups and have your support during the downs of my PhD.
To Hugo, I am grateful that you are part of team PKS. You shared so much with me about personal situations, scientific advice but also all the nonsense we did in small office and out, better out than in. We become beasts! Karo, you made me realise that when you really want something you go for it. Thank you so much for all the support about project, career and future. We had a lot of fun together and we will continue to have a lot of fun, also with the rabbits. Mirko in the short time you were present you quickly became part of team PKS and team Formula 1. After seeing all my problems with PKS cloning you chose to join me. Welcome to team cloning problems! You pushed the project in a new direction and with your positive attitude it was a pleasure to work with you. Elsa, we are team multienzymes and the French speaking connection. We keep pushing when nothing works and when you need a break, there's ovomaltine in my drawer. To Roman, you introduced me to the lab and were sharing your knowledge when needed. I always appreciated your input. Sophia it was a pleasure to have you in the group, especially during the pandemic where everything was closed. It was a strange time, but we managed because we're beasts. Miro, you're always there with creative ideas for experiments. Such a pleasure to be your microbiology lab neighbour and have fun in the lab outside of working hours. Melanie, working with you is incredibly efficient. I wish every collaboration is like this! I will always remember our first synchrotron shift, where we collected data through a hotspot from my phone. Raffaella, happy to

have you to share solutions with in team cloning problems. In the end, we did it! Maxim, I have always enjoyed you sharing historical fun facts during late evenings. Leonie, thank you for trusting me with the ACC project and laying the groundwork. Louise-Marie, good to have you in the team to teach us proper Scottish words. Simone, a pleasure to have you as a friend and for training. If you ever need a pizza recipe, I'm happy to share mine. Francesca, it was good to have someone with your power around. I appreciated all the useful advice and conversations with you. Sharath, you have always showed a lot of interest in my project and the science. The questions you asked were always relevant bringing a different perspective. Thank you, Karol Kaiser, the Bettmeralp veteran. Polina, thank you all the fun discussions during early morning, and I'm looking forward to more fun trips with you and Raphael. Florian Naatz, such a pleasure to watch F1 races with you, especially when Red Bull wins from Mercedes. To Matze, thank you for sharing your experience especially towards the end of your PhD. To Yves, it was fun to teach block courses with you. Every year we were evaluated as the best supervisors! To Shubham, thank you for the good food and it was nice to have you in small office. To the big office people, Malik, Niko and Yana all the best. To the Basler lab people, it was always nice to come upstairs for a chat or a break. Thank you for involving me in all your fun activities. To Matthias Zeug, good to have you in BayC, it would be lonely and silent without you. I appreciate all the scientific and non-scientific discussions we have. Thank you to our administrative assistants, Alexandra and Steffie, for taking care of all the administration behind the scenes so we could focus on our research. I have also enjoyed our many informal conversations. Patrick, you were our first apprentice and I'm happy that we're still in touch after all these years. Christina, you joined in a very strange time for everyone, but with your presence in the lab made it a much nicer period.

Thank you to my family and friends back in Belgium for supporting my move abroad for my PhD. Even if we didn't see each other much, I always felt welcome when coming home.

We had a lot of fun together! I wish everyone all the best.

7 References

1. Glasfeld, A., *Biochemistry: The Chemical Reactions of Living Cells, 2nd Edition (David E. Metzler).* Journal of Chemical Education, 2004. **81**(5): p. 646.

2. Skipper, L., *PROTEINS | Overview*, in *Encyclopedia of Analytical Science (Second Edition)*, P. Worsfold, A. Townshend, and C. Poole, Editors. 2005, Elsevier: Oxford. p. 344-352.

3. Anfinsen, C.B., *Principles that Govern the Folding of Protein Chains.* Science, 1973. **181**(4096): p. 223-230.

4. Venter, J.C., et al., *The Sequence of the Human Genome.* Science, 2001. **291**(5507): p. 1304-1351.

5. Moran, U., R. Phillips, and R. Milo, *SnapShot: Key Numbers in Biology.* Cell, 2010. **141**(7): p. 1262-1262.e1.

6. Pelley, J.W., *16 - RNA Transcription and Control of Gene Expression*, in *Elsevier's Integrated Review Biochemistry (Second Edition)*, J.W. Pelley, Editor. 2012, W.B. Saunders: Philadelphia. p. 137-147.

7. Ashcraft, B.A., et al., *Polymer-protomer transition of acetyl-CoA carboxylase occurs in vivo and varies with nutritional conditions.* Journal of Biological Chemistry, 1980. **255**(21): p. 10033-10035.

8. Ekman, D., et al., *Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions.* J Mol Biol, 2005. **348**(1): p. 231-43.

9. Alberts, B., et al., *Molecular Biology of the Cell 6th ed (New York, NY: Garland Science).* 2014.

10. Hawkins, A.R. and H.K. Lamb, *The molecular biology of multidomain proteins. Selected examples.* Eur J Biochem, 1995. **232**(1): p. 7-18.

11. Leibundgut, M., et al., *The multienzyme architecture of eukaryotic fatty acid synthases.* Curr Opin Struct Biol, 2008. **18**(6): p. 714-25.

12. Cortes, J., et al., *An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea.* Nature, 1990. **348**(6297): p. 176-8.

13. Bianchi, A., et al., *Identification of an isozymic form of acetyl-CoA carboxylase.* J Biol Chem, 1990. **265**(3): p. 1502-9.

14. Kosol, S., et al., *Protein-protein interactions in trans-AT polyketide synthases.* Nat Prod Rep, 2018. **35**(10): p. 1097-1109.

15. Maier, T., M. Leibundgut, and N. Ban, *The crystal structure of a mammalian fatty acid synthase.* Science, 2008. **321**(5894): p. 1315-22.

16. Perham, R.N., *Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions.* Annu Rev Biochem, 2000. **69**: p. 961-1004.

17. Evans, D.R. and H.I. Guy, *Mammalian pyrimidine biosynthesis: fresh insights into an ancient pathway.* J Biol Chem, 2004. **279**(32): p. 33035-8.

18. Wakimoto, T., et al., *Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont.* Nat Chem Biol, 2014. **10**(8): p. 648-55.

19. de Kok, A., et al., *The pyruvate dehydrogenase multi-enzyme complex from Gram-negative bacteria.* Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1998. **1385**(2): p. 353-366.

20. Tong, L., *Chapter Five - Striking Diversity in Holoenzyme Architecture and Extensive Conformational Variability in Biotin-Dependent Carboxylases*, in *Advances in Protein Chemistry and Structural Biology*, T. Karabencheva-Christova, Editor. 2017, Academic Press. p. 161-194.

21. Tong, L., *Structure and function of biotin-dependent carboxylases.* Cellular and Molecular Life Sciences, 2013. **70**(5): p. 863-891.

22. Quadri, L.E., et al., *Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases.* Biochemistry, 1998. **37**(6): p. 1585-95.

23. Pedersen, T.B., et al., *Speed dating for enzymes! Finding the perfect phosphopantetheinyl transferase partner for your polyketide synthase.* Microb Cell Fact, 2022. **21**(1): p. 9.

24. Clarke, K.M., et al., *In vivo reporter labeling of proteins via metabolic delivery of coenzyme A analogues.* J Am Chem Soc, 2005. **127**(32): p. 11234-5.

25. Nguyen, C., et al., *Trapping the dynamic acyl carrier protein in fatty acid biosynthesis.* Nature, 2014. **505**(7483): p. 427-31.

26. Konno, S., J.J. La Clair, and M.D. Burkart, *Trapping the Complex Molecular Machinery of Polyketide and Fatty Acid Synthases with Tunable Silylcyanohydrin Crosslinkers.* Angew Chem Int Ed Engl, 2018. **57**(52): p. 17009-17013.

27. Hibbing, M.E., et al., *Bacterial competition: surviving and thriving in the microbial jungle.* Nat Rev Microbiol, 2010. **8**(1): p. 15-25.

28. Weissman, K.J., *The structural biology of biosynthetic megaenzymes.* Nat Chem Biol, 2015. **11**(9): p. 660-70.

29. Hertweck, C., *The biosynthetic logic of polyketide diversity.* Angew Chem Int Ed Engl, 2009. **48**(26): p. 4688-716.

30. Firn, R.D. and C.G. Jones, *Natural products--a simple model to explain chemical diversity.* Nat Prod Rep, 2003. **20**(4): p. 382-91.

31. Carter, G.T., *Natural products and Pharma 2011: strategic changes spur new opportunities.* Nat Prod Rep, 2011. **28**(11): p. 1783-9.

32. Smanski, M.J., et al., *Synthetic biology to access and expand nature's chemical diversity.* Nat Rev Microbiol, 2016. **14**(3): p. 135-49.

33. Katz, L. and R.H. Baltz, *Natural product discovery: past, present, and future.* J Ind Microbiol Biotechnol, 2016. **43**(2-3): p. 155-76.

34. Weissman, K.J. and P.F. Leadlay, *Combinatorial biosynthesis of reduced polyketides.* Nat Rev Microbiol, 2005. **3**(12): p. 925-36.

35. Schmidt, E.W., *Trading molecules and tracking targets in symbiotic interactions.* Nat Chem Biol, 2008. **4**(8): p. 466-73.

36. Weissman, K.J., *Introduction to polyketide biosynthesis.* Methods Enzymol, 2009. **459**: p. 3-16.

37. Nguyen, T., et al., *Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection.* Nat Biotechnol, 2008. **26**(2): p. 225-33.

38. Helfrich, E.J. and J. Piel, *Biosynthesis of polyketides by trans-AT polyketide synthases.* Nat Prod Rep, 2016. **33**(2): p. 231-316.

39. Herbst, D.A., et al., *Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases.* Nature, 2016. **531**(7595): p. 533-7.

40. Smith, J.L. and D.H. Sherman, *An enzyme assembly line.* Science, 2008. **321**(5894): p. 1304-5.

41. Fischbach, M.A. and C.T. Walsh, *Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms.* Chem Rev, 2006. **106**(8): p. 3468-96.

42. Broadhurst, R.W., et al., *The structure of docking domains in modular polyketide synthases.* Chem Biol, 2003. **10**(8): p. 723-31.

43. Zeng, J., et al., *Portability and Structure of the Four-Helix Bundle Docking Domains of trans-Acyltransferase Modular Polyketide Synthases.* ACS Chem Biol, 2016. **11**(9): p. 2466-74.

44. Lohman, J.R., et al., *Structural and evolutionary relationships of "AT-less" type I polyketide synthase ketosynthases.* Proc Natl Acad Sci U S A, 2015. **112**(41): p. 12693-8.

45. Kautsar, S.A., et al., *MIBiG 2.0: a repository for biosynthetic gene clusters of known function.* Nucleic Acids Res, 2020. **48**(D1): p. D454-D458.

46. Helfrich, E.J.N., et al., *Automated structure prediction of trans-acyltransferase polyketide synthase products.* Nat Chem Biol, 2019. **15**(8): p. 813-821.

47. Jenke-Kodama, H., et al., *Evolutionary implications of bacterial polyketide synthases.* Mol Biol Evol, 2005. **22**(10): p. 2027-39.

48. Donadio, S., et al., *Modular organization of genes required for complex polyketide biosynthesis.* Science, 1991. **252**(5006): p. 675-9.

49. Herbst, D.A., *Functional architectures of polyketide synthases*, in *Faculty of Science*. 2017, University of Basel. p. 1-155.

50. Aparicio, J.F., et al., *Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: analysis of the enzymatic domains in the modular polyketide synthase.* Gene, 1996. **169**(1): p. 9-16.

51. El-Sayed, A.K., et al., *Characterization of the mupirocin biosynthesis gene cluster from Pseudomonas fluorescens NCIMB 10586.* Chem Biol, 2003. **10**(5): p. 419-30.

52. Piel, J., *Biosynthesis of polyketides by trans-AT polyketide synthases.* Nat Prod Rep, 2010. **27**(7): p. 996-1047.

53. Jenner, M., et al., *Acyl-chain elongation drives ketosynthase substrate selectivity in trans-acyltransferase polyketide synthases.* Angew Chem Int Ed Engl, 2015. **54**(6): p. 1817-21.

54. Sugimoto, Y., et al., *Freedom and constraint in engineered noncolinear polyketide assembly lines.* Chem Biol, 2015. **22**(2): p. 229-40.

55. Busch, B., et al., *Multifactorial control of iteration events in a modular polyketide assembly line.* Angew Chem Int Ed Engl, 2013. **52**(20): p. 5285-9.

56. Zhang, L., et al., *Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides.* Angewandte Chemie International Edition, 2017. **56**(7): p. 1740-1745.

57. Keatinge-Clay, A.T., *Polyketide Synthase Modules Redefined.* Angewandte Chemie International Edition, 2017. **56**(17): p. 4658-4660.

58. Vander Wood, D.A. and A.T. Keatinge-Clay, *The modules of trans-acyltransferase assembly lines redefined with a central acyl carrier protein.* Proteins, 2018. **86**(6): p. 664-675.

59. Lowry, B., et al., *A Turnstile Mechanism for the Controlled Growth of Biosynthetic Intermediates on Assembly Line Polyketide Synthases.* ACS Cent Sci, 2016. **2**(1): p. 14-20.

60. Bode, J.W. and E.M. Carreira, *Stereoselective syntheses of epothilones A and B via directed nitrile oxide cycloaddition.* J Am Chem Soc, 2001. **123**(15): p. 3611-2.

61. Muri, D. and E.M. Carreira, *Stereoselective synthesis of erythronolide A via nitrile oxide cycloadditions and related studies.* J Org Chem, 2009. **74**(22): p. 8695-712.

62. Xue, M., et al., *Structure elucidation of colibactin and its DNA cross-links.* Science, 2019. **365**(6457).

63. Charest, M.G., D.R. Siegel, and A.G. Myers, *Synthesis of (-)-tetracycline.* J Am Chem Soc, 2005. **127**(23): p. 8292-3.

64. Drufva, E.E., E.G. Hix, and C.B. Bailey, *Site directed mutagenesis as a precision tool to enable synthetic biology with engineered modular polyketide synthases.* Synth Syst Biotechnol, 2020. **5**(2): p. 62-80.

65. McDaniel, R., et al., *Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products.* Proc Natl Acad Sci U S A, 1999. **96**(5): p. 1846-51.

66. Wu, Q., Y. Zhi, and Y. Xu, *Systematically engineering the biosynthesis of a green biosurfactant surfactin by Bacillus subtilis 168.* Metab Eng, 2019. **52**: p. 87-97.

67. Klaus, M. and M. Grininger, *Engineering strategies for rational polyketide synthase design.* Nat Prod Rep, 2018. **35**(10): p. 1070-1081.

68. Zargar, A., et al., *Leveraging microbial biosynthetic pathways for the generation of 'drop-in' biofuels.* Curr Opin Biotechnol, 2017. **45**: p. 156-163.

69. Liao, J.C., et al., *Fuelling the future: microbial engineering for the production of sustainable biofuels.* Nat Rev Microbiol, 2016. **14**(5): p. 288-304.

70. Eng, C.H., et al., *Alteration of Polyketide Stereochemistry from anti to syn by a Ketoreductase Domain Exchange in a Type I Modular Polyketide Synthase Subunit.* Biochemistry, 2016. **55**(12): p. 1677-80.

71. Yuzawa, S., et al., *Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production.* ACS Synth Biol, 2017. **6**(1): p. 139-147.

72. Menzella, H.G., et al., *Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes.* Nat Biotechnol, 2005. **23**(9): p. 1171-6.

73. Klaus, M., et al., *Protein-Protein Interactions, Not Substrate Recognition, Dominate the Turnover of Chimeric Assembly Line Polyketide Synthases*.* Journal of Biological Chemistry, 2016. **291**(31): p. 16404-16415.

74. Murphy, A.C., et al., *Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation.* Chem Commun (Camb), 2016. **52**(54): p. 8373-6.

75. Kalkreuter, E., et al., *Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender Units.* J Am Chem Soc, 2019. **141**(5): p. 1961-1969.

76. Klaus, M., L. Buyachuihan, and M. Grininger, *Ketosynthase Domain Constrains the Design of Polyketide Synthases.* ACS Chem Biol, 2020. **15**(9): p. 2422-2432.

77. Kumar, P., et al., *Intermodular communication in modular polyketide synthases: structural and mutational analysis of linker mediated protein-protein recognition.* J Am Chem Soc, 2003. **125**(14): p. 4097-102.

78. Buchholz, T.J., et al., *Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains.* ACS Chem Biol, 2009. **4**(1): p. 41-52.

79. Wu, N., et al., *Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules.* J Am Chem Soc, 2001. **123**(27): p. 6465-74.

80. Whicher, J.R., et al., *Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis.* Chem Biol, 2013. **20**(11): p. 1340-51.

81. Lowry, B., et al., *In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase.* J Am Chem Soc, 2013. **135**(45): p. 16809-12.

82. Klaus, M., et al., *Engineering of Chimeric Polyketide Synthases Using SYNZIP Docking Domains.* ACS Chem Biol, 2019. **14**(3): p. 426-433.

83. Grote, M., et al., *Identification of crucial bottlenecks in engineered polyketide biosynthesis.* Org Biomol Chem, 2019. **17**(26): p. 6374-6385.

84. Alanjary, M., et al., *Computer-aided re-engineering of nonribosomal peptide and polyketide biosynthetic assembly lines.* Nat Prod Rep, 2019. **36**(9): p. 1249-1261.

85. Smith, S. and S.C. Tsai, *The type I fatty acid and polyketide synthases: a tale of two megasynthases.* Nat Prod Rep, 2007. **24**(5): p. 1041-72.

86. Malpartida, F. and D.A. Hopwood, *Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host.* Nature, 1984. **309**(5967): p. 462-464.

87. Dutta, S., et al., *Structure of a modular polyketide synthase.* Nature, 2014. **510**(7506): p. 512-7.

88. Keatinge-Clay, A.T., *The structures of type I polyketide synthases.* Nat Prod Rep, 2012. **29**(10): p. 1050-73.

89. Keatinge-Clay, A., *Crystal structure of the erythromycin polyketide synthase dehydratase.* J Mol Biol, 2008. **384**(4): p. 941-53.

90. Akey, D.L., et al., *Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway.* Structure, 2010. **18**(1): p. 94-105.

91. Ames, B.D., et al., *Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis.* Proc Natl Acad Sci U S A, 2012. **109**(28): p. 11144-9.

92. Gay, D.C., et al., *A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase.* Structure, 2014. **22**(3): p. 444-51.

93. Whicher, J.R., et al., *Structural rearrangements of a polyketide synthase module during its catalytic cycle.* Nature, 2014. **510**(7506): p. 560-4.

94. Davison, J., et al., *Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module.* Chemical Science, 2014. **5**(8): p. 3081-3095.

95. Edwards, A.L., et al., *Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase.* J Mol Biol, 2014. **426**(11): p. 2229-45.

96. Tang, Y., et al., *The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase.* Proc Natl Acad Sci U S A, 2006. **103**(30): p. 11124-9.

97. Tang, Y., et al., *Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase.* Chem Biol, 2007. **14**(8): p. 931-43.

98. Dodge, G.J., F.P. Maloney, and J.L. Smith, *Protein-protein interactions in "cis-AT" polyketide synthases.* Nat Prod Rep, 2018. **35**(10): p. 1082-1096.

99. Maier, T., S. Jenni, and N. Ban, *Architecture of mammalian fatty acid synthase at 4.5 A resolution.* Science, 2006. **311**(5765): p. 1258-1262.

100. Bretschneider, T., et al., *Vinylogous chain branching catalysed by a dedicated polyketide synthase module.* Nature, 2013. **502**(7469): p. 124-8.

101. Butcher, R.A., et al., *The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis.* Proc Natl Acad Sci U S A, 2007. **104**(5): p. 1506-9.

102. Taniguchi, Y., et al., *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.* Science, 2010. **329**(5991): p. 533-8.

103. Straight, P.D., et al., *A singular enzymatic megacomplex from Bacillus subtilis.* Proc Natl Acad Sci U S A, 2007. **104**(1): p. 305-10.

104. Gay, D.C., et al., *The LINKS motif zippers trans-acyltransferase polyketide synthase assembly lines into a biosynthetic megacomplex.* J Struct Biol, 2016. **193**(3): p. 196-205.

105. Jain, M. and J.S. Cox, *Interaction between polyketide synthase and transporter suggests coupled synthesis and export of virulence lipid in M. tuberculosis.* PLoS Pathog, 2005. **1**(1): p. e2.

106. Tillett, D., et al., *Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide-polyketide synthetase system.* Chem Biol, 2000. **7**(10): p. 753-64.

107. Mattheus, W., et al., *Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of its biosynthesis gene cluster.* Chem Biol, 2010. **17**(2): p. 149-59.

108. Porter, J.L., et al., *The cell wall-associated mycolactone polyketide synthases are necessary but not sufficient for mycolactone biosynthesis.* PLoS One, 2013. **8**(7): p. e70520.

109. Kapur, S., et al., *Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase.* Proc Natl Acad Sci U S A, 2010. **107**(51): p. 22066-71.

110. Cheng, Y.Q., G.L. Tang, and B. Shen, *Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3149-54.

111. Tang, G.L., Y.Q. Cheng, and B. Shen, *Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase.* Chem Biol, 2004. **11**(1): p. 33-45.

112. Quin, M.B., et al., *Spatial organization of multi-enzyme biocatalytic cascades.* Org Biomol Chem, 2017. **15**(20): p. 4260-4271.

113. Wei, J. and L. Tong, *Crystal structure of the 500-kDa yeast acetyl-CoA carboxylase holoenzyme dimer.* Nature, 2015. **526**(7575): p. 723-7.

114. Munday, M.R., *Regulation of mammalian acetyl-CoA carboxylase.* Biochemical Society Transactions, 2002. **30**: p. 1059-1064.

115. Abu-Elheiga, L., et al., *Human acetyl-CoA carboxylase 2. Molecular cloning, characterization, chromosomal mapping, and evidence for two isoforms.* J Biol Chem, 1997. **272**(16): p. 10669-77.

116. Abu-Elheiga, L., et al., *The subcellular localization of acetyl-CoA carboxylase 2.* Proc Natl Acad Sci U S A, 2000. **97**(4): p. 1444-9.

117. Thampy, K.G., *Formation of malonyl coenzyme A in rat heart. Identification and purification of an isozyme of A carboxylase from rat heart.* J Biol Chem, 1989. **264**(30): p. 17631-4.

118. Park, C.E., et al., *Rapid increase of cytosolic content of acetyl-CoA carboxylase isoforms in H9c2 cells by short-term treatment with insulin and okadaic acid.* Exp Mol Med, 1998. **30**(2): p. 73-9.

119. Attwood, P.V. and J.C. Wallace, *Chemical and catalytic mechanisms of carboxyl transfer reactions in biotin-dependent enzymes.* Acc Chem Res, 2002. **35**(2): p. 113-20.

120. Lee, C.K., et al., *Biotinoyl domain of human acetyl-CoA carboxylase: Structural insights into the carboxyl transfer mechanism.* Proteins, 2008. **72**(2): p. 613-24.

121. Chou, C.Y., L.P. Yu, and L. Tong, *Crystal structure of biotin carboxylase in complex with substrates and implications for its catalytic mechanism.* J Biol Chem, 2009. **284**(17): p. 11690-7.

122. Hunkeler, M., et al., *The dynamic organization of fungal acetyl-CoA carboxylase.* Nat Commun, 2016. **7**: p. 11196.

123. Hunkeler, M., et al., *Structural basis for regulation of human acetyl-CoA carboxylase.* Nature, 2018. **558**(7710): p. 470-474.

124. Brown, N.F., et al., *The mammalian target of rapamycin regulates lipid metabolism in primary cultures of rat hepatocytes.* Metabolism, 2007. **56**(11): p. 1500-7.

125. Guri, Y., et al., *mTORC2 Promotes Tumorigenesis via Lipid Synthesis.* Cancer Cell, 2017. **32**(6): p. 807-823 e12.

126. Martin, D.B. and P.R. Vagelos, *The mechanism of tricarboxylic acid cycle regulation of fatty acid synthesis.* J Biol Chem, 1962. **237**: p. 1787-92.

127. Kleinschmidt, A.K., J. Moss, and D.M. Lane, *Acetyl coenzyme A carboxylase: filamentous nature of the animal enzymes.* Science, 1969. **166**(3910): p. 1276-8.

128. Maier, T., et al., *Structure and function of eukaryotic fatty acid synthases.* Q Rev Biophys, 2010. **43**(3): p. 373-422.

129. Wakil, S.J., J.K. Stoops, and V.C. Joshi, *Fatty acid synthesis and its regulation.* Annual review of biochemistry, 1983. **52**(1): p. 537-579.

130. Munday, M.R., *Regulation of mammalian acetyl-CoA carboxylase.* Biochem Soc Trans, 2002. **30**(Pt 6): p. 1059-64.

131. Taketa, K. and B.M. Pogell, *The effect of palmityl coenzyme A on glucose 6-phosphate dehydrogenase and other enzymes.* Journal of Biological Chemistry, 1966. **241**(3): p. 720-726.

132. Goodridge, A.G., *Regulation of the activity of acetyl coenzyme A carboxylase by palmitoyl coenzyme A and citrate.* J Biol Chem, 1972. **247**(21): p. 6946-52.

133. Carling, D., V.A. Zammit, and D.G. Hardie, *A common bicyclic protein kinase cascade inactivates the regulatory enzymes of fatty acid and cholesterol biosynthesis.* FEBS Lett, 1987. **223**(2): p. 217-22.

134. Ha, J., et al., *Critical phosphorylation sites for acetyl-CoA carboxylase activity.* J Biol Chem, 1994. **269**(35): p. 22162-8.

135. Davies, S.P., A.T. Sim, and D.G. Hardie, *Location and function of three sites phosphorylated on rat acetyl-CoA carboxylase by the AMP-activated protein kinase.* Eur J Biochem, 1990. **187**(1): p. 183-90.

136. Naini, A., F. Sasse, and M. Bronstrup, *The intriguing chemistry and biology of soraphens.* Nat Prod Rep, 2019. **36**(10): p. 1394-1411.

137. Gaussin, V., et al., *Activation of hepatic acetyl-CoA carboxylase by glutamate and Mg2+ is mediated by protein phosphatase-2A.* Biochem J, 1996. **316 (Pt 1)**: p. 217-24.

138. Wang, B., et al., *Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response.* Science, 2007. **316**(5828): p. 1194-8.

139. Mehrgou, A. and M. Akouchekian, *The importance of BRCA1 and BRCA2 genes mutations in breast cancer development.* Med J Islam Repub Iran, 2016. **30**: p. 369.

140. Ford, D., et al., *Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium.* Lancet, 1994. **343**(8899): p. 692-5.

141. Brunet, J., et al., *BRCA1 and acetyl-CoA carboxylase: the metabolic syndrome of breast cancer.* Mol Carcinog, 2008. **47**(2): p. 157-63.

142. Ray, H., et al., *Cell cycle regulation of the BRCA1/acetyl-CoA-carboxylase complex.* Biochem Biophys Res Commun, 2009. **378**(3): p. 615-9.

143. Shen, Y. and L. Tong, *Structural Evidence for Direct Interactions between the BRCT Domains of Human BRCA1 and a Phospho-peptide from Human ACC1.* Biochemistry, 2008. **47**(21): p. 5767-5773.

144. Kim, C.-W., et al., *Induced polymerization of mammalian acetyl-CoA carboxylase by MIG12 provides a tertiary level of regulation of fatty acid synthesis.* Proceedings of the National Academy of Sciences, 2010. **107**(21): p. 9626-9631.

145. Ueda, K., et al., *Prolyl isomerase Pin1 binds to and stabilizes acetyl CoA carboxylase 1 protein, thereby supporting cancer cell proliferation.* Oncotarget, 2019. **10**(17): p. 1637-1648.

146. German, N.J., et al., *PHD3 Loss in Cancer Enables Metabolic Reliance on Fatty Acid Oxidation via Deactivation of ACC2.* Mol Cell, 2016. **63**(6): p. 1006-20.

147. Loenarz, C. and C.J. Schofield, *Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases.* Trends Biochem Sci, 2011. **36**(1): p. 7-18.

148. Waldrop, G.L., I. Rayment, and H.M. Holden, *Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase.* Biochemistry, 1994. **33**(34): p. 10249-56.

149. Shen, Y., et al., *A mechanism for the potent inhibition of eukaryotic acetyl-coenzyme A carboxylase by soraphen A, a macrocyclic polyketide natural product.* Mol Cell, 2004. **16**(6): p. 881-91.

150. Zhang, H., et al., *Crystal structure of the carboxyltransferase domain of acetyl-coenzyme A carboxylase.* Science, 2003. **299**(5615): p. 2064-7.

151. Zhang, H., et al., *Crystal structure of the carboxyltransferase domain of acetyl-coenzyme A carboxylase in complex with CP-640186.* Structure, 2004. **12**(9): p. 1683-91.

152. Shi, S., et al., *Improving production of malonyl coenzyme A-derived metabolites by abolishing Snf1-dependent regulation of Acc1.* mBio, 2014. **5**(3): p. e01130-14.

153. Shi, L. and B.P. Tu, *Acetyl-CoA and the regulation of metabolism: mechanisms and consequences.* Curr Opin Cell Biol, 2015. **33**: p. 125-31.

154. Yoon, S., et al., *Up-regulation of Acetyl-CoA Carboxylase α and Fatty Acid Synthase by Human Epidermal Growth Factor Receptor 2 at the Translational Level in Breast Cancer Cells*.* Journal of Biological Chemistry, 2007. **282**(36): p. 26122-26131.

155. Sinilnikova, O.M., et al., *Acetyl-CoA carboxylase alpha gene and breast cancer susceptibility.* Carcinogenesis, 2004. **25**(12): p. 2417-24.

156. Swinnen, J.V., K. Brusselmans, and G. Verhoeven, *Increased lipogenesis in cancer cells: new players, novel targets.* Curr Opin Clin Nutr Metab Care, 2006. **9**(4): p. 358-65.

157. Swinnen, J.V., et al., *Selective activation of the fatty acid synthesis pathway in human prostate cancer.* Int J Cancer, 2000. **88**(2): p. 176-9.

158. Yahagi, N., et al., *Co-ordinate activation of lipogenic enzymes in hepatocellular carcinoma.* Eur J Cancer, 2005. **41**(9): p. 1316-22.

159. Moncur, J.T., et al., *The "Spot 14" gene resides on the telomeric end of the 11q13 amplicon and is expressed in lipogenic breast cancers: Implications for control of tumor metabolism.* Proceedings of the National Academy of Sciences, 1998. **95**(12): p. 6989-6994.

160. Milgraum, L.Z., et al., *Enzymes of the fatty acid synthesis pathway are highly expressed in in situ breast carcinoma.* Clin Cancer Res, 1997. **3**(11): p. 2115-20.

161. Lei, Y., et al., *Carbohydrate response element binding protein (ChREBP) correlates with colon cancer progression and contributes to cell proliferation.* Sci Rep, 2020. **10**(1): p. 4233.

162. Wang, C., et al., *The acetyl-CoA carboxylase enzyme: a target for cancer therapy?* Expert Rev Anticancer Ther, 2015. **15**(6): p. 667-76.

163. Abramson, H.N., *The lipogenesis pathway as a cancer target.* Journal of medicinal chemistry, 2011. **54**(16): p. 5615-5638.

164. Eckel, R.H., S.M. Grundy, and P.Z. Zimmet, *The metabolic syndrome.* Lancet, 2005. **365**(9468): p. 1415-28.

165. Chen, L., et al., *Acetyl-CoA carboxylase (ACC) as a therapeutic target for metabolic syndrome and recent developments in ACC1/2 inhibitors.* Expert Opinion on Investigational Drugs, 2019. **28**(10): p. 917-930.

166. Glien, M., et al., *Stimulation of fat oxidation, but no sustained reduction of hepatic lipids by prolonged pharmacological inhibition of acetyl CoA carboxylase.* Horm Metab Res, 2011. **43**(9): p. 601-6.

167. Brusselmans, K., et al., *RNA interference-mediated silencing of the acetyl-CoA-carboxylase-alpha gene induces growth inhibition and apoptosis of prostate cancer cells.* Cancer Res, 2005. **65**(15): p. 6719-25.

168. Olsen, A.M., et al., *Fatty acid synthesis is a therapeutic target in human liposarcoma.* Int J Oncol, 2010. **36**(5): p. 1309-1314.

169. Wang, C., et al., *Acetyl-CoA carboxylase-α as a novel target for cancer therapy.* FBS, 2010. **2**(2): p. 515-526.

170. Beckers, A., et al., *Chemical inhibition of acetyl-CoA carboxylase induces growth arrest and cytotoxicity selectively in cancer cells.* Cancer Res, 2007. **67**(17): p. 8180-7.

171. Keil, S., et al., *Identification and Synthesis of Novel Inhibitors of Acetyl-CoA Carboxylase with in Vitro and in Vivo Efficacy on Fat Oxidation.* Journal of Medicinal Chemistry, 2010. **53**(24): p. 8679-8687.

172. Chow, J.D., et al., *Genetic inhibition of hepatic acetyl-CoA carboxylase activity increases liver fat and alters global protein acetylation.* Mol Metab, 2014. **3**(4): p. 419-31.

173. Ronnebaum, S.M., et al., *Chronic suppression of acetyl-CoA carboxylase 1 in beta-cells impairs insulin secretion via inhibition of glucose rather than lipid metabolism.* J Biol Chem, 2008. **283**(21): p. 14248-56.

174. Galdieri, L. and A. Vancura, *Acetyl-CoA carboxylase regulates global histone acetylation.* J Biol Chem, 2012. **287**(28): p. 23865-76.

175. Paik, W.K., et al., *Nonenzymatic acetylation of histones with acetyl-CoA.* Biochim Biophys Acta, 1970. **213**(2): p. 513-22.

176. Walker, K.A., et al., *Fluazifop, a grass-selective herbicide which inhibits acetyl-CoA carboxylase in sensitive plant species.* The Biochemical journal, 1988. **254**(1): p. 307-310.

177. Kemal, C. and J.E. Casida, *Coenzyme a esters of 2-aryloxyphenoxypropionate herbicides and 2-arylpropionate antiinflammatory drugs are potent and stereoselective inhibitors of rat liver acetyl-CoA carboxylase.* Life Sciences, 1992. **50**(7): p. 533-540.

178. Kemp, R.G., *Rabbit liver phosphofructokinase. Comparison of some properties with those of muscle phosphofructokinase.* J Biol Chem, 1971. **246**(1): p. 245-52.

179. Frey, T.G., D. Eisenberg, and F.A. Eiserling, *Glutamine synthetase forms three- and seven-stranded helical cables.* Proc Natl Acad Sci U S A, 1975. **72**(9): p. 3402-6.

180. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.

181. Menny, A., et al., *CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers.* Nature Communications, 2018. **9**(1): p. 5316.

182. Webb, B.A., et al., *The glycolytic enzyme phosphofructokinase-1 assembles into filaments.* J Cell Biol, 2017. **216**(8): p. 2305-2313.

183. Park, C.K. and N.C. Horton, *Novel insights into filament-forming enzymes.* Nat Rev Mol Cell Biol, 2020. **21**(1): p. 1-2.

184. O'Connell, J.D., et al., *A proteomic survey of widespread protein aggregation in yeast.* Mol Biosyst, 2014. **10**(4): p. 851-861.

185. Lynch, E.M., et al., *Human CTP synthase filament structure reveals the active enzyme conformation.* Nat Struct Mol Biol, 2017. **24**(6): p. 507-514.

186. Polley, S., D. Lyumkis, and N.C. Horton, *Mechanism of Filamentation-Induced Allosteric Activation of the SgrAI Endonuclease.* Structure, 2019. **27**(10): p. 1497-1507 e3.

187. Schuchmann, K., J. Vonck, and V. Muller, *A bacterial hydrogen-dependent CO2 reductase forms filamentous structures.* FEBS J, 2016. **283**(7): p. 1311-22.

188. Narayanaswamy, R., et al., *Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation.* Proceedings of the National Academy of Sciences, 2009. **106**(25): p. 10147-10152.

189. Suresh, H.G., et al., *Prolonged starvation drives reversible sequestration of lipid biosynthetic enzymes and organelle reorganization in Saccharomyces cerevisiae.* Mol Biol Cell, 2015. **26**(9): p. 1601-15.

190. Shen, Q.J., et al., *Filamentation of Metabolic Enzymes in Saccharomyces cerevisiae.* J Genet Genomics, 2016. **43**(6): p. 393-404.

191. Liu, J.L., *Intracellular compartmentation of CTP synthase in Drosophila.* J Genet Genomics, 2010. **37**(5): p. 281-96.

192. Dobro, M.J., et al., *Uncharacterized Bacterial Structures Revealed by Electron Cryotomography.* J Bacteriol, 2017. **199**(17).

193. Celler, K., et al., *Multidimensional view of the bacterial cytoskeleton.* J Bacteriol, 2013. **195**(8): p. 1627-36.

194. Noree, C., et al., *Identification of novel filament-forming proteins in Saccharomyces cerevisiae and Drosophila melanogaster.* J Cell Biol, 2010. **190**(4): p. 541-51.

195. Bitinaite, J. and I. Schildkraut, *Self-generated DNA termini relax the specificity of SgrAI restriction endonuclease.* Proc Natl Acad Sci U S A, 2002. **99**(3): p. 1164-9.

196. Barahona, C.J., et al., *The Need for Speed: Run-On Oligomer Filament Formation Provides Maximum Speed with Maximum Sequestration of Activity.* J Virol, 2019. **93**(5).

197. Woodward, J.D., et al., *Substrate specificity of plant nitrilase complexes is affected by their helical twist.* Communications biology, 2018. **1**(1): p. 1-10.

198. Prouteau, M. and R. Loewith, *Regulation of cellular metabolism through phase separation of enzymes.* Biomolecules, 2018. **8**(4): p. 160.

199. Park, C.K. and N.C. Horton, *Structures, functions, and mechanisms of filament forming enzymes: a renaissance of enzyme filamentation.* Biophys Rev, 2019. **11**(6): p. 927-994.

200. Proschel, M., et al., *Engineering of Metabolic Pathways by Artificial Enzyme Channels.* Front Bioeng Biotechnol, 2015. **3**: p. 168.

201. Bäuerlein, F.J.B. and W. Baumeister, *Towards Visual Proteomics at High Resolution.* Journal of Molecular Biology, 2021. **433**(20): p. 167187.

202. Petrov, P.N., H. Müller, and R.M. Glaeser, *Perspective: Emerging strategies for determining atomic-resolution structures of macromolecular complexes within cells.* Journal of Structural Biology, 2022. **214**(1): p. 107827.

203. Scales, B.S., et al., *Microbiology, genomics, and clinical significance of the Pseudomonas fluorescens species complex, an unappreciated colonizer of humans.* Clin Microbiol Rev, 2014. **27**(4): p. 927-48.

204. McNear Jr, D.H., *The rhizosphere-roots, soil and everything in between.* Nature Education Knowledge, 2013. **4**(3): p. 1.

205. Gross, H. and J.E. Loper, *Genomics of secondary metabolite production by Pseudomonas spp.* Nat Prod Rep, 2009. **26**(11): p. 1408-46.

206. Whatling, C.A., et al., *Identification of a 60-Kb Region of the Chromosome of Pseudomonas-Fluorescens Ncib-10586 Required for the Biosynthesis of Pseudomonic Acid (Mupirocin).* Microbiology-Uk, 1995. **141**(4): p. 973-982.

207. Thomas, C.M., et al., *Resistance to and synthesis of the antibiotic mupirocin.* Nat Rev Microbiol, 2010. **8**(4): p. 281-9.

208. Tucaliuc, A., et al., *Mupirocin: applications and production.* Biotechnol Lett, 2019. **41**(4-5): p. 495-502.

209. Fuller, A.T., et al., *Pseudomonic acid: an antibiotic produced by Pseudomonas fluorescens.* Nature, 1971. **234**(5329): p. 416-7.

210. Sutherland, R., et al., *Antibacterial activity of mupirocin (pseudomonic acid), a new antibiotic for topical use.* Antimicrob Agents Chemother, 1985. **27**(4): p. 495-8.

211. Thompson, T., *The staggering death toll of drug-resistant bacteria.* Nature, 2022.

212. Khoshnood, S., et al., *A review on mechanism of action, resistance, synergism, and clinical implications of mupirocin against Staphylococcus aureus.* Biomed Pharmacother, 2019. **109**: p. 1809-1818.

213. Antimicrobial Resistance, C., *Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis.* Lancet, 2022. **399**(10325): p. 629-655.

214. Matthijs, S., et al., *Antimicrobial properties of Pseudomonas strains producing the antibiotic mupirocin.* Res Microbiol, 2014. **165**(8): p. 695-704.

215. Hothersall, J., et al., *Mutational analysis reveals that all tailoring region genes are required for production of polyketide antibiotic mupirocin by pseudomonas fluorescens:*

pseudomonic acid B biosynthesis precedes pseudomonic acid A. J Biol Chem, 2007. **282**(21): p. 15451-61.

216. Connolly, J.A., et al., *Defining the genes for the final steps in biosynthesis of the complex polyketide antibiotic mupirocin by Pseudomonas fluorescens NCIMB10586.* Sci Rep, 2019. **9**(1): p. 1542.

217. El-Sayed, A.K., J. Hothersall, and C.M. Thomas, *Quorum-sensing-dependent regulation of biosynthesis of the polyketide antibiotic mupirocin in Pseudomonas fluorescens NCIMB 10586.* Microbiology (Reading), 2001. **147**(Pt 8): p. 2127-2139.

218. Hmelo, L.R., et al., *Precision-engineering the Pseudomonas aeruginosa genome with two-step allelic exchange.* Nat Protoc, 2015. **10**(11): p. 1820-41.

219. Rietsch, A., et al., *ExsE, a secreted regulator of type III secretion genes in Pseudomonas aeruginosa.* Proc Natl Acad Sci U S A, 2005. **102**(22): p. 8006-11.

220. Hoang, T.T., et al., *A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked Pseudomonas aeruginosa mutants.* Gene, 1998. **212**(1): p. 77-86.

221. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature Methods, 2009. **6**(5): p. 343-345.

222. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.

223. Ringel, P.D., D. Hu, and M. Basler, *The Role of Type VI Secretion System Effectors in Target Cell Lysis and Subsequent Horizontal Gene Transfer.* Cell Rep, 2017. **21**(13): p. 3927-3940.

224. Thevenaz, P., U.E. Ruttimann, and M. Unser, *A pyramid approach to subpixel registration based on intensity.* IEEE Trans Image Process, 1998. **7**(1): p. 27-41.

225. Hockett, K.L. and D.A. Baltrus, *Use of the Soft-agar Overlay Technique to Screen for Bacterially Produced Inhibitory Compounds.* J Vis Exp, 2017(119).

226. Mattson, A.M., C.O. Jensen, and R.A. Dutcher, *Triphenyltetrazolium Chloride as a Dye for Vital Tissues.* Science, 1947. **106**(2752): p. 294-5.

227. Hostettler, L., et al., *The Bright Fluorescent Protein mNeonGreen Facilitates Protein Expression Analysis In Vivo.* G3 (Bethesda), 2017. **7**(2): p. 607-615.

228. Bindels, D.S., et al., *mScarlet: a bright monomeric red fluorescent protein for cellular imaging.* Nat Methods, 2017. **14**(1): p. 53-56.

229. Haines, A.S., et al., *A conserved motif flags acyl carrier proteins for beta-branching in polyketide synthesis.* Nat Chem Biol, 2013. **9**(11): p. 685-692.

230. Rahman, A.S., et al., *Tandemly duplicated acyl carrier proteins, which increase polyketide antibiotic production, can apparently function either in parallel or in series.* J Biol Chem, 2005. **280**(8): p. 6399-408.

231. Blomfield, I.C., et al., *Allelic exchange in Escherichia coli using the Bacillus subtilis sacB gene and a temperature-sensitive pSC101 replicon.* Mol Microbiol, 1991. **5**(6): p. 1447-57.

232. Billinton, N. and A.W. Knight, *Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence.* Anal Biochem, 2001. **291**(2): p. 175-97.

233. Xu, X.P., et al., *Localization of the ActIII actinorhodin polyketide ketoreductase to the cell wall.* FEMS Microbiol Lett, 2008. **287**(1): p. 15-21.

234. Bagde, S.R., et al., *Modular polyketide synthase contains two reaction chambers that operate asynchronously.* Science, 2021. **374**(6568): p. 723-729.

235. Cogan, D.P., et al., *Mapping the catalytic conformations of an assembly-line polyketide synthase module.* Science, 2021. **374**(6568): p. 729-734.

236. Li, X., et al., *Structure-Function Analysis of the Extended Conformation of a Polyketide Synthase Module.* J Am Chem Soc, 2018. **140**(21): p. 6518-6521.

237. Reece-Hoyes, J.S. and A.J.M. Walhout, *Gateway Recombinational Cloning.* Cold Spring Harb Protoc, 2018. **2018**(1).

238. Clark, D.P., N.J. Pazdernik, and M.R. McGehee, *Chapter 7 - Cloning Genes for Synthetic Biology*, in *Molecular Biology (Third Edition)*, D.P. Clark, N.J. Pazdernik, and M.R. McGehee, Editors. 2019, Academic Cell. p. 199-239.

239. Fitzgerald, D.J., et al., *Protein complex expression by using multigene baculoviral vectors.* Nature Methods, 2006. **3**(12): p. 1021-1032.

240. Sari, D., et al., *The MultiBac Baculovirus/Insect Cell Expression Vector System for Producing Complex Protein Biologics.* Adv Exp Med Biol, 2016. **896**: p. 199-215.

241. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* Nat Methods, 2017. **14**(3): p. 290-296.

242. Hofmeister, F., *Zur Lehre von der Wirkung der Salze.* Archiv für experimentelle Pathologie und Pharmakologie, 1888. **24**(4): p. 247-260.

243. Rosenthal, P.B. and R. Henderson, *Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy.* J Mol Biol, 2003. **333**(4): p. 721-45.

244. Wang, J., et al., *Structural basis for the biosynthesis of lovastatin.* Nat Commun, 2021. **12**(1): p. 867.

245. Betancor, L., et al., *Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after coexpression with Streptomyces chaperonins in Escherichia coli.* Chembiochem, 2008. **9**(18): p. 2962-6.

246. Mutka, S.C., et al., *Heterologous production of epothilone C and D in Escherichia coli.* Biochemistry, 2006. **45**(4): p. 1321-30.

247. Gibson, D.G., et al., *Creation of a bacterial cell controlled by a chemically synthesized genome.* Science, 2010. **329**(5987): p. 52-6.

248. Gibson, D.G., et al., *Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome.* Science, 2008. **319**(5867): p. 1215-20.

249. Fredens, J., et al., *Total synthesis of Escherichia coli with a recoded genome.* Nature, 2019. **569**(7757): p. 514-518.

250. Herbst, D.A., et al., *The structural organization of substrate loading in iterative polyketide synthases.* Nat Chem Biol, 2018. **14**(5): p. 474-479.

251. Guzman, K.M. and C. Khosla, *Fragment antigen binding domains (Fabs) as tools to study assembly-line polyketide synthases.* Synthetic and Systems Biotechnology, 2022. **7**(1): p. 506-512.

252. Davis, T.D., J.M. Michaud, and M.D. Burkart, *Active site labeling of fatty acid and polyketide acyl-carrier protein transacylases.* Org Biomol Chem, 2019. **17**(19): p. 4720-4724.

253. Kosa, N.M., et al., *Reversible labeling of native and fusion-protein motifs.* Nat Methods, 2012. **9**(10): p. 981-4.

254. Beld, J., et al., *The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life.* Nat Prod Rep, 2014. **31**(1): p. 61-108.

255. Maschio, L., et al., *Cloning, expression, and purification of intact polyketide synthase modules.* Methods Enzymol, 2019. **617**: p. 63-82.

256. Tong, L., *Structure and function of biotin-dependent carboxylases.* Cell Mol Life Sci, 2013. **70**(5): p. 863-91.

257. Kaushik, V.K., et al., *Characterization of recombinant human acetyl-CoA carboxylase-2 steady-state kinetics.* Biochim Biophys Acta, 2009. **1794**(6): p. 961-7.

258. Wei, J., et al., *A unified molecular mechanism for the regulation of acetyl-CoA carboxylase by phosphorylation.* Cell Discov, 2016. **2**: p. 16044.

259. Woods, A., et al., *Yeast SNF1 is functionally related to mammalian AMP-activated protein kinase and regulates acetyl-CoA carboxylase in vivo.* Journal of Biological Chemistry, 1994. **269**(30): p. 19509-19515.

260. Ficarro, S.B., et al., *Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae.* Nat Biotechnol, 2002. **20**(3): p. 301-5.

261. Schorb, M., et al., *Software tools for automated transmission electron microscopy.* Nat Methods, 2019. **16**(6): p. 471-477.

262. Zivanov, J., T. Nakane, and S.H.W. Scheres, *A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis.* IUCrJ, 2019. **6**(Pt 1): p. 5-17.

263. Zivanov, J., T. Nakane, and S.H. Scheres, *Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1.* IUCrJ, 2020. **7**(2): p. 253-267.

264. Inc, S.B. *cryoSPARC*. 2016 [cited 2020 15.05]; Available from: https://cryosparc.com/.

265. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

266. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.

267. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1204-14.

268. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.

269. Adams, P.D., et al., *PHENIX: building new software for automated crystallographic structure determination.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 11): p. 1948-54.

270. Williams, C.J., et al., *MolProbity: More and better reference data for improved all-atom structure validation.* Protein Sci, 2018. **27**(1): p. 293-315.

271. Bai, N., et al., *Isothermal Analysis of ThermoFluor Data can readily provide Quantitative Binding Affinities.* Sci Rep, 2019. **9**(1): p. 2650.

272. Niebling, S., et al., *FoldAffinity: binding affinities from nDSF experiments.* Sci Rep, 2021. **11**(1): p. 9572.

273. Young, G., et al., *Quantitative mass imaging of single biological macromolecules.* Science, 2018. **360**(6387): p. 423-427.

274. Goddard, T.D., et al., *UCSF ChimeraX: Meeting modern challenges in visualization and analysis.* Protein Sci, 2018. **27**(1): p. 14-25.

275. PyMOL, *The PyMOL Molecular Graphics System, Version 2.0*. 2015, Schrödinger, LLC Schrödinger.

276. Khosla, C., et al., *Assembly line polyketide synthases: mechanistic insights and unsolved problems.* Biochemistry, 2014. **53**(18): p. 2875-83.

277. Klaus, M., et al., *Solution Structure and Conformational Flexibility of a Polyketide Synthase Module.* JACS Au, 2021. **1**(12): p. 2162-2171.

278. Wang, W., G. Zheng, and Y. Lu, *Recent Advances in Strategies for the Cloning of Natural Product Biosynthetic Gene Clusters.* Front Bioeng Biotechnol, 2021. **9**: p. 692797.

279. Massicard, J.-M., et al., *Engineering Modular Polyketide Biosynthesis in Streptomyces Using CRISPR/Cas: A Practical Guide*, in *Engineering Natural Product Biosynthesis: Methods and Protocols*, E. Skellam, Editor. 2022, Springer US: New York, NY. p. 173-200.

280. Hirano, M., et al., *A highly photostable and bright green fluorescent protein.* Nature Biotechnology, 2022.

281. Baumeister, W., *From proteomic inventory to architecture.* FEBS Lett, 2005. **579**(4): p. 933-7.

282. Beck, M. and W. Baumeister, *Cryo-Electron Tomography: Can it Reveal the Molecular Sociology of Cells in Atomic Detail?* Trends in Cell Biology, 2016. **26**(11): p. 825-837.

283. Al-Amoudi, A., L.P. Norlen, and J. Dubochet, *Cryo-electron microscopy of vitreous sections of native biological cells and tissues.* J Struct Biol, 2004. **148**(1): p. 131-5.

284. Rigort, A., et al., *Focused ion beam micromachining of eukaryotic cells for cryoelectron tomography.* Proc Natl Acad Sci U S A, 2012. **109**(12): p. 4449-54.

285. Basler, M., et al., *Type VI secretion requires a dynamic contractile phage tail-like structure.* Nature, 2012. **483**(7388): p. 182-6.

286. Schwartz, C.L., et al., *Cryo-fluorescence microscopy facilitates correlations between light and cryo-electron microscopy and reduces the rate of photobleaching.* Journal of microscopy, 2007. **227**(2): p. 98-109.

287. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.

288. Perrakis, A. and T.K. Sixma, *AI revolutions in biology: The joys and perils of AlphaFold.* EMBO Rep, 2021. **22**(11): p. e54046.

289. Varadi, M., et al., *AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models.* Nucleic Acids Res, 2022. **50**(D1): p. D439-D444.

290. Evans, R., et al., *Protein complex prediction with AlphaFold-Multimer.* bioRxiv, 2022: p. 2021.10.04.463034.

291. Zhang, W. and J. Liu, *Recent Advances in Understanding and Engineering Polyketide Synthesis.* F1000Res, 2016. **5**.

292. Maier, T., *Fatty acid synthases: Re-engineering biofactories.* Nat Chem Biol, 2017. **13**(4): p. 344-345.

293. Till, M. and P.R. Race, *Progress challenges and opportunities for the re-engineering of trans-AT polyketide synthases.* Biotechnol Lett, 2014. **36**(5): p. 877-88.

294. Menendez-Bravo, S., et al., *Expanding the chemical diversity of natural esters by engineering a polyketide-derived pathway into Escherichia coli.* Metabolic engineering, 2014. **24**: p. 97-106.

295. Eustáquio, A.S., D. O'Hagan, and B.S. Moore, *Engineering fluorometabolite production: fluorinase expression in Salinispora tropica yields fluorosalinosporamide.* Journal of natural products, 2010. **73**(3): p. 378-382.

296. Huo, L., et al., *Heterologous expression of bacterial natural product biosynthetic pathways.* Nat Prod Rep, 2019. **36**(10): p. 1412-1436.

297. Liu, S.H., et al., *Heterologous Expression of a Cryptic Giant Type I PKS Gene Cluster Leads to the Production of Ansaseomycin.* Org Lett, 2019. **21**(10): p. 3785-3788.

298. League, E., et al., *United in science: high-level synthesis report of latest climate science information convened by the science advisory group of the UN climate action summit 2019.* 2019.

299. Silva, B., *FIA Environmental Strategy 2020-2030.* 2020: p. 1-28.

300. Yuzawa, S., J.D. Keasling, and L. Katz, *Bio-based production of fuels and industrial chemicals by repurposing antibiotic-producing type I modular polyketide synthases: opportunities and challenges.* J Antibiot (Tokyo), 2017. **70**(4): p. 378-385.

301. Yu, Y., et al., *Reassembly of the Biosynthetic Gene Cluster Enables High Epothilone Yield in Engineered Schlegelella brevitalea.* ACS Synth Biol, 2020. **9**(8): p. 2009-2022.

302. Wilson, M.R., et al., *The human gut bacterial genotoxin colibactin alkylates DNA.* Science, 2019. **363**(6428).

303. Secher, T., et al., *Escherichia coli producing colibactin triggers premature and transmissible senescence in mammalian cells.* PLoS One, 2013. **8**(10): p. e77157.

304. Li, Z.R., et al., *Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage.* Nat Chem, 2019. **11**(10): p. 880-889.

305. Duong, M.T., et al., *Bacteria-cancer interactions: bacteria-based cancer therapy.* Exp Mol Med, 2019. **51**(12): p. 1-15.

306. Zhao, M., et al., *Tumor-targeting bacterial therapy with amino acid auxotrophs of GFP-expressing Salmonella typhimurium.* Proc Natl Acad Sci U S A, 2005. **102**(3): p. 755-60.

307. Ittig, S.J., et al., *A bacterial type III secretion-based protein delivery tool for broad applications in cell biology.* J Cell Biol, 2015. **211**(4): p. 913-31.

Appendix

Strain list

ID	Organism	Genotype
Pf NCIMB10586	*Pseudomonas fluorescens* NCIMB 10586	Wild-type
dmupR SM10	*E. coli* SM10 λpir+	dmupR-pEXG2
dmupR infr SM10	*E. coli* SM10 λpir+	dmupR-inframe-pEXG2
dmupR Kl6 Pf	*Pseudomonas fluorescens* NCIMB 10586	mupR gene KO
dmupR infr Kl 1 Pf	*Pseudomonas fluorescens* NCIMB 10586	mupR inframe KO
dmupN pEXG2 Kl2	*E. coli* SM10 λpir+	dmupN-pEXG2
dmupN infr pEXG2 Kl8	E. coli SM10 λpir+	dmupN-inframe-pEXG2
Pf dmupN 1	*Pseudomonas fluorescens* NCIMB 10586	mupN gene KO
mmpA-mSc Kl1	*E. coli* SM10 λpir+	mmpA-mScarlet-I inter acp pEX18GmR
mmpB-mNG Kl1	*E. coli* SM10 λpir+	mmpB-mNeonGreen-pEX18
mupA-xylE Kl1	*E. coli* SM10 λpir+	mupA-xylE-pEX18
Pf mupA-xylE Kl2	*P. fluorescens* NCIMB 10586	mupA fusion to xylE
mmpB 1	*E. coli* SM10 λpir+	mmpB-mNG-pEXG2
mmpA FLAG	*E. coli* SM10 λpir+	mmpA-FLAG-pEXG2
mmpB-mNG 1	*Pseudomonas fluorescens* NCIMB 10586	mmpB-mNeonGreen fusion
mmpD-mNG	*E. coli* SM10 λpir+	mmpD-mNeonGreen-pEXG2
mmpD-mSc	*E. coli* SM10 λpir+	mmpD-mScarlet-I-pEXG2
Pf mmpD-mSc	*Pseudomonas fluorescens* NCIMB 10586	mmpD-mScarlet-I N-terminal
Pf mmpD-mNG	*Pseudomonas fluorescens* NCIMB 10586	mmpD-mNeonGreen N-terminal
Pf mmpB-mNG mmpD-mSc	*Pseudomonas fluorescens* NCIMB 10586	mmpB-mNeonGreen mmpD-mScarlet-I N-terminal
mupR-pPSV35	*E. coli* SM10 λpir+	mupR-pPSV35
Pf mupR	*Pseudomonas fluorescens* NCIMB 10586	mmpB-mNeonGreen mmpD-mScarlet-I N-terminal mupR-pPSV35

mmpC-mNG	*E. coli* SM10 λpir+	mmpC-mNeonGreen-pEXG2
mmpC-mSc	*E. coli* SM10 λpir+	mmpC-mScarlet-I-pEXG2
mmpA_pDonor	*E. coli* NEB10β	mmpA-pDONR221
mmpB_pDonor	*E. coli* NEB10β	mmpB-pDONR221
mmpC_pDonor	*E. coli* NEB10β	mmpC-pDONR221
mmpD_mod34_pDonor	*E. coli* NEB10β	mmpD modules 3-4 -pDONR221
mmpA_pAB1G	*E. coli* NEB10β	mmpA-pAB1G-N-HMF
mmpB_pAB1G	*E. coli* NEB10β	mmpB-pAB1G-N-HMF
mmpC_pAB1G	*E. coli* NEB10β	mmpC-pAB1G-N-HMF
mmpD_mod34_pAB1G	*E. coli* NEB10β	mmpD modules 3-4 -pAB1G-N-HMF
mmpA_pAB1G	*E. coli* EMBacY	mmpA-pAB1G-N-HMF
mmpB_pAB1G	*E. coli* EMBacY	mmpB-pAB1G-N-HMF
mmpC_pAB1G	*E. coli* EMBacY	mmpC-pAB1G-N-HMF
mmpD_mod34_pAB1G	*E. coli* EMBacY	mmpD modules 3-4 -pAB1G-N-HMF

gBlock sequences

mmpA gBlock1:

TAGTGGTTCTGGTGACTACAAGGACGACGACGACAAGCTTACAAGTTTGTACAAAAAAGCAGGCTT
AATGAATATAGAGAGAAAGAAGGCAAGGCTCGCTGCAGCAATAAATGCATCCTTGAAAGCGCAACC
CGCTCCCAGTTTTCCAGAGCCGGTGGCTATAATCGGATTGAGCGCGAATGTCGCGCAATCAGCGTCA
GTTCGTCAATTCTGGCAAGCATTGGACGACGACCGCAGTCTCATAGAGGAGATCCCGGCAACCCGTT
TTGATTTCACAAGTTGGTACGCGGGATCCAATATTGAAGAAGGCAAGATGCGTACACGCTGGGGAG
GATTCATTCCCGCGATCGACCAATTTGACCCCGTTTTCTTCGGAATGTTGCCTGCAGAGGCCAGGAA
AATGGATCCTCAGCAACGCTTGCTGTTGATGAGTGTGAGGCAGACTTTTGAAGATGCTGGCTACCGT
CACACCGATTGGAAGGGCTCAGCGACAGGCGTTTTCATTGCGGCGGAACGCAATGAATACCATCTGA
ATCTTCTTCAAGCCCAGATAGATCCCGGCGAAGGACTCGATCAAGCCGCTAGCATGCTTGCGAACCG
CGTTTCGCACTTTTACGACCTCCGCGGTCCATCGGAAAGGATAGACGCTATGTGCGCGGGAGGTGCA
GTCGCACTTCATCATGCCGTTACAGCGCTCCGTTCTGGTCAAATAAATGCGGCCATAGTGGGCGCCT
GCAACCTGTTGCTTCGTCCCGATGTATTCGTCACCTTGTCGCAATCCGGTCAGATGTCACCAGAGCCC
ACTGTTAGGTCCTTTGGTGCCGGAGCGGACGGTTACTTGAGGGGCGAAGGCGTCTGTTCACTTCTGC
TGAAGCCTCTTTCAAAGGCGGAGGCGGATGGAGACCATATATATGGACTTATTCGCAACACTGCGGT
CAATTACAACGGAGGCGACGCTGCGTCGATTGCCGCACCATCCGTTTCAGCCCATTCAAGCCTCGTG
CAAGACTGTTATAGAAGAGCCGGTATCGATCCTAGACACGTTTCTTACATTGAGGCACAAGGTATGG
GCAACCCTGTAGCTGATATCGCCGAATGGGACGCATTGAACCACGGACTCTTGGCCCTTGGTAGAGA
GCAGGGAGTGCAGCTGCAGGAGGGCCAATGCGCGATATCTACTTTGAAGCCCATGAGTGGACATAT
GCATGCGGCCTCCGCGATAGGAGCCCTGTTCAAAATAATCCGTTCGCTTCAGACCGAAAAAATTCAC
AAAATCCTTGACTTTGAGCAGCCGAACCTTCATCTCCACACCGCGGGACAACCCTGTCGTCTGGCAA
CCCACACTGTTGACTGGCCGAGACAGGCGACCCCCCGCCTTGCTGGCCTGCATTCTTACGGAGCGGG
TGGAAACAATGCGCACATTCTCGTCGAGGAGTACGTCCACCAAGCTCCGGGCCGTGTAGTTCCGTCT
CAAGCGCCTCTTCTTTTCCCGTTGAGTGCACCTACCCCAGCCTTGCTTGCGCTGTTGGCCGCACAAAT
GCATCAAGCACTGATAGATGAGCCTACTCTCTGCCTCGAGAGTGTCTCTGACACTCTTAAATTCGGTC
GTGAGAAATTCGCAGCTGCAGTGGTCATAGTAGCAATGGAACGCCCTGGTCTTTTGGATGTACTTGC
CCAGTTGCACACAGGAGACCTGACGGGCGCGGTCGTATTCGCCAAAGACGTAGCAGCTCCGACAGA
TGCCGCCCTCGACCCCTGGGCGGAACTTTGGCTGGCGGGCGAAAGACCAATGGCGAAGGCGACTAG
GGCGCCTAGGGTTCCCTTGCCGACGACACCTTTTGACACCCAATCCTTTTGGTATGATGCCCCGGCGC
CAATGGCAGCTTATCCTGCGCCCGGCACGAGCAGAACTGGTCGCAGAGCTTGGTCCCCGGATGCTGC
TGAGCAGGCTGTTCGCGAAGCACTTGCGCAAGCTTTGGAGCAACCTGCTGCATCCCTGGATCTTGAC
GCACAGTTTTCCGAGTTGGGCTTCGATTCCATGATGGTCAGGCAGTTGTGCCGCCATATGCGCGATC
AGGATATCGTTGTAGAGCCCGCGGTTTTGTTCGAGCATGCGACGCCTGCAAGGCTCGTTGCGTGGCT
GGCGTGTGCCCCAGCTCAAAGGATTGCGGCGCCGTATCATCTGCCATCACCTGGAGCTAGAGGACAG
ATCGCGATTATAGGTATGGCGGGAGTATATCCAAAAAGCGCTGATCTTTCTGCCTTTTGGGATTGCCT
GGCTAATGCTGGTGACTGCATTGAAACTGTGCCAGAGCAGCGTTGGAGCCTGGATGAGCACTTTAAC
GCTGATCGCCAACGCGCAATACAAGAAGGCAAATCGTACGGAAAGTGGGGCGGCTTCATCGAGGGC
TTGGATGACTTTGACCCGATGTTCTTCAACTTCTCGTTGGCCGAAGCCACGTATATGCATCCCAAGGA
ACGCCAATTCATCCAGTGCGCATGGCATGCGTTGGAGGATGCAGGCTACACGCCTGCAAGCCTCGAA
CAGGAGAAAGTAGGTGTTTTCGTAGGAGTCTCAAAGGCGGGTCATGATAACTACAAAGATAGCTTTT
TCTCTATAGCAAACCGCGTGTCTTATAGGTTCGGCTTCACTGGCCCCAGTCTGCCCGTAGACACCGCA
TGTTCGTCTTCGCTGACGGCTGTCCATGAAGCCTGTCTCCACTTGCAGGCGGGAGAGTGTACAGTAG
CCGTGGTGGGTGGAGTTAATGCTTACACGCATCCGTCGACATTTGCAGAGTTTGCACGCCTCGGAGT
TTTGTCCGCCGATGGAAAATCCAGGGCATTTGGAGCAGGTGCAAATGGATTTGTACCGGGCGAAGG
AGTCGGTGCGCTCCTCTTGAAACCTCTGGAAC

mmpA gBlock 2:

GGGCGAAGGAGTCGGTGCGCTCCTCTTGAAACCTCTGGAACGCGCCTTGGCTGATGGAGATATGGTG
CATGGCGTAATCGCTGCAAGTGCCGTAAACCATGGCGGCAAAGCAAATGGTTATACAGTCCCCAACC
CAGAAGCCCAGAGGGCTATGATCAGGCTTGCCCTGGATAGAGCAGGTGTATCAGCCGACCAAGTAA
CATATGTTGAGGCTCATGGCACAGGCACGGCGCTCGGAGACCCAATAGAATTTAGAGGCCTGGTAG
AGGCCTTCCGTCAGGATACCGAAAGGACGGGCTTTTGCCGCTTGGGCTCCGTGAAATCCAATATTGG
CCACCTTGAGGCAGCTGCGGGAATAGCCGGATTGTCGAAGGTCCTTCTCCAAATGCGCCACGGACAA
ATCGCGCCTTCCCTCCATTCACAACAGATGAATCCTGATATTGATGCCTCCCTGTCCCCTTTCAGGGT
CCCACAGGCTCTGGAACCTTGGGACGTCGACGGCGAAGGCGCACAAGAGCGTATCGCATGTTTGTCC
TCCTTTGGCGCAGGCGGCGCCAATGCGCATGTCATTTTGAGACAGGCCCCGGTTAGAGCTGAGCCAG
CTGAAGACGTACGTGCAGACTATCTCATCGTACTCTCTGCTAGACACGAAGCTGCACTCGTTAGACA
GGTAGATACTCTGTTGGTCTTCCTGGAGAGAAACCCAAGTCTCCCACTCAGGGATATCGCCTACACT
CTGCAGGTGGGAAGGCAAGCAATGAATGCTCGCCTCTCTTGTATAGCCACTTCTACTGCTGATCTTAT
GGACGCGTTGCGTAGATATTGTGCCGGCGAGGCACATCCTGGAGTACAATCCACAACATTGAAGGA
CGCCGATAGGTTGAGCTTGTTTGGACAAGATGAGAGCGCACTGACCCTTCTTGATCAGTGGTACGCT
GAAAGTAAATGGGCCCAACTCGCACAACTGTGGCGTGCCGGAGAGGTGATTGATTGGGCACGCCTT
GTTGCACCCGGAACTGCTAGACGTATCTCACTCCCCGGTTACTCATTTGACAGGGAACCCATACCTA
GAGCACCACAGCTTGTTCCTCTCGGTTCTGTTGCTCGTCCTCCTGCACAGCCTGGAAGGGCGCCTGTA
CCGCCAACCACAGCGGCAACGCTTCTGGTAAGACCTAGATGGCAACCCGTACAAGCGTCTCCGGCGG
CTGCGGCACCAAGACGCTGCCACGTAATTCTCTGTGAATGGCCAGGTGAGAGAAGCGATCACCCTGG
CTGGCAAGTCTTGACACCGAAGGGTCTTCCGACCGAACGTTATGTCCACTGCGTTGGACAGATCGCG
GAACATCTGAGACAAATTATCGTGGACGCCGCGCACGAAGTAATTTCGGTTCAGGTAGTCATTCCGG
TTAACAGAAGCGCTTGGCTTTCGGGAATCCACGGTTTGTTGCGCTCCGCTCAGCTGGAATCAAGTAA
ACTTCGCACTCAATTGGTAGAGATCGATCAGTCGTTTACGTTCACACAAGTCCGCCACGCAGTAGAA
CAAGCTGCAGCCTTGGAGGTCCACCCTTGGCTTAGATTTGATGCACAGGGTGGACATGCGCGCCAAT
GGACGCAAGCTACAGTATTGGACGACACCCATAGGTGTGGTTGGAAGAACGCGGGAGTTTACTTGC
TTACCGGTGGCATGGGTGGTCTCGCTCAGTGCTTTGCCGCGGATATTTTTTCAAGAGTAGAGGGAGC
TACGGTGGTTCTGGTGGGCAGATCCCCGCTCGGTCAAGCCGGACAAGACCGCCTTGATGCGCTCCGC
AGGATGGGTGGTCAAGTGAGCTACGAGCAAGTGGACATCTCTGACGAGGCGGCGACTGAGAGACTC
GTCGCGGACGTACAAGCCCAGTATGGCGCGTTGAATGGCGTGCTGCACTGTGCAGGAACGCTGCATG
ATGAGTTTTTGAGGAACTTCAAACATCAGCATGGCGACGGTGTACTGGCTGCAAAAGTCGGTGGTAC
TTGGTCCATCGACCAAGCTACACGCCATTGCGATCTGGATTGCTTCGTTCTTTTTTCAAGCGCCGCAG
CGGCAGTAGGTCATCCCGGTCAAGCATTCTATGCTGCGGCAAATGGTTTCATGGATGGTTTTGCTGA
GCAAAGGACGCAACAGGTGCAGATGGGAGAGAGGCGTGGCCGTACAGTCAGCATCGCTTGGGGCCT
CTGGCGCGATGGAGGAATGCAACTCGACGCGCAAGCCTCGGCACTTCTCGAAAGGGAGAGTGGAAT
GATACCCCTCTCGAGAGAGCATGGAGTAGCAGCGTTCCATCAGGTCATGGACAGTGATGCCTGTCGT
ATTGTGGTCATCGAAGGAGACGAGCAGAGGTTGAGGAGACGCCTCCATCCTCTTTCGAACGAAGCTA
CTCTCCCAGCAGACACTAGAATGCTGGACGTGGTTAAGCGCTTGGCAGGCCAAGTCCTTGGTATCGC
GACGAGTGAACTGGAGGGAGATCATGACTTCGACAGTTATGGCCTGGATCAAGCAAGGTTTAACCA
ACTTATCCAAGCCCTCGACCGTGAATGCGGTCTGGAAGTTAGTCCTCTGGTCTATCGTCAAGTTCGCT
GTTTGGATGACCTCGTAGCACACCTCCCCGCCCAACTTCATGCCGGTACTGCACCCCGCGCCGTGGCC
GGTACAGCCTCAGCACAGGTGTTGCTTGAGCATGCCCTCCAGGTGCTTAAGCGCGTATTGTCGCCCG
TAGTACAATGGCCTGAGGATCGTCTGGATTCGGACGAACCACTGGAGAGGTATGGCTTGGACTCAAT
GATGGTGATGACCATAACCGCTGCCTTGGAAGCACAATTCGGAGCGCTGCCAAAAACACTCTTTTTC
GAGTATAGTACACTTCGCGCTCTGGCTGCG

mmpA gBlock 3:

CACTCTTTTTCGAGTATAGTACACTTCGCGCTCTGGCTGCGTACTTGTGCAGGGAACACGCGTCCACC
CTCGTCGCACTTGTTGCGGACGTGGCAGGAACACTTGCCTCAGGACCGGATGAGGCTCGTGTTGCCG
GCATTGCGTCACATGATCCGGTTCCTGCTGATGCTGCCCAGTCACCGGCTAGTGCAAAGGAAACTTC
TCAGCGTTTCGACGTAGCGATCATAGGACTCGCTGGACGTTACCCTCAGGCACGTACGCTTGAAGAG
TTCTGGCAAGTCCTTAGCCAGGGCCGCGATTGCATAAGTGAAATACCCACCGAAAGGTGGGATCATA
GCCGTTACTATTCGGCGGATGAGGATGCCCCGGGAAAAACGTATGCCAGGTGGGGTGGTTTCATTGA

TGGTGTCGCAGAGTTCGACCCTGCGTTCTTTGGTATATCTCCTCGTGAGGCAATGGCCATGGAACCTC
AAGAAAGGTTGTTTCTGCAGACCGCCCATGAAGCCATCGAGGACGCCGGATACACAAGACAAGGTT
TGGCTGCCTCTGCAAGACATGAAGATGCAGAGGGTATGGTGGGCGTTTTCGTGGGAGTAACCTACGA
GGAATATCAACTGTACGGAGCTCAGCAGACCGCGGAAGGACGCCCCCTCGTTTTGTCGATGAGCCCC
TCGAGTATCGCGAACAGAGTTAGTTTTGTAAACGGATTCCATGGCCCCTCTATGGCAATTGATGCTA
TGTGCGCGAGCTCCTTGACTACTATTCACCTTGCATGCCAGAGTCTGCATTCTGGCGAGTGCCAGGTT
GCGCTTGCAGGAGGTGTGAACGTCAGTATCCATCCTGCTAAGTTTTTGATGCTCGGCGCAGGAAAGT
TTGCCTCAAGACGTGGTCGTTGCGAATCATTTGGCGCGTCTGGTAGTGGTTACGTGCCGTCCGAGGG
TGTAGGAGCTGTACTTCTGAAACCGTTGCATCAAGCCCAGGCCGACGGAGACCGCATTTATGGTGTT
ATTCGCGGCTCTGCCATCAACCACGGAGGACGCACCAATGGATACGCTGTGCCGAATCCAGCTGCAC
AGGCTGCTGTTATATCTAGGGCACTCAGACAGGCTAACGTACAGCCCAGGCAAATAGGTTACGTAGA
AGCTCATGGCACAGGAACCGTACTGGGCGATCCGATCGAGGTCGCGGGCCTTTCTCGCGCCTGGCGT
GCCTATACTCCGGATAGGCAATTTTGTGCACTCGGTTCGGTAAAGTCAAACATTGGACACTGCGAAT
CAGCAGCGGGTATCGCCGGATTGACGAAGGTCCTTCTGCAGATGAAACACGGACTCTTGGTTTCCAG
TCTCCACGCGGAGACCCTCAATCCAAACATAGATTTCAGTGAGACACCGTTTAGGGTGCAAAGGACG
CTGGCGCCTTGGGAGAGGCCGCTGGGAAGTGACGGAGTCCCGGGTCTGAGGATGGCGGCGATCAGC
AGCTTCGGCGCGGGAGGAGCAAATGCACACCTGATTGTCTCCGAAGCGCCCCTTGCGCAGGCGGCTC
ATGCGGCGATGACGGTTACCGCGCCGCTGCCAATAGTTCTGAGTGCCCGTTCCCGCGAAGCCCTGCT
TGAAGTTGCTCAGCGTCTGGTGGATTGGTGGCTGGACGCGGCTGCGAATCGTGAACCAGACCTTCAA
GCCCTCGCGTACACCTTGCAAGTAGGAAGGGAGGCGTGGGAGTGGCGCGTGGCGCTTCTGGTAACA
TCAGGAGATGAGCTGGTGCGTGAGTTGAGGGCGTTTATCGATGGTGCGCTGGAGGGTCCTAGCTGGT
GGTCGGGTTGTCTGCCGGAAGCGCACAGTCTGGCGACGAGGCCTAGTGAGCAAGCATGTGCCGCCA
TCAGGCAGATGTTCGACCAAGCAGACCTTGGCGGCATACTGCGTAGATGGGTCCAGGGAGAAGCTG
TAGACTGGAGCCCGCTGTATCACTCTCACAGGCCAGCGCGTCTCGGATTGCCAACGTATCCATTCGC
CCGCCAGAGGTATTGGGCGCCCAGCGAAGTTGCGGTGAGTGTTGAATCCGTTGCGTCGGTCGCGCAA
GGTGGTAAGATCCGTCTTAGGGCGCTGGACACACTTCCCGTGGCTGATTCGGCCCCTATCGTCCCGCC
GGCTCCACGTATATTGTCTCAACCAGCGCCATTGGCTGCAAAAGCGGCACCAGTCGTTCCCGTCGCC
GATGACGAATGCGCGCAGTTTCTGAGACAGTCCCTCGCGGCCATGCTGTACTGCGAACCAGGACAGA
TTAGGGACGGTTCGAGATTTTTGGAACTCGGCCTTGATAGTGTAATTGCCGCACAGTGGATACGCGA
AATCAATAAGCATTACCAGCTCAAAATTCCCGCCGATGGAATATATACCTACCCAGTATTTAAAGCG
TTCACACAGTGGGTAGGAACGCAACTCCAACCTACGCAGGCAACTGCGGCTCCAGTGCAAAGGGAG
CCCGTAGCGACAGCACCACAACCTGGAGCACAGGCCAGCGCGCAGCGCGAGTCTATACAAGACTAT
TTGAAGCAATCCTTGGGTGAGCTCCTCTTCCTTGATCCTGGCCAGCTCCGTTCTGGAGCCCAATTCTT
GGATCTCGGCATGGATTCTGTAACAGGAACTCAGTGGATGAGGGGCGTATCAAGACATTTCAGCATT
CAATTGGCGGCTGATGCTATTTACACGTGGCCCACTCTGAAATCCTTGGCAGACGAAGTTGATAGAA
GGGTCCAGGTCGTTCATTCCAGTGACGACCTCCTGGATACGATACTTAACCAAGTACAATCTGGAGC
GCTCGCGGCAGATTCGGCGTGCGCCGTTGTTGAGGATGTTCTTGCACAACGTTAGAACCCAGCTTTCT
TGTACAAAGTGGTTCTAGAGCCTGCAGTCTCGACAAGCTTGTCGAGAAGTACT

mmpB gBlock1:

TAGTGGTTCTGGTGACTACAAGGACGACGACGACAAGCTTACAAGTTTGTACAAAAAAGCAGGCTT
AATGGTGAAGGATTTCGATTCTAACGAAAACGCCGGAAATGCCTCCGGCAGTGACACGCATCCTCAT
GCGTTTGATATAGCCATTATTGGACTCGCAGGTAGATATCCTCAAGCAGAAAATATTGAGGAGCTTT
GGGAAAATCTTAAGCTGGGACGTGATTGCATCACAACTGTTCCAAGTCAGCGTTGGGATCATGATGC
CATCTATGATCCTAGCAAGGGCGTCAGTGGCAAAACTTACAGTAAATGGGGTGGATTCCTTAGGGGC
GTGGATGAATTTGACCCTCGCTTTTTCAATATATCCCCGCGCGAGGCTGAAATAATGGATCCCCAGG
AGCGTCTGTTTCTTCAATGTGCTTATCATGTTTTGGAGGACGCTGGTTACACTCGTCAGTCTCTGAGT
GCCAAAGGTCGTGTAGGCGTCTACGTGGGAGTACAGTATACAGAGTACACAGCGTTCTCCACGGCCC
AGACGCTTGTGGCCGCACTGCCGGCCAGTATCGCCAATCGTGTCAGTTTTTTTTGCGACTTTAGAGGC
CCGTCTTTGACGCTTGACAGCATGTGCAGTTCGTCTCTCACCACGATCCATCTGGCTTGTCAGTCCCT
CAGGAGCGGAGAAAGCGAGTATGCAATAGCAGGTGGTGTTAATGTATCTATTCACCCCAACAAATA
TTTGCTGCATGCACAGGGCCGTTTCGCGTCAAGTGTGGGCTGTTGTAAAACGTTTGGCCAGGGAGGA

GACGGTTACGTCCCTGCAGAAGGCGTTGGAGCAGTGTTGCTTAAGCCACTTCCGCAGGCAATAGCAG
ACGGCGATCGTATTCATGCTGTTATTAAAGGATCGAGTATTAACCACGGCGGCAGAGCCACGGGATT
TACAGTGCCGAAGAGCTCGGCACAAGCCGTGGCAGTAAGATCTGCGCTGCATCAGGCTGGTCTGGCT
AGTAGCGATCTGTCGTACATAGAGGCCCATGGCACGGGCACCGCACTCGGAGATCTGATAGAAATA
GAAGGATTGAGAACGGTCTTTGAGGCGGATGGCTTCGAGCCGCAAACCTGTGCTATCGGTAGTATCA
AGAGTAACATAGGCCATTGCGAATCCGCCGCAGGAATCGCGGGCCTCACAAAAGTCCTGCTTCAACT
CAAGCATAGGCAATTGGTTCCGTCACTCCACTCGACTCAACTCAACAGAAACATTGACTTCGCTGGA
TCTCCATTCCATGTTCAACAGACGCTCGAGCCATGGCTTCCGAAAGGAGCGGCGGACGCAGTCCAAC
CTCGCAGAGCGGGCATATCTTCATTTGGCGCCGGCGGCAGCAATGCGCATCTGGTGGTTGAAGAATA
TTCACTCCAGGCGCAGGCGCACACCCACGTCCCAACTTTGCCTTCAGACGTTCTCATAGTCCTGTCGG
CACGCACACACGAAAGGCTGCAGGCGGTGGCCGCACTTCTCCTGGGTCGTCTTGAGCAGACGCCGCT
TGCGTGTGATCTCGCTGGACTGCATGACTTGGCATATACGCTGCAAGTGGGTCGTGAAGCATTGGAT
GCTCGCGCTGCTTTCACCGCTCAATCGGTGCAAGTTTTGAAGGAGAGACTTGTAGCACTCGCTGATG
GTGCACAGCATCCCGATGTCCTGATAGGTCAGGCTTTGAAGCCTGTACGCCTTAGAGCCGGAGAGAC
CGCCGTTCCCGTACAAAGCCCTGCTATGGATGACAACCAAGCGGTCATGCGCCATTGGGTCTCTGGC
GGCCAAGTAGAGTGGGCACAATTGTATCAAGGATTCCAACCCGCTAGGATATCGTTGCCCCTTTACC
CCTTCGTTCGTGAACGTTGCTGGGTGCCTCCGGCGCCTCAAACCGCTGTGTTGGGCAGATTCGCCGCC
CCTGGAGACGGACCACATCCATTGCTGTCTCGCCAGATTACAGCGGGAGACTCTCAACAATTTGAAG
TACAGTTCTGTGGCGATGAGTTTTTCCTGGACCATCATCGCGTTAACGGAAGGAAGGTTTTGCCAGG
AGTCGTTTACCTCGAAATGGCTCTCGCCGCGCTCTCGAGGGTCTCGGGCCAAGCCGACCTTGCTCTGA
ATTCCATTCTTTGGGTTCGTCCCTTTGAATGCTTCAACGACCCTGTCACCCTGACCTTGGC

mmpB gBlock2:

CCTTTGAATGCTTCAACGACCCTGTCACCCTGACCTTGGCAGTGGAGCAGGCGCGTAAAGGCTCTGG
TTGGGAATTTGAGTTTTTCGCGTGGCTTACACGCGAACCGGGCGGCGAACCCGTTCGCCAAGTCTAC
TGTCGCGGCCAAGCGTCCGTGCCGGCAGAACCTGCACCCCTTGCGCATCTCGATGTTAACGCCTTGG
CCGCACTGGCACAGGGTGCAGCTTCCACGACGGAGTTTTACAGACACTTCGCAGCAGTAGGTATAGA
ATACGGTCCCGGACACCAGGGCCTTCACGCGCTCCGTGGAGGCGCCGAGCAAGTCCTCGGCCACGTT
AGGCTGCCAGACTTCCTCAAACCCGGCGACGAATTTAGGCTCCATCCCTGCTTGCTCGACTCGGCACT
CCAGGCAGCAATGGGCTTTACAACAGGTAGTGAATCTAGCCAGGGTCCCCAGGTGCCATTCTCGCTC
GAGCAACTTACATTGGGTAGATCCCGTAGAACCCCGGCCTGGGTCCGTCTGACAGCGCATACATCGG
GTGGAGGTCGCGGTTCTTTCGCGTTCGACGTTGAAGTCTTCGACGCTCAGGGCGAACCTTGTGTACA
CCTCCATGGCCTGCGTTCGAGGGTCATGGCACGCGCCCCTGCCCCCCCGAGCCTTGCTCTCATGACTC
CCGGCCTCGAAGCTTTTGTCGAACCCGCTTGGGCAGGACTCAGTCACGCGCACCGCATATATCTGTG
TCCAGGTAGTGAAGCATATGCTGCACAGGTCGCTGAGCATTTTCCCCAAGCCCAGTGCGTCCTGCTC
GAATCTGGAGACGATCAGGGTCTGACCCTTCCTGAAAGATATTGCAGAGCGGCTCAGGCATTGAGCG
AACATCTTAGACAACTCCCTCATGACGGACAAGGTTGTCTCGCTCAGTTGGTGGTTGCCGACGACGA
CAGCTCCGGCCTCTTTGCTGGACTGTTTGGCCTCCTTAGGAGCTTCGCAAGGGAACAGACTTCTTTGC
GCGTCCAGTTGTTGCATCTGCAGCCCGGTGACGCATTGCCACCTCTTCTCCAATCTGCGTGCACGGCC
GATGCTGAGCTCGCATTGCACAGAAGGGATGGCCAACTGTTCAGCGCCTCTTGGCAGCGTCGTCCAG
ACGAACAACTCCCGGCAGAACCTGGCGCCGCACCTTGGAAGCCAACCGGCGTCTATCTCCTGACGGG
AGCGTCCGGCGCTCTGGCTCGTAGTATTGCTCTTGATATTGCACGTGCGGCAGAAGGAGTCAGCCTG
GTGCTTGTAGGACGTGGTGTGCCCGATGACGACATGGCGGCCCATATTGAACAACTCGAATCCCTTG
GCGCACAGGTTCTCCATCGTACAGCCGATCTTGCGGATGCCACTGCCCTTAGGGCACTCGTCCAAGA
GGTTTCAGCGCGTGTGGGTTCCCTTACAGGTGTGCTCCACATTGCCGGATGTCGTCAGGACGGACTTC
TGCGCAACAAGAGCGCTGAACATTTGCGCCAAGTGTTGCTGCCAAAGGTAGACGGTTTGTGGAATCT
TGACTGCGCGACTCAGGACCAGCCCTTGGAATTTTTCCTCGCTTTCGCTTCGATCGCAGGCGCATTGG
GTAATGCCGGTCAGAGCGATTATGCTTGTGCCAACGCTTTTATGGATCAATTTGCAGAGTACAGAAA
CAGACAGGTGGCGGACGGTACTAGACAGGGTCACACCATCGCTATAGATTGGCCGCTTTGGGAAGA
CGGAGGTATGCAACTCTCACCGGAACATATGGGCGCGATGACAAAAGCTACTGGCCTTCGTCCTCTG
GCAACAACTCAAGGCCTTCAAGTGCTCCGTCACATTATTCACCAAGGCGGAGCGAGATGGCTTGTGT
TGCCTGGCCAGGGTGCAACATTCGAAGCACTTGTTGCCGGCTTGGGAGTAGGTGTCGCCGCGGCCGT

TGCTCCTAGGTCTCAAGTCGATCCAAGTGACACTCTTCGTGGTCTCGTTGGACAAATCTTGAAAGTTG
ATGCACAAGAGATCGATGATACCACTGCCTTCTCAGACATGGGTTTCGATTCGGTTATGCTGACTGA
GCTCGCCACAGCAATCAACAGGACGTATACGCTCGAGCTGGGAACAGCCGCGCTTTTCGAACACCCA
ACTCTGCAGGCCCTCGCAGCTCATTTGCAAGGCGCACGTACAGCAGAGTCT

mmpB gBlock 3:

CCTCGCAGCTCATTTGCAAGGCGCACGTACAGCAGAGTCTCAACCACCAGCACCAGGACTTACTAGA
GCCCAAGTTGCCCAAGGTGTAAGGGAAGTTGTGGCGGAAGCTTTGAAAGTCAGGCTTGAGGATATC
GGAGATGACGATCCCTGGTCGGATTACGGCATGGATTCTGTTAGCAGCGTACAGATGACTGGCCTTC
TGAACGAACGCTTTGATATCCAACTTGCGGCGGACACGTTCCAAGCTTTTGGCAATGTGGTCGAATT
GACCACGGCTATTGCTGACATTCAAGTCGTCATGGCCTGTGCGGATATTGAAAGACAGCCAAGCGCT
GGTGGCGGTGTCATTGCCGACACTGCCTTGTTGGATGAATTGGTAGCCCTGGTTTGTCAGCTCTTGAA
AACGGTAGCCGGAGATATCGATCCGCACACTGATTTGCACGATTTTGGCTTCGACTCCGTGCTCTTGA
CCCAGCTCTTGGCACAAATATCCTCCACATACGGAGTGGAGCTCGATCCCGGTTCGGTTCTCGAGGA
TGCGACTGTAGCCGGTCTCGTAGCCCAAGTTCAGGCGCAGCGCCACGGCGCTGAACCTGCCAGCCTG
AGAGTCCCCCAGGTCCAGGAACGTCCAGCTCCGCAGCGCGTGGTAACCCTTGCAGAGCCCGAACAG
AGGGCAGACCGTCAATCGCCGGTTGTCGTATTGCTTAGTGCCGGTTCGGCGGCACAACTCAGGCAAG
TAGCTGACAACCTCCTTCACCTCATCGATGGCCAAGCCGATCACGACGTCGATCTCCACACGCTGGC
AAGAGTCACACAGGCGGGCAGTGGCCTCATGCCGGTAAGGCTTGGATTGTCCGTCGTAAGTTTGGAG
GCTCTTGCCGAACAACTCAGGGCTTATCTGCAAACTGATGAACATTCAGCTTATGCCCAGGTGGGCA
ACGCAGAGGTCGATGCTAGCGTTCTTAACCGTGTGATGGCACTTCCCGGTCTCGTTGACCAGGTACA
AGGCTGGCTTGATCAAGGTGGTACTCTGCCCCTCCTCCAACTCTGGAGTCAAGGACTCGACGTTGAC
TGGAGAAGCCTTGCGGGAGCTGATAGTCCCGCAAGTGCACTTAGTCGCATGCCCGCGTCGCCGTTGA
GGGAGTCCGCGCCGGTTGTAACTCTGTTGAACCACGCTGGTGAGCAAGGTCTCTTCTGCTTTCCTGCG
GCGGGAGCTGGTGCCCAGTCGTTCCATGCCTTGGCAGCTGCACTTGGAGATGTGGCCCAGCTGCAGG
TTCTTGAGTACCCGCCTTTGGATACACAGGCACCCCCTGACTTGTCAATGCAGCGTATGGTAGCGGC
CCATGTGCACACTGTCACACGCCGCTGTCCAGCAGGTGTCGTTCGTCTGCTTGGACATTCGTTTGGAG
CGTCAGTTGCCTTTGAAGTGGCTGTGGAGTTGCAGGCGAGGGGAAGACAGGTGCAGCTTTTTATGTT
GGACAGCTTCTTCTACGTGCCTCAAGAACTTCAAGGCGAAGTTCTGACGAGTGCCCAATTGCGTTAC
TTCTTGGGTGACACCCCACTTTGTGCACGTCTGCTGGATGCGCAAGACAAAGGTGCGGTTGTATCTCT
GAAACAGGTGCGTGAGTTGATGCGTGATGCCAGTCTGAACCAAGGACCGGCCCAGGGTCCCGGCGC
TTGGCTGGCTGCGCAGATTGCAATGTCAGTTGATTACAGACCGACCCATACATTCACCGGAGACACA
TGTGTTATTATTGCTGCTCAGTCTATTCTGGGAGCCTCGTCGCTGCCCCACACCTTGGCTAGATACTG
TGCGTACCTTACAGCGTCACCCCACGTTGCGACGGTGGACGGTGGCCACATGTCTATGCTGAATCCA
GGTAACGTCCATACCTTGGCTCGCGCTCTTAGGCTGCATTTGGACACTCCAGGAGGTAAAAGCCAAG
GTAGGAGAGCTGGAGAATTCTGAAACCCAGCTTTCTTGTACAAAGTGGTTCTAGAGCCTGC

mmpD_mod12 gBlock 1:

GTGGTTCTGGTGACTACAAGGACGACGACGACAAGCTTACAAGTTTGTACAAAAAAGCAGGCTTAA
TGAAGAACAGGTTGTCGAGTCTCATTAGGAGACTGGCTGGCCATACCCATACCGAGGGAGGTGTAG
GCGACGCGGCTGAGGCGGGTAAGCATCGCCATCCGGGTGACATCGCGATAATCGGTATGGCGTGTC
GCGTCCCTGGTGCAGGCGACTACCAACAGTTTTGGGAGAATCTTGCCAATGCGCGTGACTCCGTCAC
CGAGATTCCCGCGTCACGTTGGGATTGGCGTGCGTTCTGGGGAGACCCTACAAGTGAAGATGACCGT
TCTCATGCAAGGCACGCCGGCTGCATTGATGCTGTGGATGCATTCGATGCAAGGTTTTTTGGCATAG
CCCCAGCCTCAGCAAGGATTATGGACCCGCAACAGCGTATTATGTTGGAACTCTCTTGGGCATGCCT
TGAGGATGCCGGAGTAGTGCCTTCAGCGTTGGCAGGCTCGAGAACGGGAGTTTTTGTAGCAGCCTTT
AATTATGATTACAAACAACTCCTCGAAAGCGCCGGACTTCCAATCGATGCACACCACTCCACGGGCA
ACGCTGCCGCTGTGATAGCGAACAGAATTAGTCATTTCTATGATCTCCACGGTCCTTCAGTGCTCGTG
GACACAGCTTGCTCAGGTTCTCTCTCCGCAATTCACCATGCTGTACAAAGTCTCCGCTTGGGAGAAA
CCGAACTCGCGCTGGCTGGAGGAGTCAACCTGCTTCTGACTCCGACCCGTCATATCGCTTTTGCTAAG
ACAGGTATGCTCTCACCAACCGGAGCGTGCAAGTCATTTGACGAAGCTGCTGATGGTTACGTCCGTT
CGGAGGGCGCAGGCCTCTTGCTCTTGAAGCCGCTGGCAAAAGCGCTTGCCGACGGCGATCCTATCCA

TGGAGTCATCAAAGGATCGGCAGTCAATCACTGTGGTAAGACGCATACCTTGACCTACCCGTCGAGT
GCAGCCCAAGCTCAGGTGATAGAACAAGCGTTGGGTGATGCCCACATACCGGCCTCCTCCGTATCGT
ATATTGAAGCTCATGGTACAGGCACACCCAAGGGAGATCCGATTGAAATTCAGGGCCTTCGCCAAGC
ATTCTGCCCCCAAGCGAGCGAACCGCAAACGTGTGGCGTCGGCTCCGCAAAAAGCAACATCGGTCA
CCTTGAAGCAGCTGCTGGAGTTGCAGGAGTTATCAAAGTTCTGCTGTCATTTAAGGCAGGCCAGCTG
GCGCCCCTCTGCCACTTTAGTGCACTTTCATCTAGGATCGATCTGCAAGGATCCCCTTTTTATCCTGT
GACTAGACTGCAGCCTTGGGAGCCCGCATCGGGAGTACGTAGGGCGGGCGTGTCGAGCTTTGGCTTT
GGAGGAACAAACGGTCACTTGATCCTCGAGCAACCTCCGGCATGCGCTGAAGTGGATGTGGCAGCT
CGCCCGGGATGGCTGATAGCTCTGTCAGCCAAAACGCCGGCTTCGTTGAGAAGACAATGCTCAGCAC
TTCACGATTGGCTTCAAGGTGCTGGAGCCTCGGCGTCTCTTGCACAGCTCTCGCAGGCTTTGTTGACA
GGTAGAGAGCACCTGGCATGCCGCCAGGCCTGGGTAGTATCGGATAGAGACACGCTCATTGCTCAA
CTGTTGGCTGCTTCTCAGGCGGAGGTCCAGGCCCCAGAAATGGATGAGGAGGGAGCAGCTGTACGC
CAACTTGATGCTGCACGCTTGATTAAGGCTCTCCCCACTTTGGCTCACGAGCCATCTGCGTACCTTGC
CCACCTGCAAACCCTTGCAGAACATTATTGCTGTGCTGTGCCCGTGAAATTTTCAGTACTTTTCAGCG
ACCAGACGCGCCATCATCTTGGACTTCCCAGTTACCCCTTCGCGCGTGATAGATTCTGGCTCCCTACA
GCACCGACGGACCTTCTTGAGTACAGGCTGCAGAGACATGACCCAAGTATCGGAGACCACGTCATTG
CGGATCAGCATCGCGTAGCTGGCGCGGTAACTTTGGCTGTCGCTGCCAGGGCGTGGGGTGCGCTGCC
CGCCGGTGGATGGTGTGCCTCACAGGTGGCTTGGCTGCAAGCGCTGGACGTCCCGCTCCAGGGTGCC
CTTCTGCGCGTTAGTCTTCGCCCCGCCTCAGCAGGCGCGACAGATTTTGAGGTGCGCTTGGAGGGCCT
TGTGCACCCACTTTGTTCTGGTACTCTTGCTCCGGGCCAAACGGACGAAGTAAGAAGAGATGTACAT
GCACTGCGCAGGGCACATCCTGATGCCTGGCTGACTCGTGAGGTGCTTTATGAGCAACTGAGAAACG
TAGGTATGGCGTATGGATCAACCTATAGGGTGTTGCAAGACTCACATCTTCA

mmpD_mod12 gBlock 2:

CCTATAGGGTGTTGCAAGACTCACATCTTCAAGGTGATCAAAGTCTCTTGGCACTGGTCGTGGTACC
CGCGGGAGCGGCCGACGACCTTCCATTTTCACCCGCCGCCGTCGACGCAGCATTTCAGGCTGCTGTC
CTTTTGGCCATCAGGCAGTCTGGAGAGCAGGCTCGCCTTGGTTTTAGCGTGGGCCAACTCCAATTGTC
AGCTCCCCTTCCCGAGAGATATTGGGTACACGTTAGTGCTGCGACTACCGGCGCTCAGCAGTGTCGC
TTCAACATAGAATGGATAGCAGAGGATGGTCGTGTGTGCGCAACGGCTAAGGACTTTGTTCTCAAAG
CCGCCCCAGTCCAACAGGAGCAGTCGCCCAAAGCGAAGGTTCACTCCGTCGAATGGAGAGTTCTCCA
TCCGGTAGAAGAGGGCCATGCGGAAGCCCAGCGTAGGACTCTCATAATCTCTGCTGACCCCGCGCAA
GCCCAGGCCTGGAGCGCAGGATCAACGCAGGCTTCGGTTTTGGCGGGTCAATCTGATTGGAGCGCGG
AGCAATGGAGGGATGCGCTTGGCAACGCGGGAGCCTTCGAAAGACTTATTTGGCTTGCACCCATACC
GTCAGGAGAGGCTTGGTCAGCGCAGGCGGCCACAGCGGCGCAAGAGCACGGAGTAATGGCTCTGTT
CGCTCTGTGCCAGGGCCTTTTGATGTTGAAAGTCAGGGAGCACCCTTTGGAACTTCTGGTCGTAACC
AGGCAAGCGCTGCAGGTGCTTCCGGGAGAAAGACATGATCCTACGCATGCGCCTGTTCATGGACTTG
CAGGAACTCTCGCGAACGAGTGTCCGACATGGCAGGTGAGAGTCGTTGATTTGCCCTTGGCACCAAT
GGCGGCCTCGCCGAGTCTTGTGCCGCTGGCCGCGCTGCCTGTTGGAGCTTCAGGAGCCTGGAGGGAG
GGCCAGTGGTACTGCCAGGTGTTGCGTGAGGTGAATTATTGTGCTGCCCCCCCAGTATATAGGAAAG
GTGGCTGCTATATAATTATAGGAGGCGCGGGTGGCATAGGAGTTGCATGGACGAGAAGACTGATTG
TAGACTACCAGGCTCAGGTCTATTGGATAGGCAGGAGGCCACTTGACGAGAATATACGCGCGGACA
TTGAGAGGCTGGCATCGGAGGGTCCCGCCCCAATTTATCTGCCAGCGGACGCAGGCGATCCGCAGGC
GCTGCGTGAGGCACTGGACTGGGTACATGGACAACATGCGCAGGTACACGGCGTTATACATGCGGC
TATGGTAATGAGAGGCAAGTCCGTACAGCGCATGAGCAGGGACGAATTCTCCCAAGTTTTGGCGGC
GAAGGTCGCGACGGGCGCTGCTCTCGAGCAAGCGCTTCAACGTCCTCCCCTGGACTTTCTTGTTTTCT
TTAGTTCTATAAATGCCTTTGAGAGGAGCCCCAACCAAAGCAATTATGCGGCGGGATGTTGCTTCAT
CGATGAGCTCGGTGGCTATCTGGCTCAACACCTTCCGTGTGCCGTAGCCGTTATAGATTGGGGATATT
GGGCAGATGCCGGCGCCTTGGCTGGAAGTCCCGCATTGCAGGTCCTCAGGGAAAAAACCGGCGCTG
GACTCATAGGTCTGTCAGCTGCGATGGAGGTCGTGGATCTTGCTCTCTCCAGACCCTGGCTCCGTGTC
GCCGGAACTGGACGCGAAGCGCGTACACAGTTGGCACTCGAGTGCCCCCGTCAGCTTATCTCGGTGA
TGCCATGTGAGACACAACGTATGGCGAATATGCTGGCACTCTTTGCAACAGCTGTTGAGGGACGTGG
TTCCCTTCAACTTCGCGGCGGCCTCATGCTCGAGCCACTTCAGACTCTCCTGTGCCAAGTGCTTCTGG

TCCAACTTCAAGCCGCCGGACTCTTTGACCAGCCCACTCCCCAAACGACACAGGCCCTGTACCAACG
CCTTCAAGAGCGCTCATCTGGTGCTGTGGGCGTAGACCCTAGGTGGCTGGCCCATTCACTTCGTTGGT
TGGCGGACGAAGGCTGGCTCGCGACAGATGGAGATACCTGGCGTGTACAGAGGGTCGTAGCTGCGG
AATCGGCCTGGGCGGCCTGGGATTCGGCACTGCGCGAGGCCGAGGGAGACGCGGATCGCAGGGCTC
AGGCACAACTTGTAGACGCCTGTTTGCGCGCTTTGCCCGCTGTCCTCTCTGGTCATCTGAGGGCTACT
GAAGTTATGTTTCCCGAAGGCAGCATGGCTCGCGTCGAAGGTATCTACAAACACAACGCTGT

mmpD_mod12 gBlock 3:

CAGCATGGCTCGCGTCGAAGGTATCTACAAACACAACGCTGTTTCTGATTACTTTAATGAAGTCATT
ACACAAGCCGTCCTGGCGCGCGTCAGAGCTGGACACCAGGCCGGCGCGGCGCCCGTGCGTATACTG
GAAATAGGTGCAGGCACGGGCGGAACATCCGCGAGAGTCCTCCAGGGTTTGGACTCGCTGGGTGTG
GAGGTGGGCCAGTACGCTTACACCGATCTCTCACGCGCGTTTCTCATGCATGCCCAAACGGCCTATG
GTGAGGGAAGGTCGTATCTTGATTACAGGATCCTTGACGTAGAGAAGGACCTCGTTCAACAGGGACT
GGAATTGGGCTCGTACGATGTCGTGATCGCAACAAATGTGCTGCACGCTACTGCCGACATAAGACAG
GTAGTGGAGCACGCTAAGGCCGCTTTGAAGACTGACGGATGGCTCCTTTTGAATGAATTGTCGCAGG
CATCCCTTTTCTCGCACCTCACCTTCGGACTCTTGGAAGGTTGGTGGAGATACAGGGACGAAGCACT
CAGGCTGCCCGGAACTCCGGGTCTGTCGTCTGAACGCTGGCAGCAAGTGCTGGAAGAAGCAGGATA
TCGTGGCGTGATGTTTCCTTGTCAAGCGTTCCACGGCCTGGGACAACAAGTAGTCATGGCAAGATCT
GATGGAGTCGTCAGACACGCACGTGAGCCGCAAGCGACCCCGGCACCTGCTCCTTCAGCTGAGAGTC
ACGCACCCGTATCTGCACCTGCTCGCCGTGCCGAAGATGGCAGGCCCCTCCTCCTGGCAAGTGTCAG
GGAGCTTGTTGCTAGTTCTCTGCAGTTCGAACTGTGTGATATCAGAGATGACGAGCCCTTCGCAGAT
TACGGCGTGGACTCAATAACTGGTGTTGCTCTTGTTCGTCAGCTTAATGCCAGACTCGCGATAGATCT
CCATACCACTTGTCTCTTTGATTACCCGAGCATTAATCGCCTTGTCGATTACATGCTCGGTCAGTATC
CTCAGTTTGCTACACCAGCGACGGAACCAACTGACGCCCTGCCTGTCACGCCTCCTACAGGAGAGCC
CGCTCAGCCAGCGGTAGACGACCAAGCGATCGCTATAGTCGGAATTTCTGGTCGCTTCCCACAGTCT
CGCGACCTCGATGCGTTCTGGGAGCACCTTGCAGCAGGCGCAGATTTGGTAAGTACAGTACAAAGGT
GGGACCCGCGCCACTGTTATGCCGGAGCGGACCAAGGTTGCGATCGTGGTGGTTTCATTGAACGTAT
CGATGAGTTTGATCCCTTGTTCTTTAACATAAGCCCTCTGGAAGCCACATATATGGACCCGCAACAG
AGACTCTTTCTTCAAGAGGCATGGAATGCCCTGGCTGACGCTGGCTGTCTCGGAGAACAGCGCAGGG
CGGAGGTTTCTGTATACGTTGGCTGCGAGCAGGGCGATTACGACCAATTGTTTGACGACATGCCGCC
TCCGCAATCTTTCTGGGGAAACGCCCCGTCCATAGTTCCCGCCCGCATTGCTTATTATCTTGATTTGC
AAGGCCCGGCAATAACCGTTGATACGGCTTGTTCGAGTGCGCTCGTGGCGATTCACATGGCATGCCA
ATCGTTGCTGGCTGGCGACACTAAGGTGGCCCTCGCAGGTGGTGTGTTTATTCAATCCACCCCAGCA
TTCTATCAAAAGGCTAACAGGGCGTCCATGCTCTCTCTGACGGGAGCGTGCCATACTTTTGATAGTC
GCGCTGACGGATTTGTTCCAGGAGAGGGTGCTGGTGCCGTTGTTCTGAAACGTCTCGCCGACGCTCT
GGCAGACCGCGACCACATTCATGGACTGATAAGGGGTTCTGGAATTAATCAGGACGGTGCAAGCAA
CGGAATTACCGCCCCTAGTGCTCTGTCTCAGCAGCGTCTTCACGAGTCTGTGTACGCGCGCTACGGCA
TTGACCCTGCCACTGTTCAGCTGGTGGAAGCGCACGGCACAGGTACCCGCCTTGGCGATCCCATTGA
GTTCCAAGCTTTGAAAGGAGCATTTAGTCGTTACACCCAACGTAAAGGTTATTGCGCTATCGGCTCC
GTCAAATCAAACATGGGTCACGCGGCTACCGCTGCCGCGATGGCCGGTCTGTTCAAAGTACTGCTGG
CCATGCGCCACAGACAACTTCCGCCTTCCCTCCACTTCCAGCATACAAATCCAGCCATAGATTTCGCT
GATAGCCCCTTTTTTCTTAATCAACATCTTGGTGACTGGCCGGCGATCGATGGTGCCCCGCGTAGGGC
AGCGATCTCCTCATTTGGATTCAGCGGTACCAACGCACATATGGTACTGGAACAGCCTCCGCTCCGT
GTTCATCAAGCGTCTCTTGTACCCGCCGTACTTATTGTCTTGGATGCCCGCACGCGTGAACAGCTCCA
TCAGCGTATGGCAGATTTGTTGGCGCTTTGCGAACGCGACGCGTCATTCGCAGTCGTAGATATGTCG
TACAGCCTGCTCACAGGACGCCAAAGATTCGAGCAGCGCTTTCATGTCGTAGTAGCGGACAGGGCCG
AGCTGATTGCCGCGCTCAGACGTGGTATGCCTGCCGATACAGCAGACCTGGCAGAAGCTCATCAACG
TAGATTGCAAGGACTGTCAGCGAGGGCCACCGGACAATCCCTTTACTACCTTCAGCAGTTGGCGGTC
TTCGCTGATGCGTGGGCGGCTGGCCAAGCGGTAGACTTCGCGCCGTT

mmpD_mod12 gBlock 4:

GATGCGTGGGCGGCTGGCCAAGCGGTAGACTTCGCGCCGTTGTTCGATGCCTTGGATTGCCAAAGGA
TGCCACTTCCAGGATACCCTTTTGCTCGCGAGAGGTATTGGGTGGTGGAAGCGATGCCTGCGCAGGA
TAAGGCTACTCATCCACTCAGCGCAGCGATGATGCCCAATGATGGCGCAGCCATGGTACTTAGACTC
GAGGGCGATGAATACTTTCTTGATCAACACCGCATAAATGGCGAACCCGTCGTTCCTGGAGTCGCAT
TTCTGGAGGTGCTGGCCCTGGCCCGTGCCGCTCATACCCCACTTACCCCGCTCAGTCTCATTAATCTT
GTGTGGCAGGCTCCTCTCCGTGTTGTATCGGAACCAGTTGAAGTCAGAGTCACTTTGGATGTGCAAT
CGCGCCATACAGGCCTTGGAGGTGTGGAGGTGCACTCGGGACATCCCGCTCAAGTGCATTGCGCGGG
CCACTACGCCGCCCCAGCCACCGCGCCTGCGGCACTGGATCTGGCAGCGCTGCGTGAAGCGTGTAGA
ACAGCACCCCATTCCGCCCATGATTGCTACCGTCATTTTACTGGATTGGGTGTAGACTACGGAGTCA
ACTTTCGTTGCTTGGTAGAGTATGCCGCGTCGGCAGATATTGACCATCCTCAACAGCTCCTCGTGAGG
CTTGCCCGTAATCCAGGAGCGACGCCCAATACCGGCTGGTTGCTTGAACCGGGACTTACGGACTCGG
CTCTTCAGGCCGCTGCACACTGGTTGGCGCGCGATTGGCCAGAGCCGCGCGCATTGGTACCATTTTCT
CTCGAACGCCTGGATGTCTTTTCAGCTGCAGAACCTGTGTGGGCTTATGCCGTTCGCAGGCAGGCTG
CAGGCTGGGCCTCGGTGCCTGCCGTCGACATCCAACTGGCGGATGCAAACGGCACCATCTGCGCTCA
GCTCCAGGGTCTCTCGTTCCGCGCCCTTCCTCCTCCGCAACATCCAGGCGGTACTTCACTGGATGCGG
GAGAAAGCACGTTGTGGACAGGAGACTGGCAACCTAGGGAGGCAACGATGGCTCTGCCGGACTCAG
TAGGTGAGAGACTTCTGATGGTAGCAGCTGCACCACAAAGACTGCTGATGGAGCAAGCTCTTCACGG
TTTGCCTTGGCATTCTCTGGTCGAAAGTGAATGGCCGGTGGGAGAGGGCTATACGCAGGCCGCCATA
AGCTTGATAACACACTTGCAGGCTCACTTGCGCAAAGCCGATGGCCGTCCCCTTCACGTGCAACTGC
TCGTAGAGCCCCATGGCCCTTGGGCGCTGTTTGAAGGATTGGCGGCAATACTTGAAACAGCCGCACT
CGAGAACCCCGCGTTGACAGTGCAGTGCCTTCTGGTCGAGATTCAAGCTCCATCCGCTTCACTCGTG
AACGTAGTAAACCAAGCTCTCGCCCTGGGAGATCCCAGGCTTCGTTGGTGCGATGGACAATGGCAAT
CATTCGTATGGAGGCAACTGCCCTCCGGTCCTGGTGTAGTTCCGTGGAAAGCGACGGGTGTATACTT
GATAACAGGTGGAGCGGGCGGCATGGGCCAAGTTCTGGCAGAGGAAATTTTTCGTCATGCACCAGG
CGCGACCTTGGTACTTGCAACGCACCGCCCTACTACACTGGAGGGATTGGCCCATTTGCACCCAGCA
TTCAACCGCCTGCGTGAGAGGGGTGCTAGAGTGGAGGTAGTAACTCTTGAAGTAAGTGACCGCCTCG
CGACGCAGGCTTTGGTGGAAAGGATCCTCGAACGCTATGGTCGCTTGGACGGCGTGGTGCATAATGC
AGGAGTTACGCGTGATGCTTATCTCCTCCAGAAAACAGCCGAGGACGTAAGGGAAGTCTTTGCACCC
AAAGTCCAGGGCGTGCTCAACCTGGACGCCGCTACTGCCGAAGTAGTACTTGATTTCTTTCTCGTAA
GTTCATCGATTTCTGCACAGTTGGGCAACGTAGGTCAAGCAGACTACGCTGCTGCTAACTGCTTCAT
GGACGGCGTCGTGAGGGCAAGACAGAGGCAGGCAGAGCAAGGCCTCCGTCATGGACTCTCGATGGC
CGTGGATTGGGGATATTGGGCTGATGGTGGAATGCGCATAGACGCTTTGTCTCTGGGACTTCTGAAG
CAGGCTTGGGGCTTGCTGCCCATGCCGAGAGACGTTGGCTTGCATGCGTTGTATCAAGCTTTTAAAC
ACCGTCAAGCCAATACTCTGGTCCATTACGGCGTAGCCAGCTCTGTTCGTAGGCTCTTGACCCCGACC
ACCCATCGCGTTGAACCACTGGAACAGTGGTTGCTGGCTAATGTCGCCGCTCTCACTGGCGTACCTG
TTGCCGATATTGATCTTCAACAACCTTTTGCTGACTATGGATTTGATTTGCCAATGCGCCAGGCCCTT
GTGCAGAGGCTTAGTGAGACCAGAGACCTCGCGGGTCAAGGTGCGAGACTCCTTCGCTGCGAATGTC
TGCAAGCATTGTTGGATCATATAGTCCAACCTCCATTGACTGACCTTACCGAGAAGACTCCTGCAAT
GGATACACGCGATGACTCACAGACGCGTCACGCCAGGGATTTCATCATAGGCCTTCTCCGTACCGTA
CTCGTTACAGTTACTAGAGTCAGCCCCGCGCAACTCCAAGACGACCGTCATTTCGAACATTATGGCA
TCGACTCCTTGATGGTCATTCAGATGACAGCCGAACTGGAAAAGAGTTTCGGATCACTTTCAAAAAC
TTTGTTCTTTGAACACAGCTCTATCAAAGAGCTCGCCGACTATTTCACAAGACGCCATGCGGAAAGG
CTCGGAGAACTCGCCGGAAGTTAGAACCCAGCTTTCTTGTACAAAGTGGTTCTAGAGCCTGCAGT

mmpD_mod34 gBlock 1:

TAGTGGTTCTGGTGACTACAAGGACGACGACGACAAGCTTACAAGTTTGTACAAAAAAGCAGGCTT
AATGGGGCCGGTTGCCCCGGCACCAACTGTCGCGGAGAGCACCCCTGTTGCTCCTGTGAATGCACTG
CCAGATGCGTCTCTCGCTCAACCGATAGCCATAGTAGGTCTTTCAGGACGCTACCCTAAAGCAGAAG
ATTTGCAAGCGTTCTGGCGCAACCTTGCAGAGGGCGTGGACTGTGTAACAGAGGTACCAGCAGAAC
GCTGGAGCCTGGGTGACTTGGCACAACCTGGCAGGCTGCCCGGAGGCAAGACGCTCTCTAGATGGG
GCGGCTTTCTCGATGATGTAGATAAATTCGACCCGCTCTTTTTTAATATATCGCCACACGATGCCGAG
TACATGGACCCTCAGGAGCGTGTGTTCCTGGAGTGTGCATATGCAACTTTCGAAGATGCCGGTATGG

CCCTTGGTCTTCGTAGCCCATCGGCCCTCCGTCGCACTGGCGTTTACGTAGGTGTTATGTACCAAGAG
TATCAACTCTTCGGCGCCGAGCAGACGCTTTTGGGCCGCCCGATGGCTCTTTCAGGTAGTAGTGCAA
GCATCGCGAATAGAGTATCATGGAGCCTCGGCCTTAGCGGTCCGTCCCTTGCCGTGGACTCCATGTG
TTCCGCTTCACTTACCGCAATTCACCTGGCCTGCCAAGGTATCAGAAGTGGTGATTGTGATGTGGCGT
TGGCAGGTGGTGTAAATGTTTCTGTTCATCCAAACAAATATCTGGGTCTCGCCCAAGGCAACTTCGCT
AGCAGTCAAGGTAGATGTCAATCCTTCGGCGAGGGCGGAGATGGTTATGTTCCCGCAGAAGGCGTG
GGAGCCGTATTGCTGAAGTCATTGAGTCAAGCCGTCGCAGACGGAGACAGGATTTACGGCGTCATA
AGAGGCTCCCAAATCAACCACGGTGGAAAAGCGAACGGTTTCACTGTGCCGAACCCCAACGCGCAA
GCTGATGTGATTACGCAGGCCCTGGCAGCGGCGGGAGTTGACGCAGCCGATGTGGGCTATGTAGAA
GCACATGGTACAGGAACGGCACTTGGTGATCCTATCGAGTTGGCAGGCTTGGCGAAAGCATACCGTC
ACTACACAGATCAGACTGCCTTCTGCGCTTTGGGCTCGGTCAAGTCTAACCTCGGCCATTGTGAATCA
GCAGCTGGAATAGCCGGCTTGACCAAGGTCCTCCTCCAAATGCAGAAGGGTCAGCTTGTACCTTCTC
TTCATTCCGAGGTGTTGAACCCTGCCATTGACTTTGCTCAATCTCCCTTCGTTGTCCAACGCCACTTG
GCTCCATGGCCCCGTCGCGTCAGAAGGGTTGAGGGTGTCACGGAGGAGTGCCCTCGTCTGGCTGGAC
TGTCATCCTTTGGAGCTGGTGGCTCCAATGCGCACCTGATAATTGAGGAGTATCTTGCGCCCACCCA
ACCGGAGTCACAGAGCCGCCAACCCCTTCCCGTGGCCGTACTCCTGTCAGGTGCCAATGCCGAGCGT
CTCCAAGCCAGTGCGAGGCGTCTCCTGGCCTTCTGTGCTTCGCCTGAGGCTCACGACTTGTCCTTGCA
TGACTTGGCGTACACTTTGCAAACGGGCAGGGAAGCGTTGAATGCCAGACTTGGTTTCCTCGCGCAC
AGCATCGACGACGTGCAGGCCTGCTTGCGTGAATACCTGCAGGGAGCGCTTACGTCTGGCCGCGTAC
AAGTGGGCTCTGCACGCCAGGACGAAAACCCGCTTGTCAGGCTTCTGGGCGAGGACGACCTGTCTGC
TATGGTAGCCCAATGGGCAGCCACGGGTCAACAGGACAAGCTTCTTGCACTTTGGGTCAGCGGCGGT
GACATAGACTGGCAAGCCCTGCTGCCGGGACGTTGGGGAAGAATCATATCCCTTCCCGGTTATCCGT
TTGCCAGGGAGAGATTTTGGTACACGGGTATAAACAGTGATATTGGTCAACAACATCCCCGTTCACC
CGGTGTACAGGCCGCTCCGTTGTCGGTGCTCAACTGTAGTGACCTCACGGGCTTGGGCTTTACGTGCC
AGGTAAGTGGTAGGGATTCATTCCTCGCGGACCACCATGTCCAGGGTAAGGCAATGATGGCAGGTTC
AGCATTTCTGGAGCTGGCCAGAAAAGCGGCCGTGGCCGTCGCAACGCCGGATCAGCAAGGTGCCCT
CGTTAGACTGAGCGACGTGGTATGGTTGGCACCCCTCGTACTCGAAGATGGCGATGCACAGTTGAGA
GTCAGGCTGTTTCCCGATGATACGGCGAGGGTGCGTTTCGAAATCGCGACTTCAGGAGCTCAAGCAC
AGGCGCTTAATGTGAGGGGAGTTATTGAACAAGTGCCTATGCCTACAAGGCCTCAGGTTGATCTCCC
TGCGCTTCTGGCAAGATTGCGTTCATCTGCGGATTTTCACGAAAGGAATCCGGGTCAGCTCTACGAC
GCCTATGCGGGCCTTGGCATAGACTACGGTGCAACACATAGATGCCTCAGCGCTGTCTACGTCGGTG
ACGGT

mmpD_mod34 gBlock 2:

AACACATAGATGCCTCAGCGCTGTCTACGTCGGTGACGGTCATGCGCTGGCACTCATAGAGTGGCCGGCCGCACAACCAT
TTGAAGCCTCGTACGGATTGCACCCAAGTATGTTGGACGCAGCTCTGCAAGCGACCCGTGTTCTTTCCGATGAGAGTGCC
TTGCTCCAAGTTCCGTTCGCGCTCGACCAATTGCAAGTCTTTGCTGCCTGTGAACCGAGGATGTGGGTGCAGGTTAATCGC
GATCCTCAAAGGGAAAGTTACGACCTTGAACTCTTCGACGAACTCGGATGGCTCTGTGCGTCATTGAGAGGTTTGAAGGT
ACGCCAGTGGGCGCCGCCCGGTAATGGTGGACTGGTAGAGGGAGATGAGGAAGTTGGACTTCCCGCAGGCGAGGTATT
GCTTGTTCCCCAATGGAGCCCACTTGCTATGCAGGCGGTCACGGCTAGTCCTACCGCGACTTTGGTAATCGGCGCTGATG
CGAGCCAACTTCGTTCTCTGGAGGGCGCCCAAGGACCCCTTCTCAGCTGTGCGAGTGTTCCTAGTGATTCAGTCGAGCAG
TGGGTGACGAGATTTCAGGGACTGCCTGCCTTCGATAGGCTTGTGCTCGTTATCCCACCATCGGCTGCCCGTCACGCAGG
CGACGAGTGCCTTATTGATGAGCAACACGCATGTGCGCTGGAGCAACTGAGGATTGTCCAAGCACTTCTCCGTTGTGGAT
TTGACAAGAGAGCATTGCACTACGTTGTGGTAACAAGAGATGCATGGACCCCGCATGACGCCCAGCCTGTACATGCAGCA
GTACATGGTCTTGCTGGCAGTCTTGCTAAGGAAATGCCTCATTGGACAGTTCAGGCCGTAGACCTCGCCCATCAGGAGCG
TCTCGCCCCGACTCACCTTGCGGTCATGCCCGTGGATCCTGCGCACACTTCGCGCAGGCTTAGAGACGGCGTTTGGGAGG
GACGTCAGCTGCTCCTCCATCGTCCCCAGGCTTCGGCACAGCCCGCTTATCGCTCTAATGGTGTATACGTAGTTATTGGCG
GCTCTGGTGGAATTGGACAAGTGTGGACGGAGTCGGTAGCAAGGGAACATGCTGCTCAAGTAGTTTGGCTCGGTCGTCG
TGCTCTGGATGCATCAATACAAGCAGCACTTGATCAAGTGGCGCTGCATGGTCCACGCCCATTTTACATGAGTGTTGACGC
AACAGATGCGGGACAACTGCAACAAGCGTGTGAGCAAATCCTTAGCCGTTTTGGTCGCATAGACGGATTGGTCCATAGTG
CAATCGTACTTGACGATCGTTCTCTGGCAAATATGAGTGAAGACCAGCTGCGTAGGTCCCTCCAAGCAAAAGTCGATATTT
CAGTTAGGATGGCACAGGTGTTTGGCGGACAGGCATTGGACTTTGTACTCTTCTTTTCTTCATTGATCGCATTTACAAGAA
ACGCGGGCCAAGCCAACTACGCAATGGGTTGTGTCTTCAAAGACGCACTGGCTCGCCGCTGGGGATCGAGCGGCTCGGT
CCCGGTAAAGACTCTCAATTGGGGATATTGGGGTTCAATTGGTGTTGTATCATCGGCTGCATATCGCCAAAGGATGGAGC

GCGAAGGCATAGGTAGCATTGAGCCAAGGGAGGCGATGCAGGCAGTACGCACTCTTCTCGCGAGCCCTCTTGACCAGAT
GGCCTTTATCAAAACATTGAGGCCAATTACTTTGGACGGAAACGCCGAGGCGGACCAATGGGTATCGCAAGCTATGGTA
GCTAATGGTCCGCTCCATATAGAAGCCGACGTCGCGCAATGGGTGGCAGCTGACGGAGCTCACGCCGCCGCGCTGCAAG
ATAAGTATCGCAGAGACGCCCAACGCGAGAGACTCCCGCAGTGCCAAGTTCTTCTTGTGCAATTGCAAGCTGCAGGCCTG
TTCGATCAACCTACCCCGCAAACGACTCAGGCGCTTTACCAGCGTTTGCAAGAGAGGTCAAGTGGAACCGTTGGCGTGGA
TCCGCGTTGGCTCGCACACTCCCTGAGATGGCTGGCCGATGAAGGTTGGTTGGCGACCGATGGCGACACGTGGCGTGTA
CAACGCGTCGTAGCCGCAGAGTCGGCTTGGGCCGCGTGGGATTCCGCCTTGCGTGAAGCAGAGGGCGATGCGGACAGG
CGCGCCCAGGCCCAGCTGGTGGATGCCTGTTTGAGGGCTTTGCCCGCAGTACTCTCAGGACATTTGAGAGCCACGGAAGT
GATGTTTCCCGAAGGCTCTATGGCTAGAGTTGAAGGCATTTATAAGCACAATGCAGTTTCGGACTACTTTAATGAGGTCAT
AACCCAAGCGGTGTTGGCACGTGTCCGTGCGGGACACCAGGCAGGCGCGGCGCCTGTACGTATACTCGAGATCGGCGCT
GGAACCGGCGGAACATCTGCTAGAGTCCTCCAAGGCCTGGATTCTCTCGGTGTCGAAGTTGGTCAGTATGCCTATACAGA
TCTGTCGAGGGCCT

mmpD_mod34 gBlock 3:

CGAAGTTGGTCAGTATGCCTATACAGATCTGTCGAGGGCCTTTCTCATGCATGCGCAAACTGCCTAC
GGAGAAGGTCGTTCGTACCTTGATTATCGCATCCTTGACGTAGAGAAGGATCTGGTGCAACAAGGAC
TTGAACTTGGTTCCTATGATGTCGTTATAGCGACAAATGTCCTTCACGCCACAGCGGATATTCGCCAA
GTTGTTGAACACGCCAAGGCAGCCTTGAAGACCGATGGCTGGCTTCTGCTGAACGAGCTTAGCCAAG
CAAGCTTGTTTTCCCATCTCACGTTCGGCCTCCTGGAGGGCTGGTGGCGCTACAGGGATGAGGCTTTG
CGCCTCCCCGGCACACCAGGCCTGTCAAGCGAAAGGTGGCAACAAGTCCTTGAAGAAGCTGGCTAT
AGAGGAGTCATGTTTCCATGCCAGGCGTTTCACGGACTCGGTCAACAGGTAGTCATGGCCAGGAGTA
ATGGCGTCGTACAGGTTCCGAGAAGAGTGTCCCAACCGGCCGCTCCTAAAGTAGCATTGGCAGCGCA
ACCAGCCGCGGCACCACAACACGGCGTGGCGGCGCCTGTGCCTGCAGCTGCAGCTGACCAGTTCACA
CAAGTGCGCCAGGAAGTCATGGCAAGCGTGGCGAAAGCATTGAAAGTAGCCCTTGAGGTCATAGAT
CCTCAGGAGTCGTTCAGTGACTACGGCCTTGACTCGATAACAGGTGTCAACCTTGTACGCGTTTTGA
ACGAGAGACTTGGCATAGATCTGGGAACGACAGCACTGTTCGACTTTTCCACGGCAACGAGACTTGC
AAGACACATCGTTGAACAACACGGCGCACACTTGCCGCAACCCGCCGTGACTAGGGCCGCTGCTGCC
CCTGTTGCTCCCCTGGCGGTGGCAGTTCCGACTGCCAGTTCGCCCGCTGCAGCACCCGCACCTCGCGT
CCAGTCCCCAGAATCATTCGTACGCGAACCAATTGCTATTATTGGCATGTCAGGCAGGTTTCCGGGC
GCAGAATCTCTTGAGCAGTTGTGGGAACACTTGGCTCAAGGTCGCGACCTCACTACGCCCGTTGAAC
GCTGGGATCTCAGGGGTAGTCAGGCTGATGTTTCGGGCAAGGGTCTTTGCCAAAGGGGAGGACTGCT
CTCAAGGATTGACGAGTTCGATGCGTCTTTTTTTAACATTTCGGGACTGGAAGCTACATGTATGGACC
CGCAACAGAGGTTGTTCCTTGAACAAGCGTGGACAGCCTTGGAAGATGCCGGTTATGTAGGTCAGGC
GACGGAAGGTCAGCCAGTGGGAGTCTACGTCGGTGCGTCGGCATCTGACTACAGGTCGCTGTTTGCG
GAAGCCGCCCCAGCGCAAGCCTTCTGGGGTAATGCATCTAGCATTATCCCCGCAAGGATTGCTTACC
ACTTGGATCTGCAAGGCCCAGCGATTGCGGTGGACACAGCATGTTCGAGCTCGCTTGTAGCTATCCA
CCTGGCCTGCCAAGCTTTGTGGACCGAGGAAATTGGTATGGCTTTGGCTGGAGGAGTCTTTGTACAA
AGCACAGCGGGTTTCTACCAGGCAGCTCAGACCGCAGGCATGTTGTCACCCTCTGGTCATTGTCACG
CTTTCGACGCGAGGGCGGACGGTTTTGTGCCTGCCGAGGGAGTAGGAGTGCTGATCTTGAAACGCCT
GAGGCAGGCACAGGCTGACGGAGATTACATACACGGCGTTATCTCAGGCATCGGCTCGAATCAAGA
TGGTGCGTCGAATGGAATCACGGCACCCTCTGCACAGAGCCAGGAGCGCCTTATAAGGCAGGTTCAT
AGGGACTTCGCAATTGAACCGGGTACGATATCCATGGTCGAAGCCCATGGCACAGGAACACCCTTG
GGAGACCCGATCGAGTTTGAAGGTCTTACTAGGGCCTTTGCACGCCCTGGACAGCAAGCTGGCTATT
GCGCTCTGGGCTCGATCAAGTCCAATCTGGGTCACTGTGTAACGGCCGCCGGAGTCGCGGGTGTTCT
CAAGGTTGCGCTGGCACTTAGACATGGTCAACTTCCTCCGGCTGCAGGTTTCACACAAGCGAATGCT
GCAATCGCCGTGCAGGGATCGCCGTTTTATGTTCCGGACCAACTGCGTCCATGGGCGACCCTTGACG
GAGTTCCTCGCAGGGCGTCCGTGAGCAGTTTCGGATTTAGTGGAACTAACGCGCACTTGGTGATGGA
ACAAGCGCCACTTCCCCAACAGAGGGGAGCTTCCGGCGCGGGTCCGTGGCTTGTAGTATTGTCGGCC
CGTTCGGCCCAGCAGCTTCGTGAACAGGCCTCGAACCTGGCCGCCCATTGTGTGGCGGTACAGAGGC
TTGACATGGCTGCAATGGCTTATACCCTCATGGCAGGAAGAAAACACCATGAATGGCGTTTGGCAAT
GGTTACCCACGATGCTGGTCAATTGCAGCAATCTCTGGAGGGTTGGCTGCAAGGAAGAGACGAAGC
AACTGTTCATAGCGGCCATTGGGATGTCCGTCGTTTTGTCGAGCAGCAAGAGG

mmpD_mod34 gBlock 4:

CGGCCATTGGGATGTCCGTCGTTTTGTCGAGCAGCAAGAGGTTCTCCAAGCCGCACGCGAGTGTCTT
ATCCGCCTTATGGGTAACAGGGACCAAAGTACGAGGGGCGAGGCCCTCACAAGCCTTGCTCAGGCCT
TCGTCCAAGGATATGAGCTGGATTACAGCAGTTTGTTCATGGCGGACGAAAGACGTCGTGTCCCCCT
TCCTACCTACCCTTTCCTGCGTCAGTCTTATTGGGTGTCAGCACCTGAAGTAGAGGAACCAACAGCTG
TACCGGCCGAGGAGGTCGCCGGTCAGTGGCTTTTTTGTGCGGAACAATGGAGAGCTCATGACCTTCC
CTCAAACTTTGATTGGGCTGCGTCAATAAAGGCTCGTAGCGGAGCAAGGGTTGCTGTGGTAACCCGC
GAGCAGGCCCGTTTCGACGCTTTTTCGCAACTGTTGTGTACCTTGTGGGCCCAAGCTGGTTCCCCCCC
CCCGTCTGTGCAGAGGCTCTTGCCTGACGCTCTCCCTGCACGCGGAGCTCTTCCAGAGGTCATCTTCT
TTATAGGAGAGAACTCGTCTGATGGACGTGCGTCACCACCGGGAGGACTGGACGCCATTAAACAGG
TTCTTGAGGTCTCTAGACACTTGATGGCGACCGATTGGGAGCATGCGGTCGAATGTTTCCACGTGTA
CCGTGCGTCCGGTTCGGGCTCGGACTGTGACGCTCAGGCGATAAGTGGCTTTGCCGCTAGCGCGTCT
CTTGAAAATGCCCGCCATGGCTGGACTCTTTTGGAGTGCGACGCCGTTGCTGGTCCTGCGTCAGACTT
GCAATTGATAGTCCGTGAGTGGCTCGCGGCAGATAGGCAGGGCAGTGCTCCTGCCGTAGTAAAGTAT
CAAGGAACTCAGCGTCTGAGGCGTCTTCTTGTTCAGCGTGAAGCCGCCGAAGCACAGGCACCAAGAT
TCAGACGTGACGGCCACTACCTCGTTGTTGGCGGACTGGGTCCTGTAGGTGAACTCTTGTGTCAAGA
GCTTGCCGGTACGTTTCAAGCTCGCCTCTCCATAGTCAGTAGGTCGGCACCGAACGCAGAGCAACAA
GCTCGCTGCGACGTTTTGAGGGCCCTGGGTGCAACCGTCGATTACAGCTCTGTTGATATCACCGATCT
CCCGGCCCTTCAGCAGGTTGTCGAGAGGCGCGTGCAGGCAGCGGGTCCCCTCCAAGGTGTAATACAT
TTGGCACGTCTCGTTGATGATGGCTTGATAGTGGCTAAGCAATGGGAGGCATTTGAAAGGAGTATTG
GCGCAAAAGTCAAGGGTACCATAAATCTGGATCTGGCAACTGCTGATCAACCACTGGCCTTTTTCAC
CGTTTTTAGCAGTATGGCGGCCTATGGAATACGTGGTAGCGCAGACTACGGATACGCGGCGGCTTTC
CAGAACGCTTTCGTGAGGTACAGACAGACACTGGAAAGGCAAGGAAAGCGCCATGGCGTGAGTAGC
GCATGCACGTGGGGAGCTTGGTGCGTAGACAGGTACATGCCGGCAAATAGGCCCGCACACTTGCAG
GCTTTGGGCCTCGGCCTCATAGACATGCAGTCAGCGTTTGAGTATCTTCACACTGGCGCTGAACCAG
GAGTTGTCGGTGCGATGAAGGTGCTCGATCCAGCTCTTGCTTGCCGCGGCCTTGGCATAGAGCCACT
TCCAGCTTCTTCCGTACCCCCCGCGGGCGATGCCATAGGTCTTCACGCACAGCTTCACGCATGGGAG
CAAGCTTGCGAGGCAGGAGCAACGCCGGACGCTGCACAATGGTGCAAAGTGTTGTCCAGGTTCGAA
CTCGACGCCTTGGACGAGGCCACGATCGATAGGCTCGATCGTTTGTTTTTTCCAAGCGCACCAAGGC
CAGCCACCGTTGGCGCCGGAGCCTTGCTTGAACAAGTTAGGGAAGTGATTGAGAGAGTGCTCGTGGT
AGACGAGCCTGATTTGGACACAGCATTTTCCCGCTATGGTATGGACAGTGTAGGTGCAATGCAGGTT
TCATCAGCATTGTCAAGGGCGCTTGGCTGGCTCGTTGAACCCCGTTGGCTCGTCCAACACGCTACGA
TCCGTGCGTTGGCTGAATTTCTGCAAAGTAGAAATGAAGCAGCTACGCAGTGAAACCCAGCTTTCTT
GTACAAAGTGGTTCTAGAGCCTGCAGTCTCGACAAGCTTGTCGAGAAGTACTAG