Structural and Functional Studies of Polyketide Synthases

Ieva Drulyte

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Astbury Centre for Structural Molecular Biology

September 2018

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I have been extremely fortunate to have had the amazing support of the university, family, friends and colleagues near and far. Without this support this thesis would not have been possible.

First and foremost, I would like to thank my supervisor Alan Berry for his continuous encouragement and patience during the last four years. I would like to express my gratitude to my unofficial co-supervisors, Glyn Hemsworth and Neil Ranson, who have been irreplaceable mentors to me and taught me more about structural biology than I could give them credit for here. I would also like to thank Sheena Radford and Ryan Seipke and all their useful advice during many corridor conversations.

A lot of the work described in this thesis would not have been possible without the help from the fantastic people running the Astbury facilities. Firstly, a huge thanks to the ABSL staff Rebecca Thompson, Emma Hesketh, Daniel Maskell and Martin Fuller. You are an EM super team, and I am honoured to call you my friends. Secondly, grateful thanks to Chi Trinh for all his help with X-ray crystallography, especially during the late evening data collection sessions, and to Arnout Kaveldra, Gary Thompson and Anastasia Zhuravleva for their input with the NMR spectroscopy. Finally, thanks to James Ault, Rachel George, Iain Manfield, Nasir Khan and Maren Thomsen for their help with various protein biochemistry analyses.

My collaborators from the University of Tübingen, Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute (HKI) in Jena and the University of Bristol have been invaluable for their contribution to my thesis work. I would like to thank Marc van der Kamp for joining the IdmH project and for his input with the *in-silico* analysis. Next, I would like to thank the rhizoxin PKS collaborators, Christian Hertweck, Maria Dell, Georg Zocher and their colleagues, for allowing me to work on this exciting project and for being great hosts during our visit to Jena.

My grateful thanks are also extended to Sheena Radford, David Brockwell and Glyn Hemsworth for letting me use items of their equipment. I also wish to acknowledge the financial support from The Wellcome Trust, without whose

generous consumables fund I would not have had an opportunity to study such complex biological molecules in such unprecedented detail.

My PhD experience would undoubtedly have been a lot less enjoyable without the amazing people I met along the way. So firstly, big thanks to all Hemrry lab members, past and present, for keeping the lab spirit up and putting up with my questionable lab music choices. A particular thanks to Jana Obajdin for being the *best supervisee ever*, best of luck to you in your academic career! Thanks to all level 10 people for good chats, whether it be in the microwave queue or at the yearly Christmas do! Special thanks to Alex, Hugh and Anton for always finding time for me, whether it be for a friendly chat or to help me troubleshoot my experiments. Finally, thank you to all EM office people, past and present, for welcoming me and always being willing to help me with my EM problems!

I wish to thank my Mum, Dad, my grandmothers, Andrius, Liena and the rest of my family for their love and support in everything I set my mind to do. I will try my hardest to continue to make you proud!

A big thanks goes to Jess and Samantha, thank you for always being there for me and for keeping me in the real world!

Most importantly, I wish to thank my best friend, team mate and a husband-to-be, Dan. Thank you for all your love and support, and for putting up with me through the harder times of my PhD. I love you!

“Clarke’s Second Law: The only way of discovering the limits of possible is to venture a little way past them into the impossible.”
Arthur C. Clarke, 1973 (Clarke, 1973)

Abstract

Polyketides, natural products produced by multi-domain polyketide synthases (PKSs), have proven to be excellent starting points for drug discovery. Rational engineering of PKSs holds much promise for the generation of novel polyketide pharmaceuticals, however to enable this we need to gain a better understanding of how the mature polyketides are generated and how individual modules within a polyketide synthase assemble and interact. Here, work was performed to investigate three polypeptides from natural product indanomycin and rhizoxin biosynthesis, including the candidate polyketide cyclase IdmH, the fourth subunit of the indanomycin megasynthase, IdmO, and the branching module from the rhizoxin PKS.

Indanomycin needs to undergo several transformations by post-PKS tailoring enzymes. One such enzyme, IdmH, has been hypothesised to act as a cyclase and catalyse the formation of the indane ring via a Diels-Alder reaction. Crystal structure of the wild-type IdmH was determined to 2.7 Å resolution and the interactions between IdmH and its proposed product indanomycin were characterised using NMR spectroscopy and *in silico* methods. Fully-reducing IdmO module was successfully expressed and purified. Characterisation by negative-stain electron microscopy resulted in a low-resolution model of IdmO, while attempts to carry out cryo-electron microscopy (cryo-EM) analysis revealed a number of difficulties associated with the denaturation of this large complex during cryo-EM grid preparation. A similar cryo-EM approach was utilised to study the branching module from the rhizoxin PKS. A 3.7 Å resolution map was determined for this module containing the ketosynthase, branching and acyl carrier protein (ACP) domains. Two ACP binding sites were identified, which can help explain the unorthodox activity of this module.

This research has provided valuable insights into different aspects of PKS biology ranging from polyketide tailoring and branching to the assembly of the intact modules and forms a solid basis for future studies of these fascinating biosynthetic machines.

Table of Contents

List of Tables

List of Figures

Abbreviations

ACP	Acyl carrier protein
αKGDH	α-ketoglutarate dehydrogenase
AT	Acyltransferase
B	Branching domain
BGC	Biosynthetic gene cluster
BS3	Bis(sulfosuccinimidyl)suberate
BSA	Bovine serum albumin
CD	Circular dichroism
CoA	Coenzyme A
CP	Carrier protein
Cryo-EM	Cryo-electron microscopy
CSP	Chemical shift perturbation
CTF	Contrast transfer function
DD	Dimerisation domain
DDM	Dodecylmaltoside
6-dEB	6-deoxyerythronolide B
DEBS	6-deoxyerythronolide B Synthase
DH	Dehydratase
DMSO	Dimethyl sulfoxide
dNTPs	Deoxynucleotide triphosphates
DTT	Dithiothreitol
EM	Electron microscopy
ER	Enoyl reductase
FAS	Fatty acid synthase (mFAS = mammalian FAS)
FSC	Fourier shell correlation

FT	Fourier transform
HTS	High-throughput screening
HPLC	High pressure (or performance) liquid chromatography
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JCSG	Joint Centre for Structural Genomics
KR	Ketoreductase
KS	Ketosynthase
LC/MS	Liquid chromatography/ mass spectrometry
LM	Loading module
MAS	Mycocerosic acid synthase
6-MSA	6-methyl-salicylic acid
MS	Mass spectrometry
MWCO	Molecular weight cut off
NADH	Reduced nicotinamide adenine dinucleotide
NEB	New England Biolabs
NMR	Nuclear magnetic resonance
NP	Natural product
nPCR	Nested polymerase chain reaction
NUS	Non-uniform sampling
NRPS	Non-ribosomal peptide synthetase
NRPS-PKS	Non-ribosomal peptide synthetase – polyketide synthase
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
PAA	Poly(acrylic acid sodium salt)
PCR	Polymerase chain reaction

PDB	Protein data bank
PEG	Polyethylene glycol
pI	Isoelectric point
PKS	Polyketide synthase (iPKS = iterative PKS, mPKS = modular PKS)
PPTase	Phosphopantetheinyl transferase
QM/MM	Quantum mechanics/ molecular mechanics
RMS	Root Mean Square
RMSD	Root Mean Square Deviation
rpm	Revolutions per minute
SAD	Single wavelength anomalous dispersion
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC-MALLS	Size-exclusion chromatography-multi angle laser light scattering
SLS	Static light scattering
TE	Thioesterase
TEV	Tobacco Etch Virus
TOF-ESI-MS	Time-of-flight electrospray ionisation mass spectrometry
TPP	Thiamine pyrophosphate
TROSY	Transverse relaxation-optimised spectroscopy

Chapter 1 Introduction

1.1 Natural products as therapeutics

Natural products (NPs) are secondary metabolites produced by the Earth's flora and fauna. While not essential for an organism's survival, these complex molecules (Figure 1-1) are a product of natural selection and often convey a competitive advantage to the organism that produces them (Williams et al., 1989).

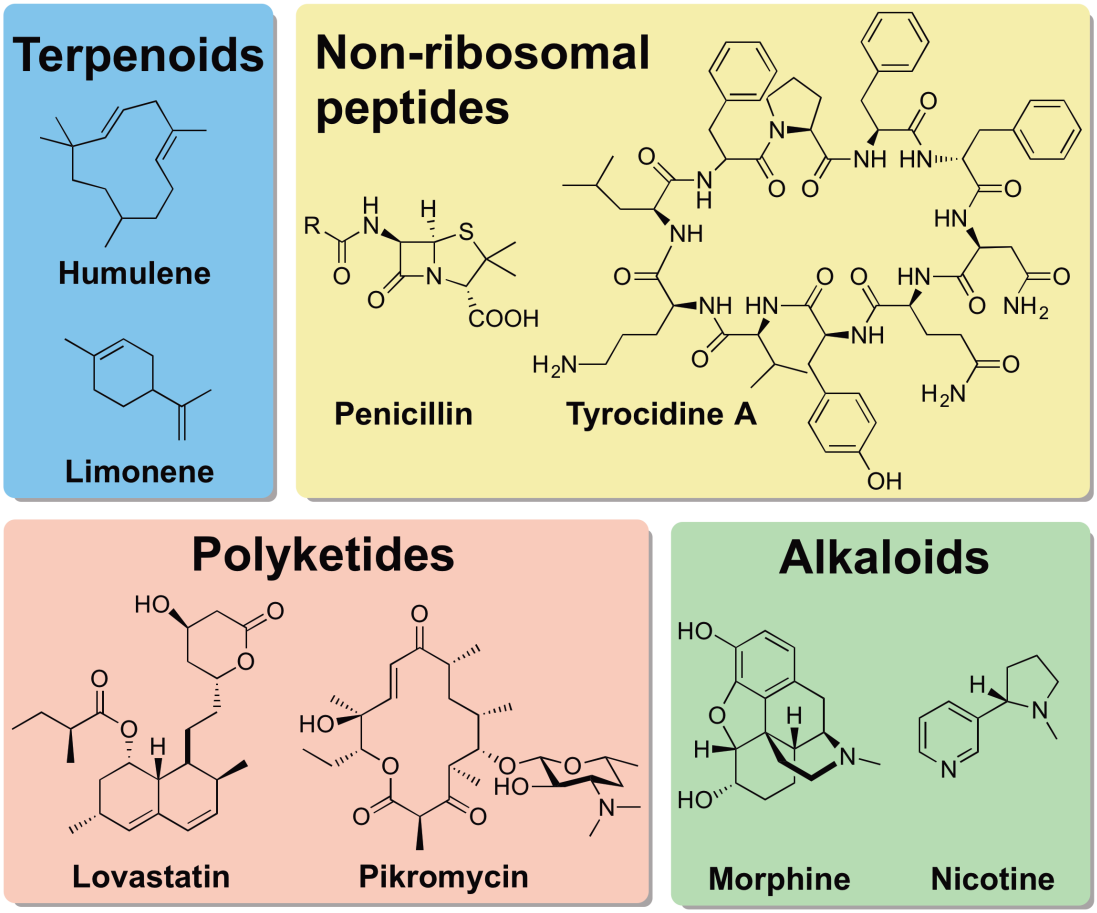


Figure 1-1 The structural diversity between different natural products. *Natural products are secondary metabolites produced by living organisms. Four main classes of natural products are terpenoids, non-ribosomal peptides, polyketides and alkaloids.*

For example, the polyketide antibiotic erythromycin A has evolved to bind the competing organism's ribosomes to impede its growth (Bulkley et al., 2010) while other natural products are known to be involved in strengthening of the cellular wall (Gavalda et al., 2014) and signalling (Zucko et al., 2007). For thousands of

years, people have relied on natural products for the treatment of human diseases, and approximately one-third of the current pharmacopoeia is derived from natural products (Patridge et al., 2016; Dias et al., 2012; Newman and Cragg, 2012).

The discovery of penicillin in 1928 (Aldridge et al., 1999) led to the start of a period from the 1950s to 1960s known as the Golden Age of natural product drug discovery (Shen, 2015). During this time big pharma companies established high-throughput screening (HTS) programs which led to the development and marketing of many currently-used natural product antibiotics (Wenzel, 2004) and other blockbuster drugs, such as statins (Tobert, 2003), the antimalaria drug Artemisinin (Miller and Su, 2011) and the anticancer drug Taxol® (Wani et al., 1971). In fact, between the years 1981 and 2013, 50% of all approved pharmaceuticals originated from natural sources (David et al., 2014).

There are an estimated two million species of living organisms on our planet; yet less than 5% have been evaluated for therapeutic potential, which highlights the potential of natural product drug discovery pipelines (Charlop-Powers et al., 2014; Colegate and Molyneux, 2007). However, the decline in new drug approval, reduction in new lead compounds, rising costs and other difficulties associated with isolation, purification and scaling-up procedures led to most pharma companies abandoning their natural product drug discovery programs (Kingston, 2011; Li and Vederas, 2009; McChesney et al., 2007; Dickson and Gagnon, 2004; David et al., 2014).

While there are very few pharmaceutical HTS R&D programs left, the future of natural product research is bright. Recent advances in metagenomic and metabolomic analysis, decreasing costs of high-throughput sequencing and innovations in culturing previously unculturable organisms make natural product research more accessible to academia and start-ups than ever before (Ling et al., 2015; Beutler, 2009). Fast and efficient gene synthesis, CRISPR and other recombination technologies have revolutionised the field of synthetic biology (Kim et al., 2015; Smanski et al., 2016). It is now possible to re-engineer complete biosynthetic pathways to alter or increase the production of the natural product, as exemplified by successful engineering of a complete artemisinic acid biosynthetic pathway in *Saccharomyces cerevisiae* (baker's yeast) (Paddon et al., 2013). With today's technology, rational engineering of natural product

biosynthesis is also feasible (Goss et al., 2012). However, understanding the architecture and dynamic properties of natural product biosynthetic machinery remains a central challenge in natural product research.

In this chapter, I will be introducing polyketide natural products and the progress made to date in understanding the fascinating biosynthetic machinery that produces them. I will also give an overview of polyketide engineering efforts and discuss the reasons why they have rarely paid off to date.

1.2 Polyketide-derived drugs

One particular class of natural products are the polyketides (Figure 1-1). Polyketides exhibit a stunning array of structural and functional features and are known to possess many useful properties, such as antimicrobial, antitumour, anticholesterol, antifungal and antiparasitic activities among others. Natural screening has given us nearly forty polyketide-derived drugs (as of 2006), such as Rapamune® (anticancer, immunosuppressant), Zithromax® (antimicrobial) and Mevacor® (hypocholesterolemic) (Figure 1-2) (Zhang and Wilkinson, 2007).

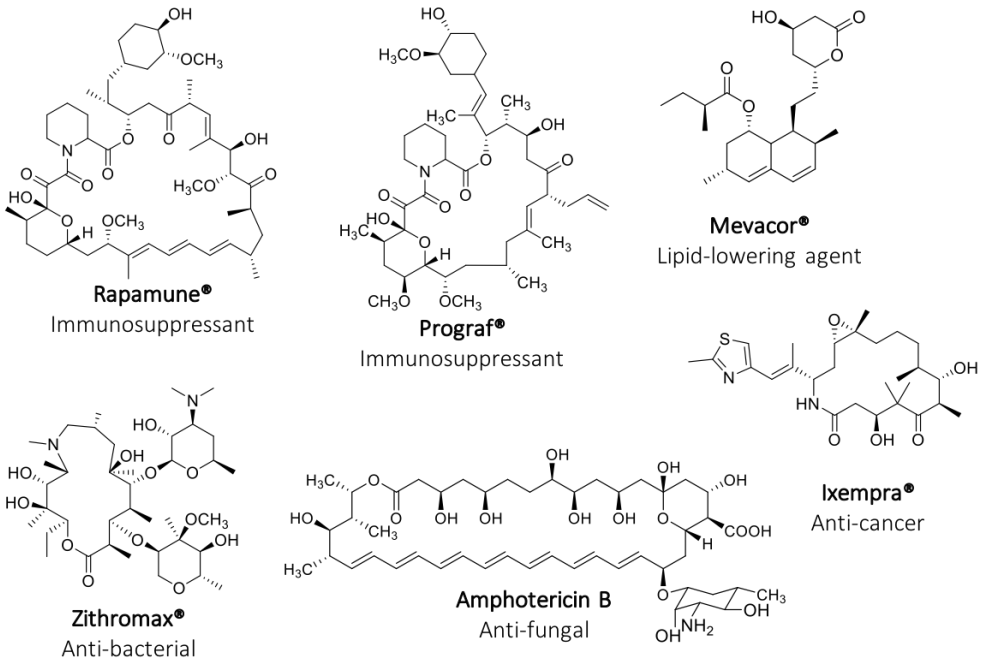


Figure 1-2 Several examples of polyketide-derived pharmaceutical drugs. *Rapamune® is derived from rapamycin, Prograf® from FK506, Mavacor® from lovastatin, Zithromax® from erythromycin and Ixempra® from epitholone B.*

Due to complex stereochemistry and many sensitive groups, total synthesis of polyketides remains extremely challenging (Wong and Khosla, 2012). For example, synthetic generation of rapamycin involves approximately one hundred organic chemistry steps and results in a yield of less than 0.5% (Zhang and Wilkinson, 2007; Goss et al., 2012). In nature, these molecules are produced by enzymes known as polyketide synthases (PKSs).

1.3 Polyketide synthases

1.3.1 A brief history of the discovery of polyketide synthases

The first polyketide, orcinol, was serendipitously discovered by a chemist named James Collie back in 1893 (Collie and Myers, 1893). He came up with a hypothesis (which turned out to be correct) of orcinol formation via a polyketone intermediate. Unfortunately, his ideas about the biosynthesis of polyketides appeared too far-fetched to other chemists and nearly half a century passed until the polyketide field revived again (Staunton and Weissman, 2001).

The next significant breakthrough in polyketide research was a result of thirty years of organic chemistry research by Sir Robert Robinson, which ultimately culminated in a book called “The Structural Relations of Natural Products” (Robinson, 1955). Arthur Birch, who trained in Robinson’s laboratory, took his supervisor’s ideas forward and came up with a hypothesis of head-to-tale fusion of acetate units to form natural products, and his ground-breaking results on 6-methyl-salicylic acid (6-MSA) were published by *The Australian Journal of Chemistry* (Birch et al., 1955).

Following on from the work of Robinson and Birch, researchers gained many valuable insights into polyketide biosynthesis from isotopic labelling of precursor and starter molecules, nuclear magnetic resonance (NMR) spectroscopy studies on the fate of hydrogen atoms in intermediates, and by introducing biomimetic precursors to study specificity (Staunton and Weissman, 2001). The latter led to the hypothesis that the *in vivo* biosynthesis of polyketides must be controlled enzymatically.

The first polyketide synthase, responsible for the synthesis of 6-MSA, was isolated from a fungal source in 1984 (Malpartida and Hopwood, 1984). However, despite the advances in molecular biology techniques and great efforts to identify

the genes responsible for the production of other polyketides, it took another six years before erythromycin synthase, and the biosynthetic cluster of genes responsible for the production of this synthase, were discovered by Peter Leadlay and Leonard Katz's groups (Cortés et al., 1990; Tuan et al., 1990; Caffrey et al., 1992; Donadio et al., 1991). These two discoveries mark the start of the modern field of polyketide synthase research.

1.3.2 Types of polyketide synthases

Broadly speaking, all polyketide synthases can be classified into three classes (Ridley et al., 2008). Type I PKSs are large multi-domain enzymes in which each module is comprised of three to six catalytic domains and is responsible for a single polyketide chain elongation step. Type I polyketide synthases are further divided into modular (termed mPKS) and iterative (termed iPKS) types. The molecular mass of fully assembled type I mPKSs often reaches the megadalton scale as exemplified by 6-deoxyerythronolide B synthase (DEBS) (Caffrey et al., 1992). Other examples of modular type I PKSs are those responsible for the production of rapamycin (Schwecke et al., 1995), FK506 (Motamedi et al., 1997), the avermectins (Ikeda et al., 1999) and pikromycin (Xue et al., 1998).

Iterative PKSs have identical catalytic domains to the mPKS; however, they only consist of a single module. The polyketide product is synthesised through iterative use of the same active sites (Staunton and Weissman, 2001). Many iterative PKSs are found in fungi and are known as fungal iPKS. To complicate matters further, iPKSs are divided into a further three classes – non-reducing, partially-reducing and fully-reducing. Notable examples of type I fungal iPKS include ones involved in 6-MSA (Shoolinghin-Jordan and Campuzano, 1999) and lovastatin (Hendrickson et al., 1999; Kennedy et al., 1999) biosynthesis. The human fatty acid synthase (FAS) is actually very similar to an iterative type I PKS (Maier et al., 2008; Keatinge-Clay, 2012). In unimodular FAS, the same catalytic domains are employed iteratively to build a sixteen-carbon length fatty acid, palmitate.

Type II PKSs are dissociable systems when in their inactive state, but form a highly structured modular arrangement when in their active state (Staunton and Weissman, 2001). Characteristically, type II PKSs are involved in the biosynthesis of phenolic compounds, such as actinorhodin (Malpartida and Hopwood, 1984). Type III PKSs, also known as chalcone synthase-like PKSs

(Ferrer et al., 1999), are found in higher plants (Schröder, 1999; Ferrer, 2008). They are fundamentally different from types I and II in that they comprise a simple homodimer of ketosynthases and the growing polyketide is never attached to the protein directly (Shen, 2003; Staunton and Weissman, 2001).

Though beyond the scope of this introduction, it is worth mentioning that the diversity in polyketide synthases extend beyond the three common types described above (Shen, 2003). For example, type I modular PKSs in myxobacteria lack a cognate acyltransferase (AT) domain and require an *in trans* acting AT domain instead (Cheng et al., 2003). In addition, some type II PKSs lack a cognate acyl carrier protein (ACP) and utilise the substrate directly (Kwon et al., 2001).

This thesis focuses on the type I modular polyketide synthases.

1.3.3 Polyketide biosynthetic gene clusters

As a result of the next-generation genome sequencing technology (Janitz, 2011) and the advances in the *in-silico* detection of biosynthetic gene clusters (BGCs) (Zazopoulos et al., 2003), the number of identified BGCs has risen dramatically since the 1990s (with a total of 1172 clusters with known end products identified as of 2015) (McDaniel et al., 2005). It would be impossible to cover all (or most) known polyketide BGCs; therefore, throughout this introduction, I will focus on the model PKS system responsible for the production of erythromycin.

The erythromycin BGC is approximately 60 kb in length and can be divided into three parts (Figure 1-3). The middle part contains open reading frames (ORFs) for the genes *eryAI to III*, encoding three large polypeptides, DEBS 1-3, responsible for the production of 6-deoxyerythronolide B (6-dEB) (precursor of erythromycin). On both sides of the *ery* genes, there are a number of non-PKS genes encoding proteins responsible for antibiotic resistance, biosynthesis of precursor molecules and tailoring steps to produce the final erythromycin product from 6-dEB (Staunton and Weissman, 2001; Summers et al., 1997).

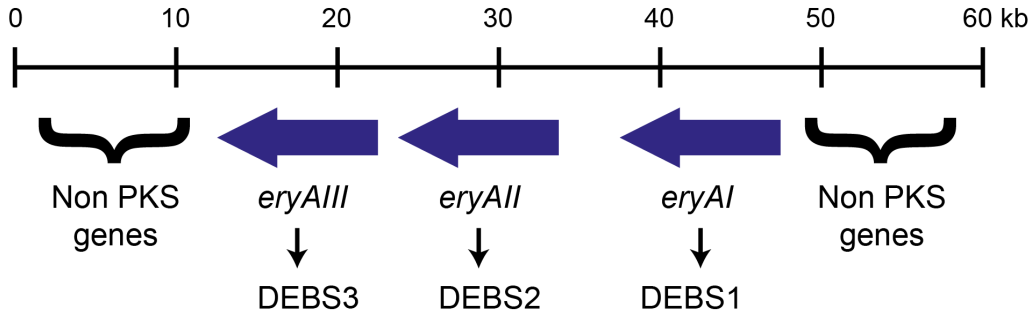


Figure 1-3 The erythromycin biosynthetic gene cluster. *The size of the cluster is approximately 60 kb. Three large ORFs in the middle of the BGC comprise genes eryAI - III, which encode erythromycin polyketide synthase subunits termed DEBS1 - 3. Ery genes are surrounded by two regions comprising non-PKS genes involved in erythromycin resistance and tailoring reactions to produce the mature erythromycin.*

1.3.4 Type I modular PKS organisation

Minimally, each module consists of at least three different domains: an acyltransferase (AT) responsible for an extender unit selection from the cellular pool, a ketosynthase (KS) which catalyses a Claisen-type condensation reaction between the selected extender unit and the growing polyketide chain and an acyl carrier protein (ACP) which acts as a covalent attachment point for the growing polyketide intermediate. The resulting polyketide intermediate can be further tailored by one or more of the accessory domains – a ketoreductase (KR) which yields a hydroxyl group, a dehydratase (DH) which yields a double bond or an enoyl reductase (ER) which produces a fully-reduced methylene group (Weissman, 2015b).

Modular PKSs containing the AT domains are called *cis*-AT PKSs, and the modules of this type of PKS can be classified into four types: KS + AT + ACP (minimal, or α-module), KS + AT + KR_s + KR_c + ACP (β-module), KS + AT + DH + KR_s + KR_c + ACP (ɣ-module), KS + AT + DH + KR_s + ER + KR_c + ACP (δ-module) (where KR_s denotes structural part of a ketoreductase and KR_c denotes catalytical part of a ketoreductase) (Keatinge-Clay, 2012).

The modular architecture of the model system, 6-deoxyerythronolide B synthase (DEBS) is shown in Figure 1-4.

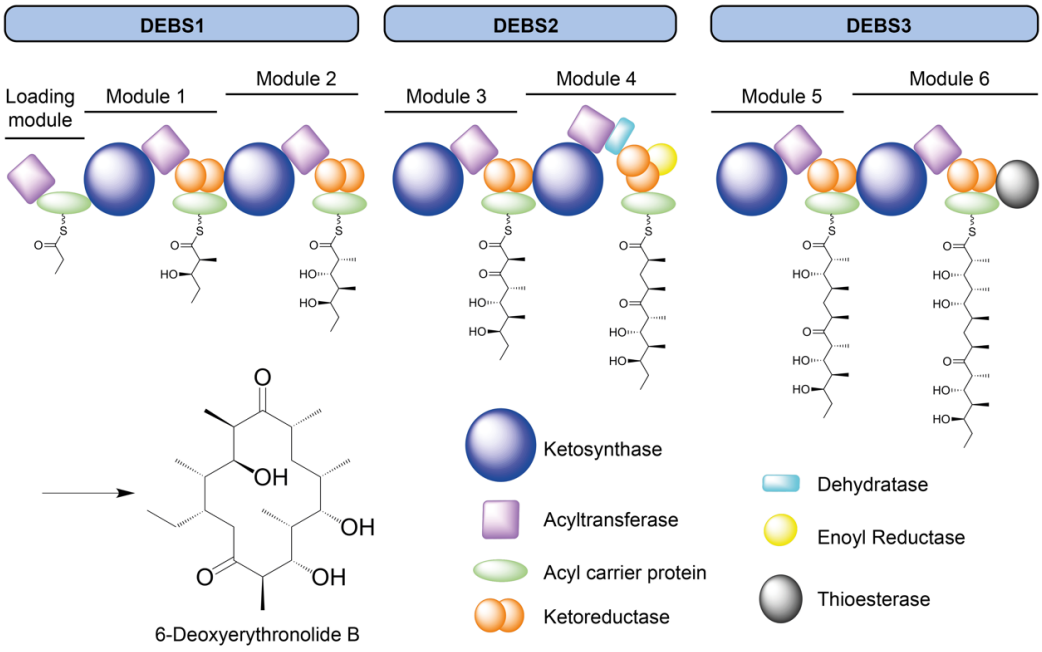


Figure 1-4 The architecture of DEBS. *The synthase is encoded in three large genes which result in the production of three large polypeptides – DEBS1, DEBS2 and DEBS3. The three subunits further comprise a loading module, six extension modules and a terminal thioesterase domain. Predicted polyketide intermediates from each module are shown attached to the acyl carrier proteins (Khosla et al., 2007).*

This section will draw closer focus to each of the type I modular polyketide synthase domains and give an overview of their structure and function.

1.3.4.1 Acyl carrier protein (ACP)

The ACPs are small, ~80-residue proteins located at the C-terminal end of the PKS modules. Their primary responsibility is the intra- and intermodular transfer of the growing polyketide intermediates between the active sites of the PKS catalytic domains. Polyketide intermediates are covalently attached to a conserved serine on the ACP via a flexible phosphopantetheinyl group, which is added onto the ACP by an enzyme, phosphopantetheinyl transferase (PPTase), using CoA as a substrate. The phosphopantetheinylation of the ACP converts the ACP from its *apo*- to the *holo*- form (Figure 1-5) (Fischbach and Walsh, 2006).

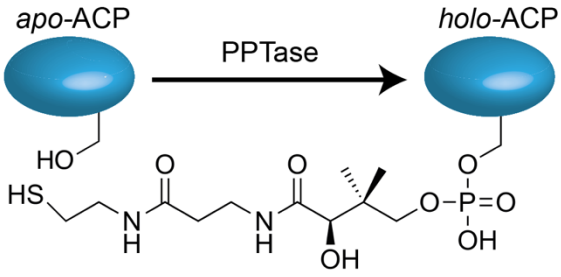


Figure 1-5 Post-translational modification of the ACP domain. *Inactive (or apo-) APC is converted to its active (holo-) form by an enzyme PPTase, which modifies the conserved serine residue with a covalent phosphopantetheinyl group.*

ACPs from different types of PKSs, FASs, NRPSs (non-ribosomal peptide synthetases) and other systems are highly conserved and share the same four-helix bundle architecture (Keatinge-Clay, 2012). There are four X-ray crystallography and many solution NMR spectroscopy polyketide synthase ACP structures deposited in the protein database (PDB). Selection of these include β-branching ACP, PDB code 2L22 (Haines et al., 2013), frenolicin synthase ACP, PDB code 1OR5 (Q. Li et al., 2003), actinorhodin synthase ACP, PDB codes 2AF8 (Crump et al., 1997), 2K0X, 2K0Y (Evans et al., 2008) and 2KG6 (Evans et al., 2009), and DEBS ACP, PDB codes 2JU1 and 2JU2 (Alekseyev et al., 2007). The solution NMR ACP structure of the *E. coli* DEBS (PDB code 2JU2) is shown in Figure 1-6.

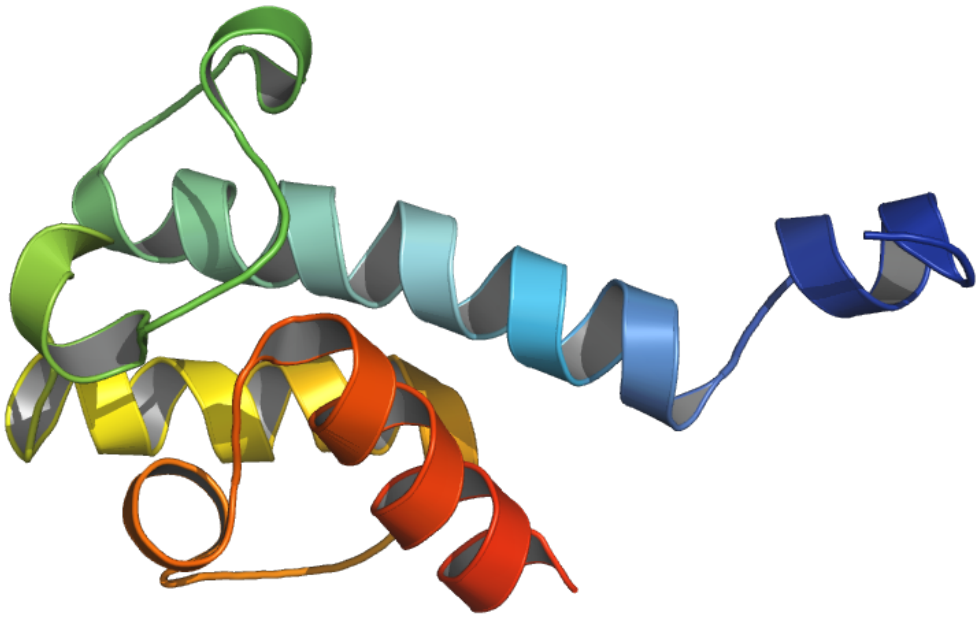


Figure 1-6 The acyl carrier protein from the second module of DEBS (PDB code 2JU2). *The ACP comprises a bundle of four-helices and an additional four-residue helix (known as "helix 0") shown in dark blue. Helix 0 is thought to be involved in docking to the proceeding module of the PKS (Keatinge-Clay, 2012).*

Besides shuttling the intermediates, the ACPs have also been shown to play a role in stabilising the polyketides bound to them (Castonguay et al., 2007) and, through protein-protein interactions, determining the specificity of the KS domain (Kapur et al., 2010; Chen et al., 2006). What is more, mPKS ACP domains have been shown to contain an intriguing additional helix, comprising no more than four residues. This helix, termed “helix 0”, does not interact with the rest of the ACP and is thought to play a role in the docking to the downstream PKS module, facilitating a forward transfer of polyketide intermediates within the assembly line (Keatinge-Clay, 2012; Chandran et al., 2006).

1.3.4.2 Acyltransferase (AT) domain

The AT domains are comprised of ~300 residues and form two subdomains: a ~240-residue catalytic subdomain with an α/β-hydrolase fold and a ~60-residue ferredoxin-like (β/α/β/α/β) subdomain, which is thought to be involved in the interactions with the ACP (Park et al., 2014; Tang et al., 2006).

There are two types of AT domains found in type I PKSs: *cis*- and *trans*-. *Cis*-ATs are located within a PKS module and are specific to the ACP of that module, while *trans*-ATs are standalone proteins which interact with multiple module ACPs.

A number of *cis*- and *trans*-AT domains have been characterised by X-ray crystallography. A selection of these include AT from enediyne synthase, PDB code 4AMM (Liew et al., 2012), AT from splenocin synthase, PDB code 5YDA (Y. Li et al., 2018), malonyl-CoA:ACP transacylase (MAT), PDB code 1NM2 (Keatinge-Clay et al., 2003) and AT from disorazole synthase, PDB code 3SBM (Wong et al., 2011). A crystal structure of the AT domain from the type I iPKS from enediyne biosynthesis (PDB code 4AMM) is shown in Figure 1-7 (Liew et al., 2012).

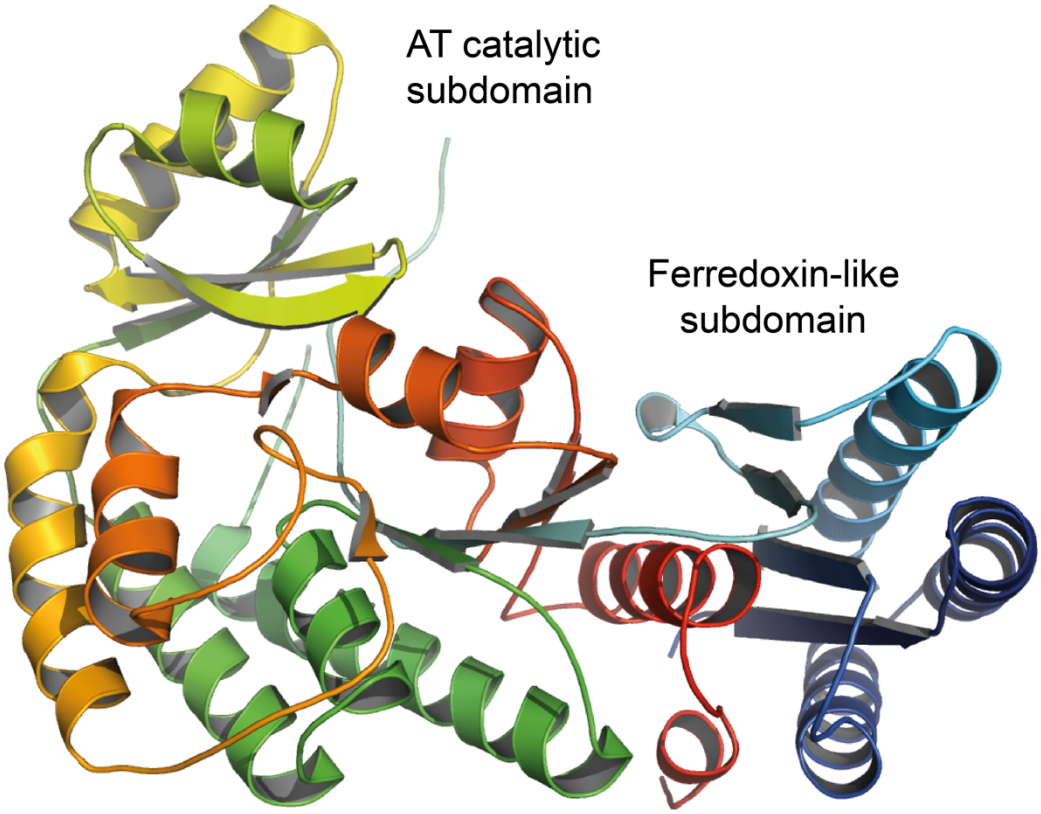


Figure 1-7 A crystal structure of the AT domain from the type I iterative polyketide synthase from enediyne biosynthetic pathway (PDB code 4AMM) (Liew et al., 2012). *The protein is comprised of two subdomains: a catalytic subdomain and a ferredoxin-like subdomain, which is thought to be involved in the interactions with the ACP (Park et al., 2014).*

The role of the AT is to select and load the extender unit, such as malonyl-CoA, methylmalonyl-CoA, ethylmalonyl-CoA, methoxymalonyl-CoA, among others, to the ACP (Chan and Thomas, 2010). The reaction mechanism has been shown to be similar to that of a serine protease and occur via a Ping-Pong mechanism through an AT-acyl intermediate (Dunn et al., 2013). In brief, the AT binds an extender unit and forms an AT-acyl intermediate. This intermediate is then subjected to nucleophilic attack by the thiol residue of the phosphopantetheinyl prosthetic group of the *holo*-ACP (Figure 1-8) (Khosla et al., 2007).

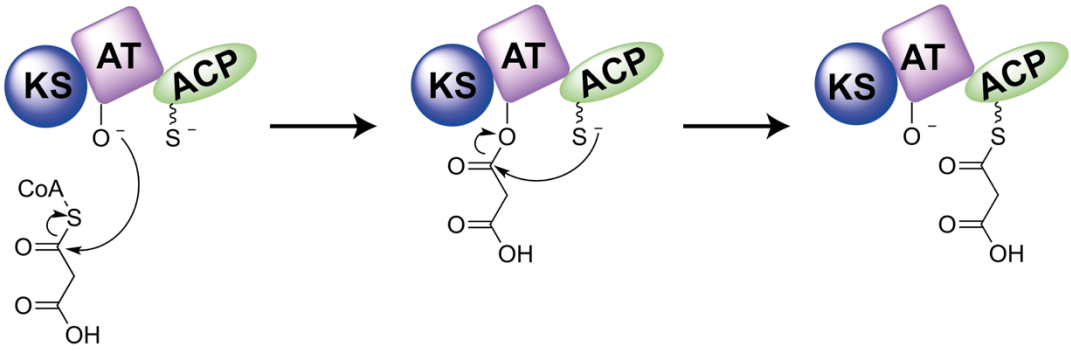


***Figure 1-8 The catalytic cycle of the AT**. First, the conserved serine of the AT attacks the extender unit (malonyl-CoA in this case) to form an AT-malonate intermediate. This intermediate is then subjected to a nucleophilic attack by the thiol moiety of the phosphopantetheinyl arm of the intramodular holo-ACP giving rise to the ACP-malonate complex.*

The specificity of the AT is determined by a number of conserved residues, known as the “fingerprint” residues. For example, a HAFH motif in the active site confers specificity towards malonyl-CoA, while a YASH motif determines the specificity towards methylmalonyl-CoA (Yadav et al., 2003; Starcevic et al., 2008). The ATs have also been shown to perform “proof-reading”, i.e. if the AT is presented with an incorrect extender unit, it behaves as a hydrolase as opposed to a transferase and corrects the mistake (Bonnett et al., 2011).

1.3.4.3 Ketosynthase (KS) domain and the extension chemistry

The approximately 340 residue-containing KSs are the most fascinating parts of polyketide synthases. These homo-dimeric enzymes are highly conserved within FASs and PKSs and are responsible for constructing the carbon skeletons of all complex polyketide natural products. The architecture of KS comprises a thiolase (α/β/α/β/α) fold, and both active sites of the homo-dimer are connected via a tunnel at the two-fold axis (Huang et al., 1998). The architecture of the DEBS module 5 ketosynthase domain fused to its cognate acyltransferase domain is shown in Figure 1-9 (PDB code 2HG4) (Tang et al., 2006).

The KS is connected to the AT domain via a ~140-residue linker comprising a ~90-residue folded linker domain and ~50-residue loop. These linkers are conserved between FASs and PKSs and have been shown to be important in orientating the AT correctly with respect to the KS (Maier et al., 2008).

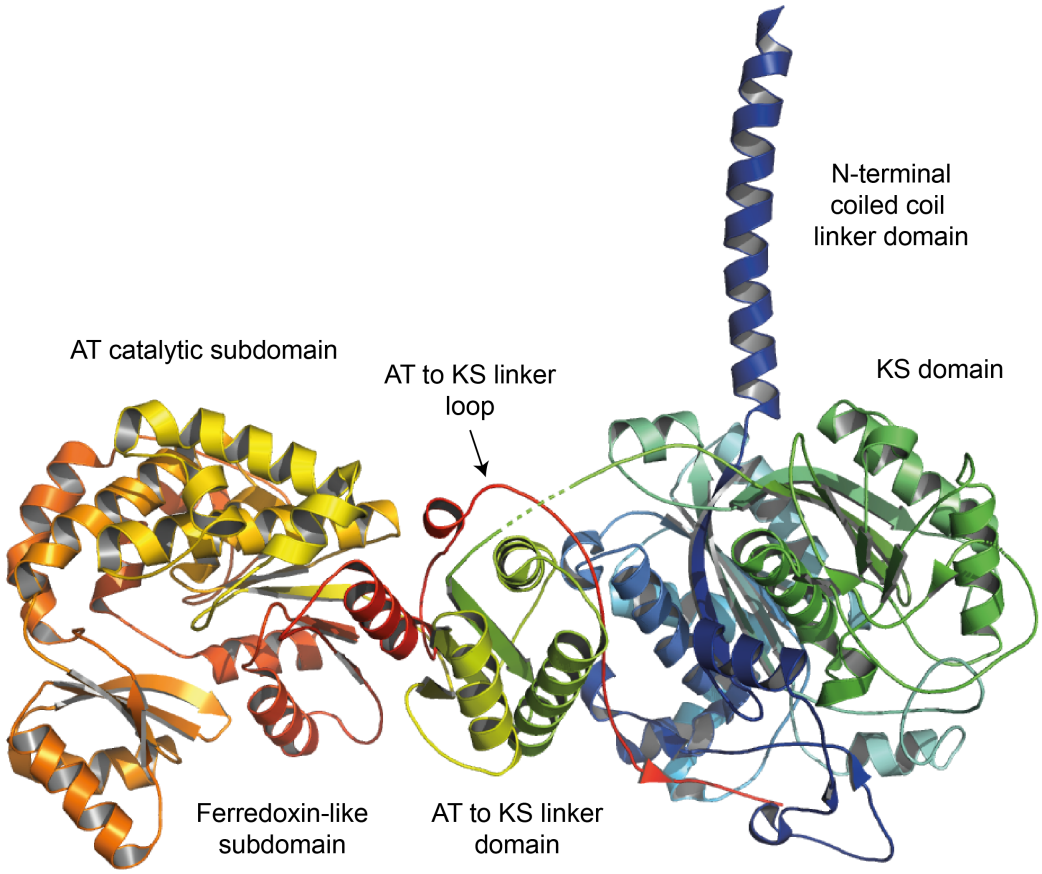


Figure 1-9 Crystal structure of the ketosynthase-acyltransferase didomain from DEBS module 5 (PDB code 2HG4). *The AT domain comprises catalytic and ferredoxin-like subdomains, the latter of which is involved in binding to the ACP. The KS is linked to the AT via a folded 90-residue linker and a 50-residue loop. Both of the linkers are thought to play a role in positioning the AT in the right orientation with regards to the KS. Finally, the N-terminal coiled-coil linker domain is involved in binding to the adjacent module (Tang et al., 2006).*

The active site of KS contains a well-characterised catalytic triad of Cys-His-His (Wettstein-Knowles et al., 2006). The catalytic cycle (Figure 1-10) begins with a *trans*-thioesterification, during which the catalytic triad cysteine acts as a nucleophile and attacks the thioester of the polyketide intermediate from the preceding module. Subsequently, the extender unit, covalently bound to the intramodular ACP is decarboxylated to form a carbanion intermediate. This carbanion intermediate then attacks the polyketide intermediate bound to the KS giving rise to the elongated polyketide chain.

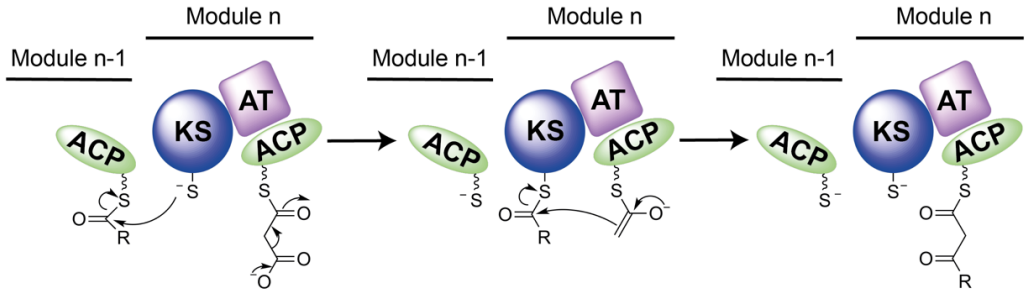


***Figure 1-10 The catalytic cycle of the KS**. First, the KS attacks the polyketide intermediate from the preceding module and decarboxylates the extender unit bound to the intramodular ACP to form a carbanion. This carbanion then attacks the polyketide intermediate bound to the KS giving rise to the elongated polyketide chain (Wettstein-Knowles et al., 2006; Fischbach and Walsh, 2006).*

Of the two histidines present in the catalytic triad, one has been shown to be essential for the decarboxylation of the extender unit, while the other is thought to act as a general base (Witkowski et al., 2002; Keatinge-Clay et al., 2004; Zhang et al., 2006). The chain extension reaction has been shown to invert the configuration at C2 of the extender unit moiety (Weissman et al., 1997). KSs have also been shown to be selective for their substrates and, if presented with a non-natural polyketide intermediate, will perform the catalysis reaction very slowly, if at all (Watanabe et al., 2003; Wu et al., 2004).

1.3.4.4 Processing enzymes

Some modules contain additional domains, known as “processing” domains, which are responsible for modifying the β-keto group of the growing intermediate.

Ketoreductase (KR) domain

The size of the KR domain is variable depending on which module the KR is found at: β-module KR contains ~460 residues, and KR found in ɣ- and δ-modules comprises ~420 residues (Keatinge-Clay, 2012). What is more, KRs from β-modules (which do not contain any other processing enzymes) contain an additional ~60-residue dimerisation element on the C-terminus, which is hypothesised to be involved in the dimerisation of the whole assembly line (Zheng et al., 2010).

Primarily, KR utilises NADPH to stereoselectively reduce the β-keto group of the β-ketoacyl intermediate to produce a hydroxyl group. There are three types of KRs: A-type generates a hydroxyl group with an $_{L}$-orientation, B-type produces a

hydroxyl with a $_D$-orientation, and C-type is incapable of reduction (Keatinge-Clay, 2007). Interestingly, some KRs can switch between types A and B depending on the nature of the intermediate they are processing (Zhou et al., 2012).

KRs are comprised of two subdomains: the N-terminal structural domain (denoted as KR_s) and the C-terminal catalytic domain (denoted as KR_c), both of which exhibit a Rossmann-like fold found in the NADPH-binding enzymes (Keatinge-Clay and Stroud, 2006). It is thought that the KR_s domain's primary responsibility is to stabilise the KR_c, enabling the entire KR to be positioned away from the two-fold axis of the polyketide synthase. Crystal structures for KR types A and B have been determined, PDB codes 3MJS (Zheng et al., 2010) and 2Z5L (Keatinge-Clay and Stroud, 2006), respectively. The architecture of the DEBS module 1 KR domain is shown in Figure 1-11 (PDB code 2HG4) (Tang et al., 2006).

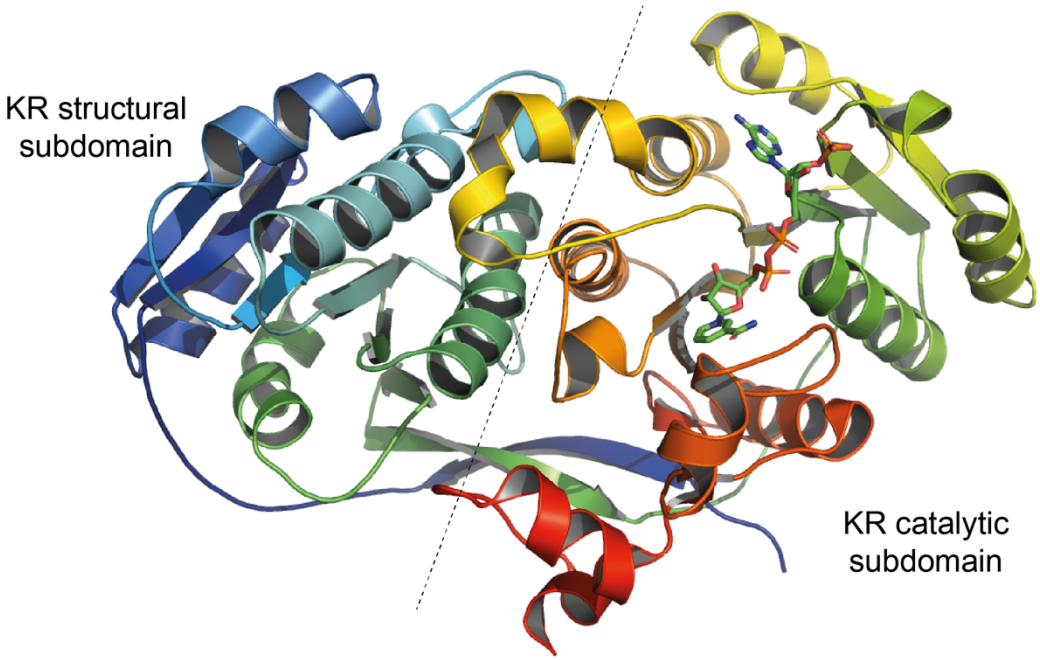


Figure 1-11 The crystal structure of the ketoreductase domain from DEBS module 1 (PDB code 2FR0) (Castonguay et al., 2007). *KR is comprised of two subdomains: structural and catalytic (separated by a dashed line). The catalytic subdomain contains a bound NADPH, which is shown by stick representation.*

Dehydratase (DH) domain

The DH monomers, each comprised of ~280 residues, make contact across the two-fold axis of the polyketide synthase and hence play a role in the dimerisation of the full assembly line (Keatinge-Clay, 2012). The majority of DH domains catalyse the dehydration of the polyketide intermediate to yield a *trans*- double bond between the α- and β- carbons; however, some DHs have been shown to form *cis*- double bonds (Gay et al., 2013). What is more, some DHs also possess isomerase activity which shifts the double bond from the α-/β- carbon to the β-/ɣ- carbon (Kusebauch et al., 2010). Unfortunately, no significant sequence differences have been identified between different types of DHs, and therefore bioinformatic analysis to predict whether the DH will form a *cis*- or a *trans*- double bond and whether it possesses epimerase activity is not yet possible.

DH domains exhibit a double hot dog motif and dimerise along the 25-residue stretch of the N-terminus. Many crystal structures of polyketide synthase DHs have been determined, including DEBS DH, PDB code 3EL6 (Keatinge-Clay, 2008), curacin synthase DH, PDB code 3KG6 (Akey et al., 2010) and rifamycin synthase, PDB code 4LN9 (Gay et al., 2013). The architecture of the DEBS module 4 DH domain is shown in Figure 1-12 (Keatinge-Clay, 2008).

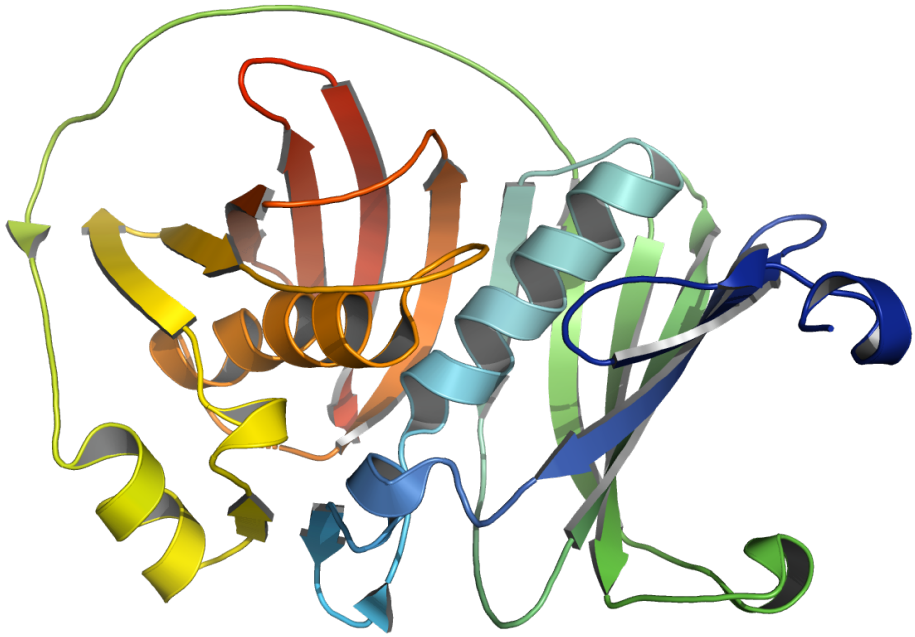


Figure 1-12 Crystal structure of the DEBS module 4 dehydratase domain (PDB code 3EL6) (Keatinge-Clay, 2008). *This domain exhibits a double hot dog fold and catalyses the dehydration of the polyketide intermediate.*

Enoyl Reductase (ER) domain

The ~310-residue ER domains are the least understood parts of polyketide elongation. The fold of ERs is characteristic of the medium-chain dehydrogenase/reductase (MDR) enzyme superfamily, however, differences between the ERs from different PKSs exist (Persson et al., 2008). Within modular PKSs, the ER domain is located in a loop connecting the catalytic and structural subdomains of the KR. Analysis of the oligomeric state of many ERs suggests that a monomeric form of ER evolved from the dimeric form to enable intramodular ACPs to reach their cognate domains (Keatinge-Clay, 2012; Zheng et al., 2012). The active site of ER is located in a cleft between core cofactor (NADPH)-binding and substrate-binding (or catalytic) subdomains. ERs utilise NADPH to stereoselectively reduce *trans*-α,β-double bonds formed by DH (Keatinge-Clay, 2012).

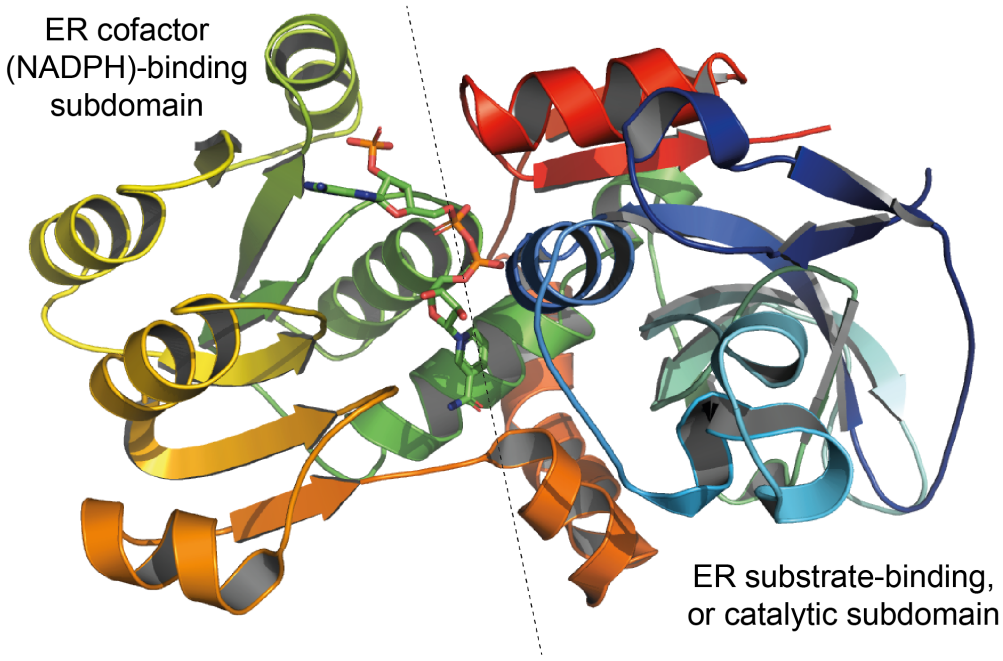


Figure 1-13 Crystal structure of CurF enoyl reductase from curacin A biosynthetic pathway (PDB code 5DP2) (Khare et al., 2015). *ER belongs to the MDR superfamily of enzymes and catalyses the reduction of the trans-α,β-double bonds in polyketide intermediates. ERs are comprised of substrate binding and cofactor-binding subdomains (separated by a dashed line). The cofactor-binding subdomain contains a bound NADPH, which is shown by stick representation.*

Several ERs have been characterised by X-ray crystallography, including curacin A synthase ER, PDB code 5DP2 (Khare et al., 2015), fungal fatty acid synthase ER, PDB code 4CW5 (Bukhari et al., 2014) and spinosyn synthase ER, PDB code 3SLK (Zheng et al., 2012). The architecture of CurF ER from the curacin A biosynthetic pathway is shown in Figure 1-13 (PDB code 5DP2) (Khare et al., 2015).

1.3.4.5 Chain initiation and termination

Chain initiation by a dedicated loading module (LM)

Each polyketide synthase comprises a dedicated loading module responsible for loading a polyketide starter unit onto the assembly line. The LM always contains AT and ACP domains and sometimes a KS domain as well (Fischbach and Walsh, 2006). Loading ATs and ACPs are homologous to their chain-extension counterparts and possess the same fold and catalytic machinery. Loading KSs are highly related to the elongating KSs as well, but they are referred to as “KSQ”, due to the catalytic cysteine being replaced with a glutamine residue. KSQs are thus catalytically-inactive in terms of the condensation reaction; however, the glutamine catalyses the decarboxylation of the extender unit to form a starting acyl group (Keatinge-Clay, 2012). In the absence of the KSQ domain, chain initiation is carried out by a loading acyltransferase which selects the starter unit and loads it onto the LM ACP (Wilkinson et al., 2001).

Chain termination by a thioesterase (TE) domain

A ~240-290 TE domain is found at the C-terminus of many PKSs. The modular type I PKS TE also contains an additional ~35-residue N-terminal helix which is thought to be involved in the dimerisation of the assembly line (Keatinge-Clay, 2012). The TE comprises an α/β-hydrolase catalytic core and loop regions which form a substrate-binding lid. A number of TE domains have been structurally characterised, including TE from DEBS, PDB code 1KEZ (Tsai et al., 2001), pikromycin synthase, PDB code 1MNA (Tsai et al., 2002) and tautomycetin synthase, PDB code 3LCR (Scaglione et al., 2010). The architecture of the DEBS thioesterase domain (PDB code 1KEZ) is shown in Figure 1-14 (Tsai et al., 2001).

The primary function of a TE domain is linear polyketide product release from the synthase, however, in some polyketide biosynthesis pathways (such as

erythromycin and pikromycin), the TE is also responsible for the intramolecular cyclisation reaction to create a macrolactone ring (Tsai et al., 2002; Scaglione et al., 2010). This occurs in a large active site chamber where the bound polyketide cyclises into a ring structure (Akey et al., 2006; Tsai et al., 2002).

In DEBS TE, the substrate channel (Figure 1-14) does not appear broad enough to accommodate a circular DEBS product; however, analysis of the channel interior revealed a wide cavity around the catalytic residues. This cavity was calculated to have a diameter of approximately 15 Å and be big enough to accommodate the 14-membered macrolide ring without changing the conformation of the enzyme (Tsai et al., 2001).

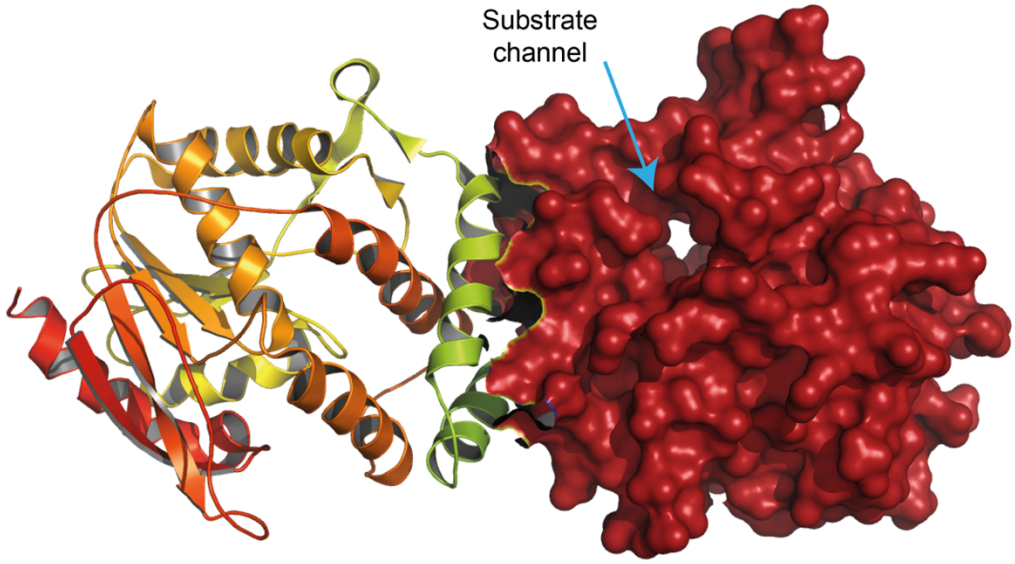


Figure 1-14 Crystal structure of the DEBS thioesterase dimer (PDB code 1KEZ) (Tsai et al., 2001). *TE comprises an α/β-hydrolase catalytic core and loop regions which are involved in substrate-binding. The roles of TE involve substrate cyclisation and release from the synthase. Here, chain A is shown as a cartoon representation, and chain B is shown as a surface representation. The substrate channel is indicated by an arrow.*

1.3.4.6 Post-PKS tailoring enzymes

Either following the release from the synthase, or while still tethered to one of the ACPs, polyketide intermediates can undergo a wide variety of modifications to further alter their functionality. These modifications are often essential for the polyketide's biological activity and are performed by dedicated enzymes encoded within the same BGC (Fischbach and Walsh, 2006). Common post-PKS

polyketide tailoring reactions include oxidation to introduce hydroxy or carbonyl groups, methylation at oxygen, nitrogen or carbon centres, cyclisation, glycosylation, acylation and alkylation (Staunton and Weissman, 2001). Mixing and matching this stunning variety of polyketide processing enzymes presents an enormous opportunity to produce novel natural product structures (Gaisser and Leadlay, 1998).

1.4 Generating the diversity in polyketide natural products

Natural products have proved to be a matchless starting point for drug discovery. During the last three decades, the majority of antimicrobials and anticancer agents entering clinical trials were based on natural products (Goss et al., 2012; Newman and Cragg, 2012). Advances in next-generation sequencing and various bioinformatics tools are aiding the discovery of many novel PKS genes without a need for traditional bioactivity-guided screens (Helfrich et al., 2014). Such genome mining gives access to a myriad of new starting points for drug discovery. However, the physicochemical properties, for example solubility, stability or permeability, of such natural products are often poor. Such difficulties, paired up with rapid development of drug-resistance, has lead us to a significant and urgent need for development of existing natural product analogue libraries, which can then be screened for improved pharmaceutical properties (Goss et al., 2012).

1.4.1 Hybrid approaches

Several strategies have been successfully employed to generate polyketide natural product analogues. A traditional approach would be to use total synthesis. However, the complexity and number of chiral centres present in complex polyketides, combined with poor yields, make this method of generating diversity far from ideal.

Increasing attention is placed on hybrid methods combining biological and chemical approaches. For example, in semi-synthesis, the polyketide is produced by a native organism, extracted and purified, before is it modified synthetically. One of the key benchmarks in this area was the work performed on rapamycin, which resulted in the increased potency of this therapeutic (Ruan et al., 2008). However, the number and complexity of the functionalisable groups in the natural

products available for the modification using semi-synthesis are often very limited.

In precursor-directed biosynthesis, an organism producing the natural product is fed a synthetic substrate leading to the production of both the natural and the unnatural products. A great example of this approach is a study by Khosla *et al.* which led to the discovery of epothilone C, a novel anticancer reagent (Boddy et al., 2004). The disadvantages of this method include limited substrate flexibility and the need to separate the end-products chromatographically (Goss et al., 2012).

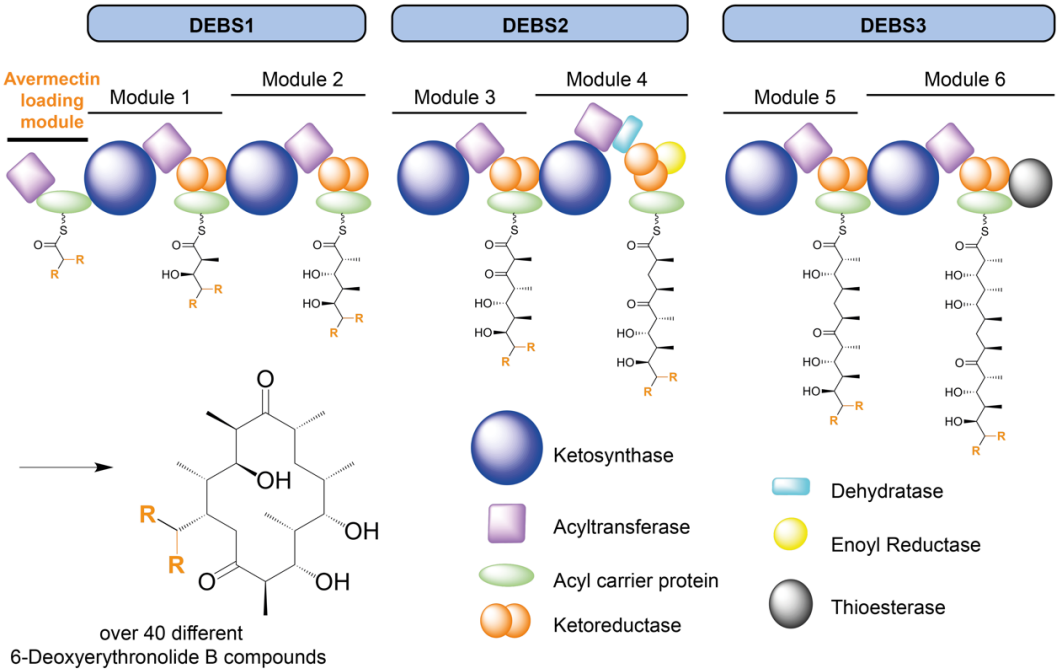


Figure 1-15 An example of mutasynthesis approach used to generate polyketide diversity. *The loading module from 6-deoxyerythronolide B synthase (DEBS) was replaced with a more substrate-flexible loading module from the avermectin PKS which resulted in the generation of over forty novel macrolide derivatives (Marsden et al., 1998).*

Mutasynthesis involves modifying megasynthase enzymes through either gene disruption or gene exchange. Following genetic modification, a synthetic substrate or intermediate is fed into the system (Goss et al., 2012; Gupta et al., 2008). An elegant example of gene exchange approach was the study by Leadlay and colleagues at Cambridge which involved replacing the loading module of DEBS with the more substrate-flexible loading module from the avermectin PKS

(Figure 1-15). This allowed the incorporation of numerous branched starter units to produce over forty novel erythromycin derivatives (Goss and Hong, 2005; Marsden et al., 1998).

The mutasynthesis approach has also been successfully used to gain insights into the structure-activity relationships of the polyketide geldanamycin, a potent heat shock protein 90 (Hsp90) inhibitor. In one particular study, a 3-amino-5-hydroxybenzoic acid (AHBA)-blocked mutant of *S. hygroscopicus*, the natural producer of geldanamycin, was fed aromatic and heteroaromatic amino acids which led to the generation of a new geldanamycin analogue with improved properties (Eichner et al., 2009).

Unfortunately, the success of mutasynthesis-employing studies varies significantly between different systems, and, therefore, combinatorial biosynthesis is regarded as a more promising approach to generate diversity in the existing natural products.

1.4.2 Combinatorial biosynthesis

By definition, combinatorial biosynthesis involves interchanging secondary metabolism genes between different BGCs to generate novel natural products. The main difference of combinatorial biosynthesis from the hybrid approaches discussed in the previous section is that it must yield an organism capable of natural product synthesis “from scratch”, i.e. without the need to supplement the system with the additional precursor or extender units.

Since the discovery of PKSs in the early 1990s and figuring out the modular arrangement, as well as the correlation between the domain organisation and the structure of the polyketide product, scientists deemed that these systems could be rationally engineered in a mix-and-match approach (Wong and Khosla, 2012). Over the years, engineering efforts could be classified into two different branches. The first involves the introduction of genes from other pathways which will operate either prior to (generation of starting units) or following (tailoring of the polyketide product) the biosynthetic assembly line. The second branch involves mixing-and-matching of PKS domains and modules from either the same, or different, biosynthetic pathways (Goss et al., 2012).

Introducing post-PKS tailoring enzymes to characterised biosynthetic pathways has led to the generation of hundreds of novel polyketide compounds (Olano et

al., 2010). For example, complementing *S. hygroscopicus* with different methyltransferase-encoding genes resulted in a number of rapamycin derivatives (Gregory et al., 2006).

DEBS is probably one of the most engineered systems using combinatorial approaches. One of the most famous examples of this comes from the early studies led by Leadlay's group, which showed that the chain length of the DEBS product could be successfully shortened by placing a DEBS TE domain at the C-terminus of DEBS1 subunit (Cortes et al., 1995; Kao et al., 1995). The generation of this construct, termed DEBS1-TE, is shown in Figure 1-16.

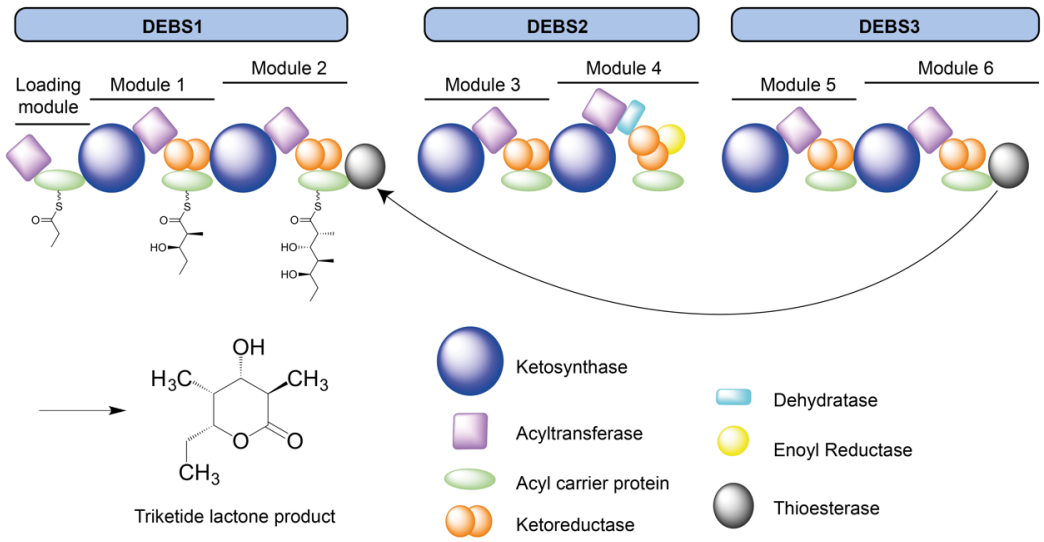


Figure 1-16 Construction of triketide lactone synthase (DEBS1-TE) from DEBS (Cortes et al., 1995). *The DEBS3 terminal thioesterase domain was repositioned to the C-terminus of the DEBS1 subunit. This resulted in the generation of a triketide lactone product instead of the full-length 6-deB.*

In another study by the Leadlay group, the length of the native polyketide erythromycin was successfully extended by introducing an additional module (Rowe et al., 2001). In this study, a module from the rapamycin-producing polyketide synthase was covalently inserted between the two chain-elongating modules of DEBS1, which resulted in the generation of the anticipated octaketide macrolactone products. Unexpectedly, the production of the extended macrolactones was accompanied by significant amounts of the natural product of DEBS, 6-deB, which was a sign of a module skipping action by the synthase.

Module skipping as well as stuttering, in which a single module performs one or more rounds of chain elongation iteratively, has been observed in other studies

of the DEBS system as well (Wilkinson et al., 2000; Thomas et al., 2002). Unfortunately, due to the lack of systematic studies on these phenomena and on the rules governing the linearity of PKS assembly lines, it is currently not possible to predict or control the skipping and stuttering (Khosla et al., 2007).

1.4.3 Challenges in combinatorial biosynthesis

Even though the feasibility of combinatorial biosynthesis has been shown by several successful studies dating from the 90s (McDaniel et al., 1999; Tang et al., 2004; McDaniel et al., 1995), many challenges need to be addressed to make the generation of novel natural products efficient and worthwhile (Poust et al., 2014).

Firstly, there are issues associated with the size and the large number of PKS genes. Next generation sequencing and bioinformatic tools gave rise to a genome mining approach, which made the identification of PKS clusters, especially those from unculturable marine organisms, more accessible than ever (M. H. Li et al., 2009; Nett, 2014). However, the lack of tools for automated assembly of large (20 - 100 kb) gene contigs from complex GC-rich metagenomic samples makes it hard to study these systems effectively.

Secondly, domain swapping to alter specificity often results in a significant kinetic penalty; and, therefore, in order to enable rational engineering of PKSs, a better understanding of enzyme specificity and the forces driving the unidirectional movement of the substrate along the PKS production line is needed (Wong and Khosla, 2012).

Finally, obtaining high-resolution structural information of the whole PKS complexes is of extreme importance (Weissman, 2015b). Previous efforts of engineering chimeric PKS assembly lines were based on the assumptions that the individual domains within the PKS structure do not depend on the neighbouring domains and can be interchanged. However, the majority of rational engineering efforts result in dysfunctional PKS lines, which either do not produce a polyketide, or produce alternative, truncated, products. Now it is known that intermodular linker pairs are essential for the generation of active PKS lines (Del Vecchio et al., 2003; Weissman, 2015b; Tang et al., 2006) and, therefore, studying complete PKS systems with the linker areas intact, as well as designing accurate constructs, are of paramount importance.

1.5 Structural studies of whole polyketide synthase modules

As discussed in section 1.3.4, the individual domains from type I *cis*- and *trans*-AT polyketide synthases have been successfully characterised by X-ray crystallography and solution NMR spectroscopy techniques (Lundqvist et al., 1994; Tsai et al., 2001; Liew et al., 2012; Tang et al., 2006; Keatinge-Clay, 2008; Park et al., 2014; Gay et al., 2013; Robbins et al., 2016).

Until 2014, there was no structural data available on intact polyketide synthase modules, and therefore the understanding of the architecture of these molecular machines was based on a PKS "relative", fatty acid synthase (FAS). In 2014, the situation improved drastically when the first modular structures from pikromycin, erythromycin and virginiamycin synthases were characterised by cryo-electron microscopy (cryo-EM) and small-angle X-ray scattering (SAXS) (Leadlay, 2014; Whicher et al., 2014; Dutta et al., 2014; Tang et al., 2007). Since then, a structure of a fully-reducing fungal iPKS module was also reported (Herbst et al., 2016).

The progress towards visualising intact PKS modules, starting with the model of FAS, will be discussed in this section. However, even today, a three-dimensional structure of a whole PKS assembly line remains an enigma.

1.5.1 Fatty acid synthase – a model for polyketide synthase module arrangement

The mammalian FAS (mFAS) comprises a single module with the same six catalytic domains as a fully-reducing type I PKS module (KS, AT, ACP, KR, DH and ER) and therefore could essentially be regarded as a type I iPKS responsible for the production of fatty acids. The 4.5 Å crystal structure of the mFAS was first reported in 2006 by Maier et al. (Maier et al., 2006). The same group published a higher resolution, 3.2 Å, re-refined model of mFAS in 2008 (Maier et al., 2008).

The homodimer of mFAS was shown to exhibit an X-shape architecture with the bottom condensing region containing the domains responsible for chain elongation (KS and AT) and the upper modifying region containing all processing domains (KR, DH and ER) (Figure 1-17). The structure also revealed two catalytically-inactive pseudo domains: a ketoreductase (KR^x) and a methyltransferase (ME^x) (Maier et al., 2008). Neither the ACP nor the TE domains were resolved in the structure, which is not surprising given the distance the ACP

needs to travel between different active sites of the synthase (Weissman, 2015b). The extraordinary flexibility and continuous movements within the FAS were indeed observed using negative stain electron microscopy (negative stain EM) and published the following year (Brignole et al., 2009).

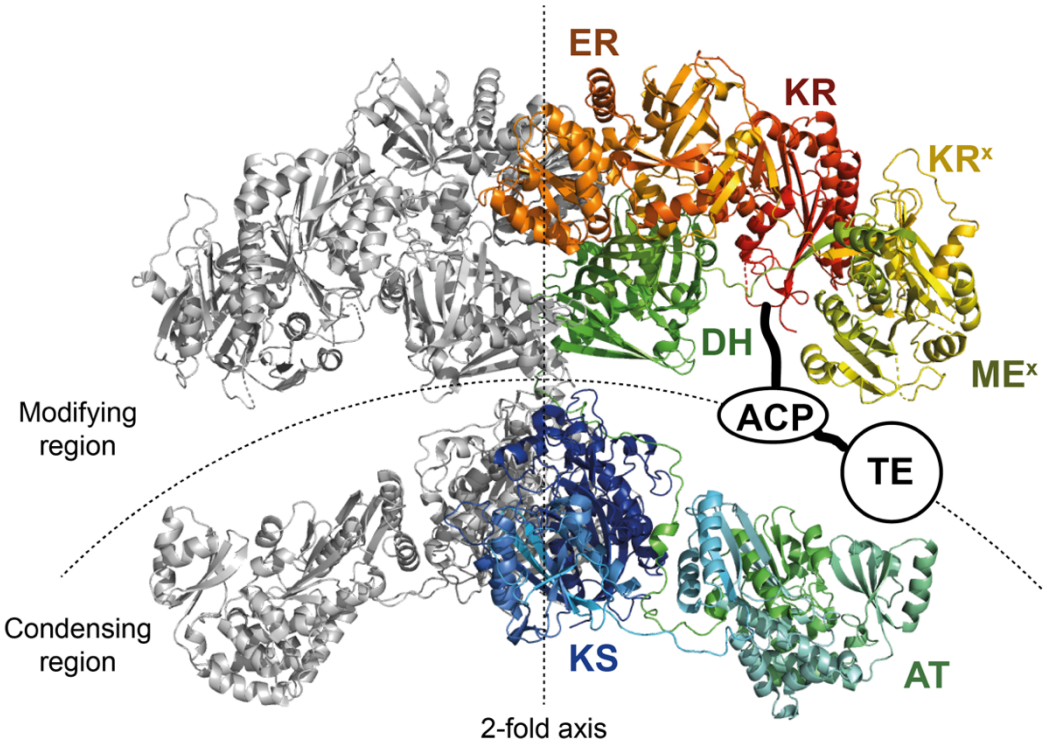


Figure 1-17 The architecture of the mFAS. *The mFAS comprises two monomers (the two-fold axis shown as a straight dashed line), with one monomer shown in grey and the other one coloured according to the domains it comprises. The whole complex can be divided into two regions: the lower condensing region comprising the KS and AT domains, and the upper modifying region, comprising all other domains. In total five catalytic (KS, AT, DH, ER and KR) and two pseudo-non-catalytic (KRx and MEx) domains were resolved. Both the ACP and TE domains (illustrated as white circles at their predicted location) were present in the construct but were not resolved in the crystal structure. The figure was made using PDB entry 2VZ8 (Maier et al., 2008).*

Due to the domain sequence homology and the atomic detail revealed by the crystal structure, mFAS was regarded as a highly-relevant model for the structure of modular polyketide synthases. Indeed, the comparison between DEBS KS-AT didomain (Figure 1-9) and mFAS (Figure 1-17) crystal structures revealed unprecedented similarities in the arrangement and the angles between the KS and AT domains (Maier et al., 2008; Weissman, 2015b; Tang et al., 2006).

It is worth noting, however, that the comparison of the mFAS structure to the structures of the individual PKS domains revealed some notable differences. For example, PKS TE domains are homodimeric (Tsai et al., 2002), while the FAS TE domains are monomeric (Chakravarty et al., 2004) (which is not surprising since the predicted location of mFAS TE domains is on the opposite ends of the mFAS dimer). The difference in the oligomeric states was also observed in ER domains: FAS ER domains are homodimeric (Maier et al., 2008) and PKS ERs have been shown to be monomeric (Zheng et al., 2012).

1.5.2 Cryo-EM analysis of PikAIII reveals a dramatically different architecture from the mFAS model

In 2014, Skiniotis, Smith, Sherman and colleagues published a number of cryo-EM structures of module 5 (known as PikAIII) from the pikromycin PKS (EMDB accession codes 5647, 5648, 5649, 5651, 5653 and 5662) (Dutta et al., 2014; Whicher et al., 2014). PikAIII is a distinct module comprising three essential domains (AT_5, ACP_5 and KS_5), one accessory (KR_5) domain as well as N-terminal docking and post-ACP dimerisation helices. The global resolution of PikAIII maps in different states was calculated to be 7.3-12.5 Å which allowed the fitting of the homologous domains from DEBS (Tang et al., 2006; Tsai et al., 2001; Keatinge-Clay, 2008; Castonguay et al., 2007) to create pseudo-atomic models of *holo*-PikAIII and its several mutants.

Two distinct conformers were observed for the *holo*-PikAIII (Figure 1-18). Both cryo-EM reconstructions revealed a drastically different architecture compared to previous studies of mFAS and the double helical “Cambridge model” proposed for the erythromycin PKS (Goss et al., 2012; Staunton and Weissman, 2001; Weissman, 2015a). Previous models predicted PKS to contain two isolated reaction chambers (one for each of the two monomers), however, the overall shape of PikAIII can be described as a horseshoe, or a symmetrical arch, containing a single reaction chamber in the middle. In the first conformer of PikAIII, representing 57% of the data, the ACP was found at the bottom of the arch, forming direct interactions with the KR domain. In the second conformer, representing 43% of the data, the ACP was found at the top of the arch, close to both KS and AT domains.

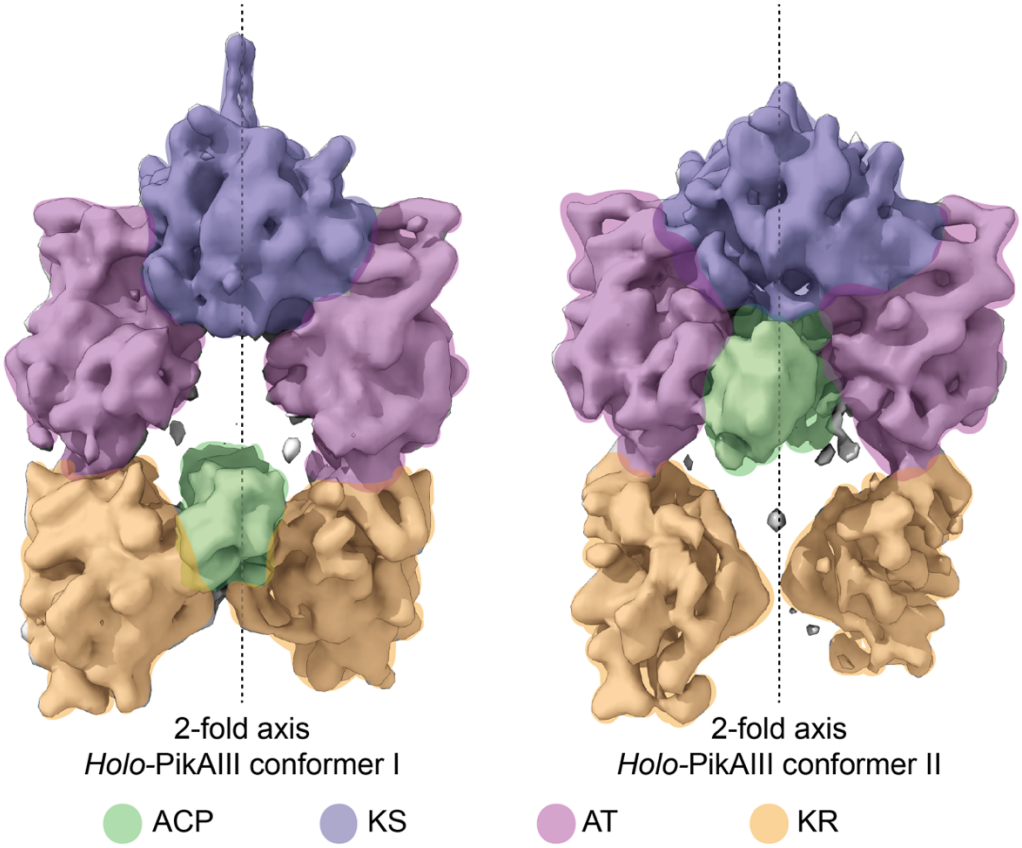


Figure 1-18 Cryo-EM analysis of holo-PikAIII module from pikromycin polyketide synthase (Dutta et al., 2014). *Left: cryo-EM map of conformer I, in which the ACP (green) is located at the bottom of the arch, bound to the KR domain (found in 57% of data). Right: cryo-EM map of conformer II, in which the ACP (green) is located at the top of the arch, bound to the AT domain (found in 43% of data). The figure was made using EMDB entries 5647 (conformer I) and 5648 (conformer II).*

While radically different from the mFAS, the arch-like shape of the PikAIII was regarded as highly plausible since the ACP would have to travel shorter distance between the different catalytic sites within the module. Nonetheless, several drastic differences from previously determined data raised some concerns regarding the PikAIII models. First, the crystal structure of the DEBS KS-AT didomain revealed an extended linear arrangement (Figure 1-9) (Tang et al., 2006), which was in agreement with both crystallographic and negative stain EM data of mFAS (Maier et al., 2008; Brignole et al., 2009). In contrast, the AT domain in PikAIII EM maps was found to be rotated a full 120° relative to the previously observed position. As a result of this, the interface between the KS and AT is too small in PikAIII models to accommodate a KS-to-AT linker domain observed previously.

The second issue comes from the steric arrangement of the ACPs. While the ACPs are clearly visible in the EM maps of both conformations, modelling of homologous ACP structures into the EM density revealed that the conserved serine of the ACP (which is post-translationally modified to contain a phosphopantetheinyl group to serve as an attachment point to the growing polyketide) is facing away from both the AT and the KR domains. As a result, at least in the absence of a substrate, productive interactions between the ACP and these domains are not likely (Weissman, 2015b).

Finally, some questions regarding the dual functionality of the KS domain were raised. The KS performs two distinct reactions: transesterification of the polyketide intermediate from the upstream module (which results in the transfer of this intermediate onto the KS catalytic cysteine) and chain extension between the cysteine-tethered intermediate and an extender unit tethered to its cognate ACP. Therefore, the KS has to discriminate between the two ACPs it interacts with in order to avoid the mis-transfer of the catalysed elongated substrate back onto itself leading to an unwanted iteration. The old models, created using mFAS and DEBS KS-AT didomain crystal structures (Maier et al., 2008; Tang et al., 2006), proposed that both ACPs have their own distinct binding sites in a deep cleft at the opposite sides of the KS-to-AT linker domain, which explains how KS distinguishes between the ACPs. However, this model does not agree with the PikAIII data reported by Dutta and colleagues (Dutta et al., 2014).

To investigate how PikAIII KS distinguishes between the two ACPs, the PikAIII variant lacking ACP_5 but linked with the ACP_4 was generated and analysed using cryo-EM. The resulting cryo-EM map revealed that the ACP_4 sits at the side entrance to the KS_5 outside the reaction chamber, and is located nearly 30 Å away from the KS the active site. These data established a distinctive structural basis for the KS discrimination between the upstream and downstream ACPs (Dutta et al., 2014); however, the mechanism of the upstream ACP travelling the 30 Å distance to the KS active site remains unclear.

In an accompanying paper, by Whicher et al. (Whicher et al., 2014), the PikAIII catalytic cycle was characterised by obtaining a number of cryo-EM “snapshots”. Interestingly, the data suggested that the ACP performs “random walking” within the catalytic chamber and the active sites of different domains engage with the ACP once they recognise their native substrate. What is more, the same findings

suggested that the ACP only leaves the catalytic chamber once all the processing is complete, which explains how a broad-specificity TE domain (Aggarwal et al., 1995), located outside the catalytic chamber, does not prematurely off-load the polyketide intermediates.

1.5.3 Lower resolution structural characterisation of DEBS3 and VirA module 5 by SAXS

Prior to the PikAIII studies, the Khosla laboratory reported SAXS structural data of DEBS modules 5 and 6 (termed DEBS3) and DEBS3-TE (Edwards et al., 2014). Even though of lower resolution than PikAIII, these 40-50 Å models also allowed the rigid-body fitting of existing high-resolution crystal structures of individual DEBS domains (Tang et al., 2006; Tsai et al., 2001; Keatinge-Clay, 2008; Castonguay et al., 2007).

DEBS

The DEBS3-TE molecular envelopes calculated from the experimental data, similarly to PikAIII, revealed multiple configuration states (Figure 1-19a); however, the overall arrangement of the domains was not consistent with the PikAIII models. For example, the KS-AT didomain retained its extended form, as was observed in mFAS and DEBS KS-AT crystal structures. What is more, the ACP was observed near the TE domain, rather than docked at the KS or the KR domains, like in PikAIII (Dutta et al., 2014). The discrepancies observed could have resulted from either the low-resolution nature of the SAXS technique, from rigid body restraints or, more intriguingly, from fundamental differences that exist between the different PKS systems (Weissman, 2015b).

The molecular envelopes calculated for the DEBS3 construct (missing the TE domain) were fundamentally analogous to the ones calculated for the DEBS3-TE. The main difference observed was that the modules 5 and 6 were rotated 70° compared to their counterparts in the DEBS3-TE calculations. These observations suggest that in the absence of specific dimerisation motifs, docking regions or dimeric terminal proteins (such as TE) even covalently-linked modules do not possess strong intermodule interactions.

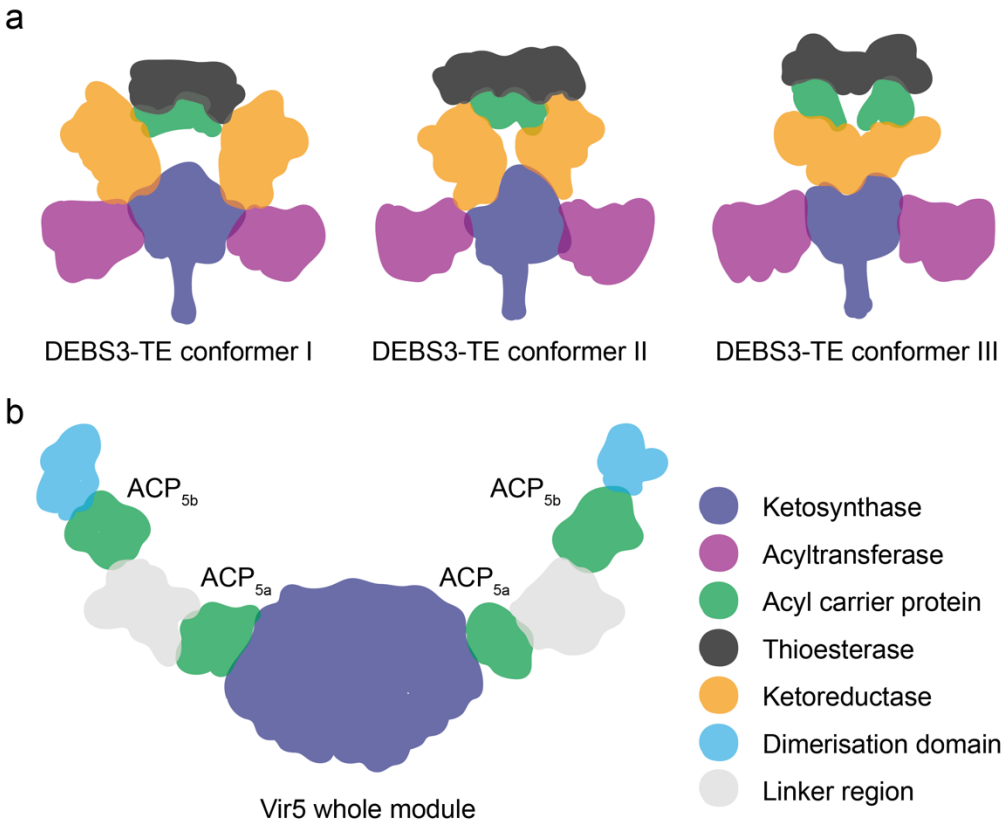


Figure 1-19 Analysis of DEBS3-TE and Vir5 by SAXS. *(a) Three conformers of DEBS3-TE derived from modelling of the high-resolution homology models into the ~50 Å molecular envelopes produced by SAXS. In all three cases, the KS-AT was kept as a rigid body (Edwards et al., 2014). (b) Model of the Vir5 whole module derived by superposition of the homology models into the ~20 Å Vir5 molecular envelope calculated by SAXS (Davison et al., 2014).*

VirA

Davison and colleagues employed a similar SAXS approach to study VirA, the fifth module of *trans*-AT modular PKS responsible for the synthesis of the antibiotic virginiamycin. Each monomer of the dimeric VirA module 5 comprises one KS and two ACP domains, with all three parts of this minimal module connected with linkers of variable length. This study involved a divide-and-conquer approach, whereby this module was characterised as a whole, and as the smaller fragments it comprises (KS+linker, ACP-ACP, KS-ACP-ACP and KS-ACP-ACP-DD, where DD denotes dimerisation domain), and a number of ~ 20 Å molecular envelopes were calculated (Davison et al., 2014).

The molecular envelopes of smaller fragments were in agreement to the one of the whole module and revealed an open, arch-like architecture (Figure 1-19b),

like the one seen in PikAIII (Figure 1-18). However, the two arms of the Vir5 arch-like structure do not come back together via the post-ACP dimerisation domains, like they do in PikAIII structures. One of the two ACPs (termed ACP_{5a}) was revealed to be tucked near the KS domain and postulated to be involved in the extension chemistry. The second ACP (termed ACP_{5b}) was shown to be located at the very end of the extended arm and therefore postulated to be involved in the modification of the β-centre as well as the transfer of the growing polyketide intermediate to the downstream module.

One of the concerns with the proposed model includes the conserved serine of the ACP_{5a} being located at least 50 Å away from the active site of the KS, and this distance being too large to permit the interaction. It is also worth noting that the position of the KS and its subdomain (analogous to the KS-to-AT linker domain in *cis*-AT PKSs) was significantly different from the one seen in the PikAIII model. It is, of course, a possibility that these discrepancies arose from visualising the system in its *apo*- form, and this hypothesis could be tested by studying the module in the presence of its native substrate.

1.5.4 Hybrid crystal structure of the iterative mycocerosic acid synthase-like PKS

Type I iterative PKSs comprise the same catalytic domains and are closely related to type I mPKSs. Following the success with mFAS, Maier and colleagues investigated the three-dimensional architecture of the first fully-reducing fungal iPKS from *Mycobacterium smegmatis* (Herbst et al., 2016). The PKS was initially identified as a mycocerosic acid synthase; however, was later renamed as mycocerosic acid synthase-like (MAS-like) PKS (Etienne et al., 2009). To determine the structure of the whole module, the complex was split into two parts, corresponding to its condensing and modifying regions, and analysed using X-ray crystallography independently (Figure 1-20).

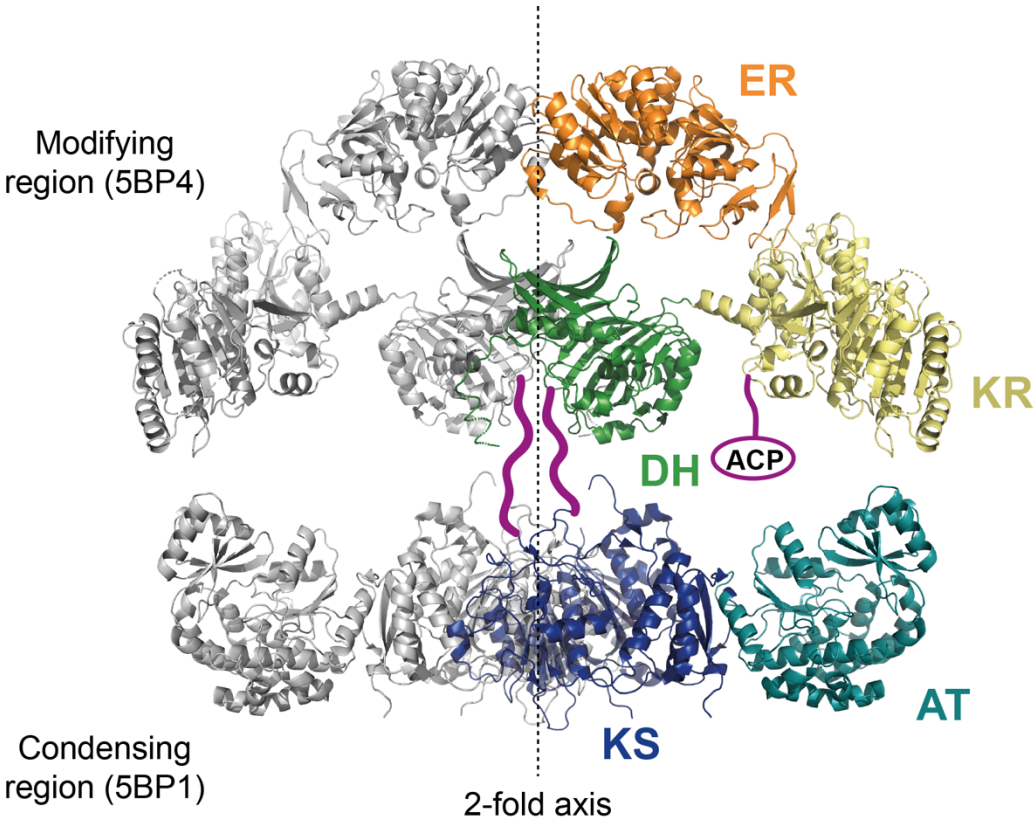


Figure 1-20 Hybrid crystal structure of MAS-like PKS (Herbst et al., 2016). *The MAS-like PKS is homo-dimeric with the two-fold axis shown as a dashed line and one monomer shown in grey and the other monomer coloured according to the domains it comprises. The hybrid crystal structure was constructed from two separate crystal structures, one of the condensing region (PDB code 5BP1) and one of the modifying region (PDB code 5BP2). In total five catalytic domains (KS, AT in the condensing region and DH, ER and KR in the modifying region) were resolved. To improve the crystallisation, the ACP was deleted from the construct but has been shown in its predicted location in this figure. The linker between the condensing and modifying regions is shown in purple. The figure was made using PDB entries 5BP1 and 5BP4.*

The comparison of the condensing region to previous structures (Maier et al., 2008; Tang et al., 2006) revealed close resemblance to both mFAS and DEBS KS-AT didomain; however, the AT was found to be slightly rotated with regard to the KS domain. To facilitate the formation of crystals, the highly-flexible ACP was deleted from the modifying region construct, and SAXS was used to confirm that this deletion did not impact the overall architecture of this region (Herbst et al., 2016). The overall shape of the modifying region was found to be similar to that of mFAS; however, the domains did not appear to be interacting as closely as previously observed. This suggested high structural and conformational variability of this region (Figure 1-20). To obtain structural information of the linker region between the condensing and modifying parts of the MAS-like PKS, a

crystal structure of DH domain, comprising overlapping sequence to the KS-AT didomain, was also determined. Finally, all structures were aligned in accordance to mFAS (Maier et al., 2008) and linked together *in silico* to produce the hybrid model of the full-length MAS-like PKS shown in Figure 1-20.

One of the main observations of the hybrid structure was the shortening of the gap between the modifying and condensing regions; however, it could also be a result of the hybrid experimental approach employed. Overall, these data agreed with the structural and biochemical data of mPKS fragments (Tang et al., 2006; Gay et al., 2013) and mFAS (Maier et al., 2008; Brignole et al., 2009), and reported a structural template for mPKSs. Subtle differences in the didomain angles and the lack of stable interfaces also suggested high conformational flexibility of this system. Further analysis needs to be performed to investigate the different conformational states of the complex during its catalytic cycle.

1.5.5 Unresolved questions in PKS structural biology

The last four years have seen a number of studies marking breakthroughs in the field of structural biology of PKSs. They highlighted that it is possible to study the large and flexible modules of PKSs using cryo-EM and SAXS, techniques deemed to be more dynamic than X-ray crystallography. However, they also highlighted that much is still unknown about these fascinating molecular machines of Nature.

First, despite the visualisation of the ACP in both *cis*-AT (PikAIII, DEBS3, MAS-like PKS) and *trans*-AT (Vir5 module 5) PKS modules, the distances between the catalytic domains and the covalent attachment points of the ACPs were found to be inconsistent for successful catalysis to occur (Dutta et al., 2014; Davison et al., 2014). Therefore, the collection of more reliable high-resolution data, as well as studying these systems in their *holo*- form with appropriate substrates, needs to be carried out.

Second, perhaps a more challenging, question to answer is the structure and positioning of linker domains. In many of these studies, the linker regions were either resolved poorly, or not resolved at all. The determination of minimum structural features required to facilitate accurate docking between different domains and modules will allow the successful engineering of these systems.

Finally, none of the studies discussed above focused on multi-module structural analysis. The first model for PKS assembly line organisation, termed double-helical “Cambridge model”, was proposed back in the 1990s (Staunton and Weissman, 2001), yet no study has been able to prove or deny this model definitively.

Structural biologists continue to push limits of specimen size, flexibility and resolution. That, paired with the continuing technological advancements, suggests that the field of PKS structural biology is still in its early days and many fascinating discoveries are yet to be made.

1.6 Project background

This thesis will focus on the enzymatic machinery responsible for the production of the polyketide natural products indanomycin (chapters 3 and 4) and rhizoxin (chapter 5). This section will give a brief overview of both systems as well as outline the aims of the project.

1.6.1 Indanomycin non-ribosomal peptide synthetase-polyketide synthase

Indanomycin (Figure 1-21) is an antibiotic agent from the pyrroloketoindane family. Indanomycin and other polyether antibiotics have been widely used in animal husbandry (Brimble, 2000). These agents act as ionophores and kill Gram-positive bacteria by chelating its essential metal ions. It is produced by a *Streptomyces antibioticus* hybrid non-ribosomal peptide synthetase – polyketide synthase (NRPS-PKS) NRRL 8167. The BGC encoding this biosynthetic machinery was identified in 2009 and sequenced to ~ 80 kb (C. Li et al., 2009).

The backbone of indanomycin is generated from L-proline, malonyl-CoA, ethylmalonyl-CoA and methylmalonyl-CoA building blocks (Figure 1-21). The total synthesis of indanomycin was achieved by the William group in 1984 (Edwards et al., 1984).

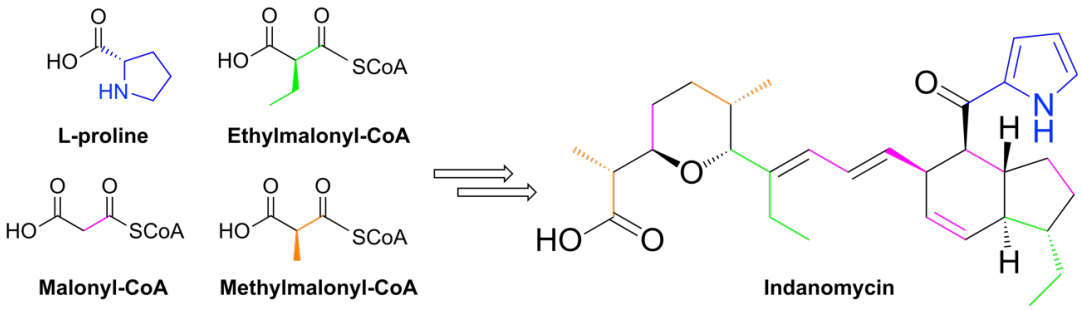


Figure 1-21 Chemical structure of the polyketide indanomycin (C. Li et al., 2009). *The starter unit pyrrole-2-carboxylate in indanomycin biosynthesis is generated from L-proline by the NRPS portion of the hybrid synthase. Chain elongation is performed by the eleven module-containing polyketide synthase from ethylmalonyl-CoA, malonyl-CoA and methylmalonyl-CoA building blocks. Finally, the linear polyketide product is cyclised to generate the final indanomycin product shown here.*

The NRPS part of the biosynthetic machinery is located immediately prior to the start of the eleven-module polyketide synthase region (Figure 1-22) and is responsible for the formation of the starter unit pyrrole-2-carboxylate from L-proline. This starter unit is then loaded onto the first PKS module by a carrier protein (CP) domain, and chain elongation commences.

The PKS portion of the hybrid indanomycin synthase belongs to the *cis*-AT type I modular PKS family. The PKS assembly line is encoded by five large ORFs translating into the polyketide synthase subunits IdmL-P (Figure 1-22). When assembled, the indanomycin PKS has been estimated to reach the size of 2 MDa. The PKS is arranged into ten catalytic modules and the cryptic eleventh module containing a ketosynthase, a non-functional acyltransferase and a final domain, termed Cyc11.

The final module does not contain a TE domain, and therefore the release of the indanomycin intermediate from the assembly line is not completely clear at this stage. The final domain, termed Cyc11, appears to exhibit high sequence similarity to the known ring-opening epoxide cyclases, which are involved in the biosynthesis of other polyether antibiotics (C. Li et al., 2009).

Many other proteins are encoded in the same biosynthetic gene cluster, and play essential roles in polyketide precursor biosynthesis, regulation, resistance and tailoring of indanomycin (C. Li et al., 2009).

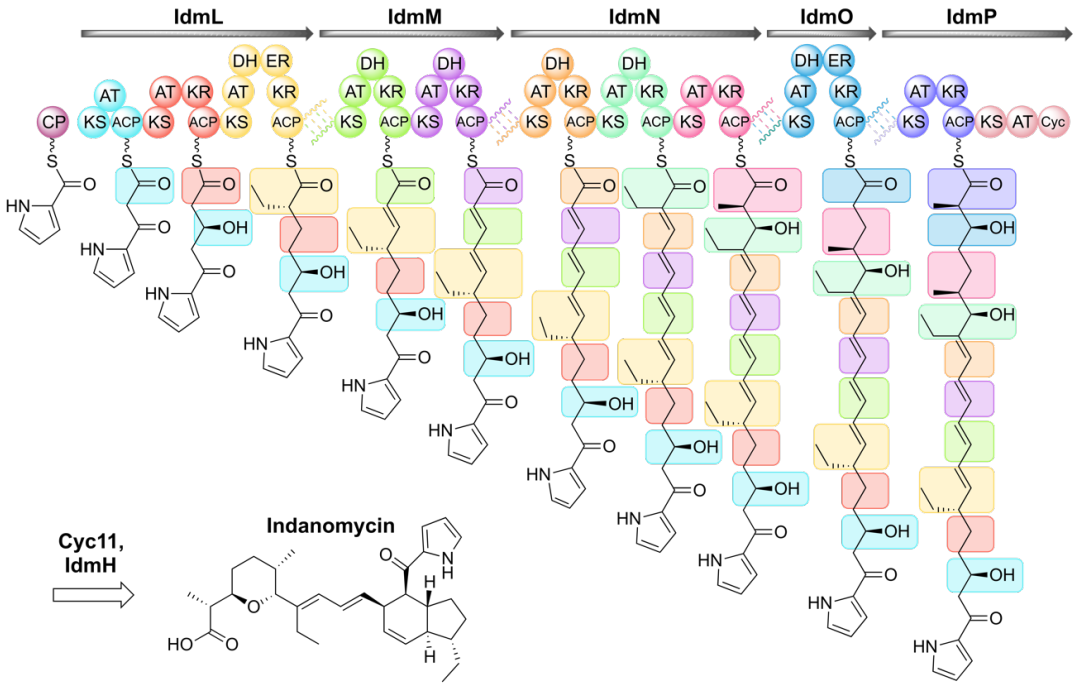


Figure 1-22 Schematic diagram showing the PKS portion of the hybrid indanomycin NRPS/PKS. *This PKS is encoded by five genes translating to subunits IdmL to IdmP. These subunits contain a total of eleven modules, the first ten of which are catalytic and the eleventh module is not thought to be involved in the chain elongation. Finally, at least two proteins (Cyc11 from the module 11 and IdmH) are hypothesised to be involved in the maturation of the final product (C. Li et al., 2009). CP denotes the carrier protein from the NRPS module, which is responsible for loading the pyrrole-2-carboxylate starter unit onto IdmL.*

At least two post-PKS enzymatic modifications of the linear indanomycin intermediate **2** are needed to produce the mature indanomycin **1** (Figure 1-23) (C. Li et al., 2009). This includes the incorporation of two distinct structural features: a tetrahydropyran ring and the indane rings. The tetrahydropyran ring has been hypothesised to be formed by a direct nucleophilic replacement mediated by Cyc11 which results in the production of **3**.

The lack of a TE in the final module of the PKS suggests an alternative mechanism of release of the full-length indanomycin precursor **2** or **3** from IdmP module 11 (Figure 1-23). It could be possible that the formation of the tetrahydropyran ring in **3** allows diffusion of a water molecule into the active site of Cyc11, which is then followed by hydrolysis resulting in **4**.

The indane ring formation is postulated to be mediated by a putative cyclase, IdmH, a 16 kDa protein encoded separately in the indanomycin BGC. IdmH shows distant sequence similarity (<40%) to a number of putative polyketide cyclases that have been shown to catalyse ring closure steps in the biosynthesis

of the anthracycline antitumour antibiotics produced in *Streptomyces* (C. Li et al., 2009; Sultana et al., 2004).

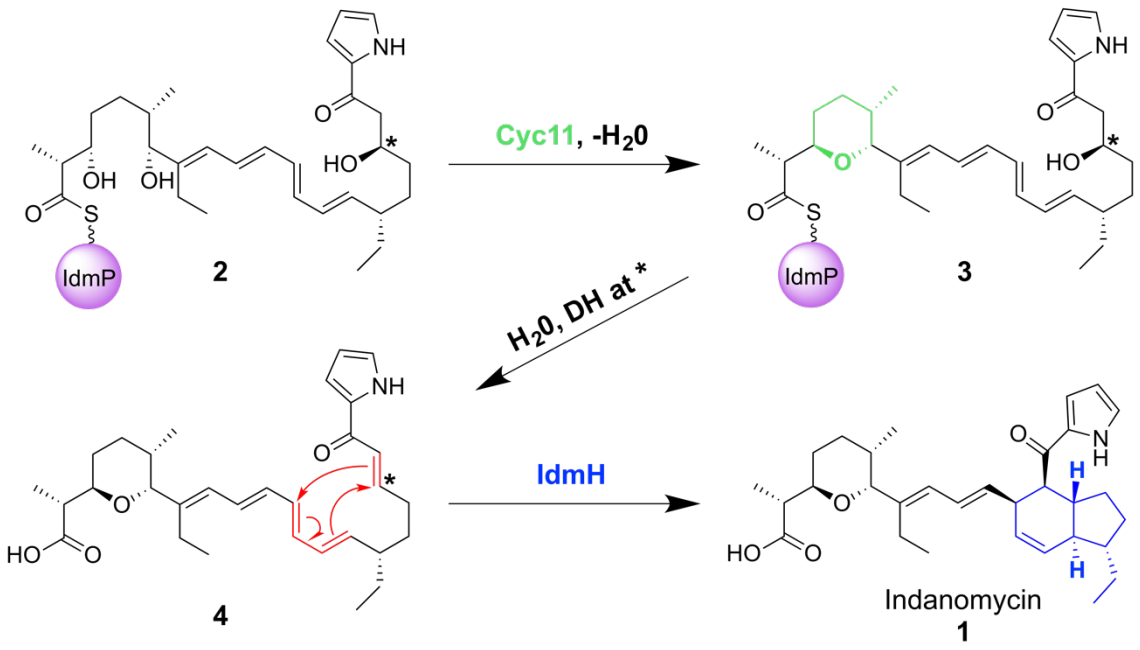


Figure 1-23 The proposed mechanism for the formation of tetrahydropyran and indane rings (C. Li et al., 2009; Rommel et al., 2011). *Cyc11 is thought to mediate the formation of the tetrahydropyran ring through a direct nucleophilic replacement to generate 3. This reaction could then be followed by hydrolysis to release 4 from the IdmP subunit. Finally, indane ring formation in the mature indanomycin 1 is thought to be mediated by a putative cyclase IdmH. For a successful [4+2] cycloaddition to occur, C19 of 2 (denoted with an asterisk) needs to be dehydrated to produce a double bond. Since module 2 does not contain a DH domain, a hypothetical dehydration step was included between intermediates 3 and 4.*

A mechanism proposed for the indane ring formation is consistent with a Diels-Alder [4+2] cycloaddition (Kim et al., 2012). In this reaction, a diene and a dienophile from **4** react through a single electrocyclic transition state to allow the formation of the five- and six-membered rings in **1** (Figure 1-23). It is also, however, possible that the cyclisation reaction occurs in two non-concerted transition states, like in a two-step Michael-aldol sequence (Fage et al., 2015).

It is also important to note that the predicted linear indanomycin PKS product **2** retains a hydroxyl, rather than a double bond, at C19 (denoted with an asterisk in Figure 1-23). This is a result of module 2 of the synthase not containing a DH domain. For the Diels-Alder reaction to be possible, a suitable dienophile is needed, and, therefore, C19 needs to be dehydrated by either *a trans*-acting DH

or a DH domain from another module. It is also possible that module 2 is skipped altogether and module 3 performs the elongation cycle twice, like in iPKSs.

Diels-Alder reactions are widely used in synthetic organic chemistry; however, until recently only a few naturally evolved enzymes have been suggested to act as Diels-Alderases (Auclair et al., 2000; Linder et al., 2012). In 2016, the first crystal structures of two natural Diels-Alderases, spirotetronate cyclase AbyU and spiro-conjugate synthase PyrI4, were reported (Byrne et al., 2016; Zheng et al., 2016). Enzyme kinetic, molecular dynamics and NMR spectroscopy data from these studies provide compelling evidence that enzyme-catalysed Diels-Alder reactions can be found in natural product biosynthesis.

The first results chapter of this thesis will focus on the putative indanomycin cyclase IdmH. To determine the exact role of this enzyme in the biosynthesis of indanomycin, IdmH will be characterised using a number of structural and biochemical techniques.

The second results chapter will focus on IdmO, the discrete ninth module of the indanomycin megasynthase (Figure 1-22). It contains all six possible domains, and structural characterisation of such fully-reducing modular PKS module would provide us with valuable information regarding the assembly and operation of complex polyketide synthase assembly lines.

1.6.2 Rhizoxin PKS

The third results chapter focuses on a *trans*-AT type I mPKS responsible for the biosynthesis of the natural product rhizoxin.

Rhizoxin is a macrolytic polyketide metabolite produced by *Burkholderia rhizoxinica*, an intracellularly living bacterium isolated from plant pathogenic fungi belonging to the genus *Rhizopus*. The infection of this pathogen results in a plant disease known as rice seedling blight. In this fascinating fungi-bacterium symbiotic-pathogenic relationship, *B. rhizoxinica* produces rhizoxin, which the fungi use to infect the rice. Rhizoxin binds to rice β-tubulin resulting in inhibition of mitosis and cell cycle arrest (Partida-Martinez and Hertweck, 2005).

Rhizoxin possesses remarkably strong antimitotic activity and therefore has been a focus of interest as a potential antitumor drug (Tsuruo et al., 1986; Jordan et al., 1998). The strong antimitotic activity has been found to be utterly dependent on a short (C2) carbon chain branching away from the main macrolide ring

(Scherlach et al., 2006; Kusebauch et al., 2011; Schmitt et al., 2008; Hong and White, 2004).

Rhizoxin is synthesised by a large and complex *trans*-AT type I polyketide synthase (Figure 1-24) (Bretschneider et al., 2013). The modular portion of the synthase is comprised of six large polypeptides, RhiA to RhiF, which contain one loading, one non-ribosomal peptide synthetase (NRPS1), one branching, one off-loading and eleven catalytic modules. Additionally, three other genes from the same biosynthetic cluster encode smaller vital proteins RhiG, RhiI and RhiH (Partida-Martinez and Hertweck, 2007). RhiG is a *trans*-AT and is essential for the chain elongation catalysis in each PKS module. RhiI, a methyltransferase, and RhiH, a cytochrome P450 monooxygenase, are both crucial for the generation of the mature rhizoxin from the linear rhizoxin intermediate released from the assembly line.

It was thought initially that the branching reaction was a result of a *trans*-acting processing enzyme. However, the sequencing of the *B. rhizoxinica* genome failed to detect any enzymes which could act *in trans* to introduce this branching (Lackner, Moebius, Partida-Martinez and Hertweck, 2011; Lackner, Moebius, Partida-Martinez, Boland, et al., 2011). Further bioinformatic analysis and investigation of the prematurely liberated products of the rhizoxin assembly line led to the discovery of a novel type of PKS module which itself catalyses a branching reaction in the polyketide backbone (Partida-Martinez and Hertweck, 2007; Kusebauch et al., 2009) (Figure 1-24).

This new type of module, termed “branching”, is part of the RhiE subunit and is located in between the tenth and the eleventh catalytic modules of the PKS. It consists of three domains, a ketosynthase, an acyl-carrier protein and a branching domain (B), and has a molecular mass of approximately 115 kDa.

This module has been extensively studied using classical biochemical techniques (such as site-directed mutagenesis and *in vitro* reconstitution), and a three-dimensional structure of part of this module, containing two out of three domains, was solved using X-ray crystallography (Bretschneider et al., 2013).

The third results chapter focuses on the full-length branching module from the rhizoxin PKS. The highly flexible ACP present in the full-length polypeptide hinders the formation of crystals required for X-ray crystallography studies, and

hence a cryo-EM approach has been chosen instead. Past studies of the branching domain will be discussed together with the results.

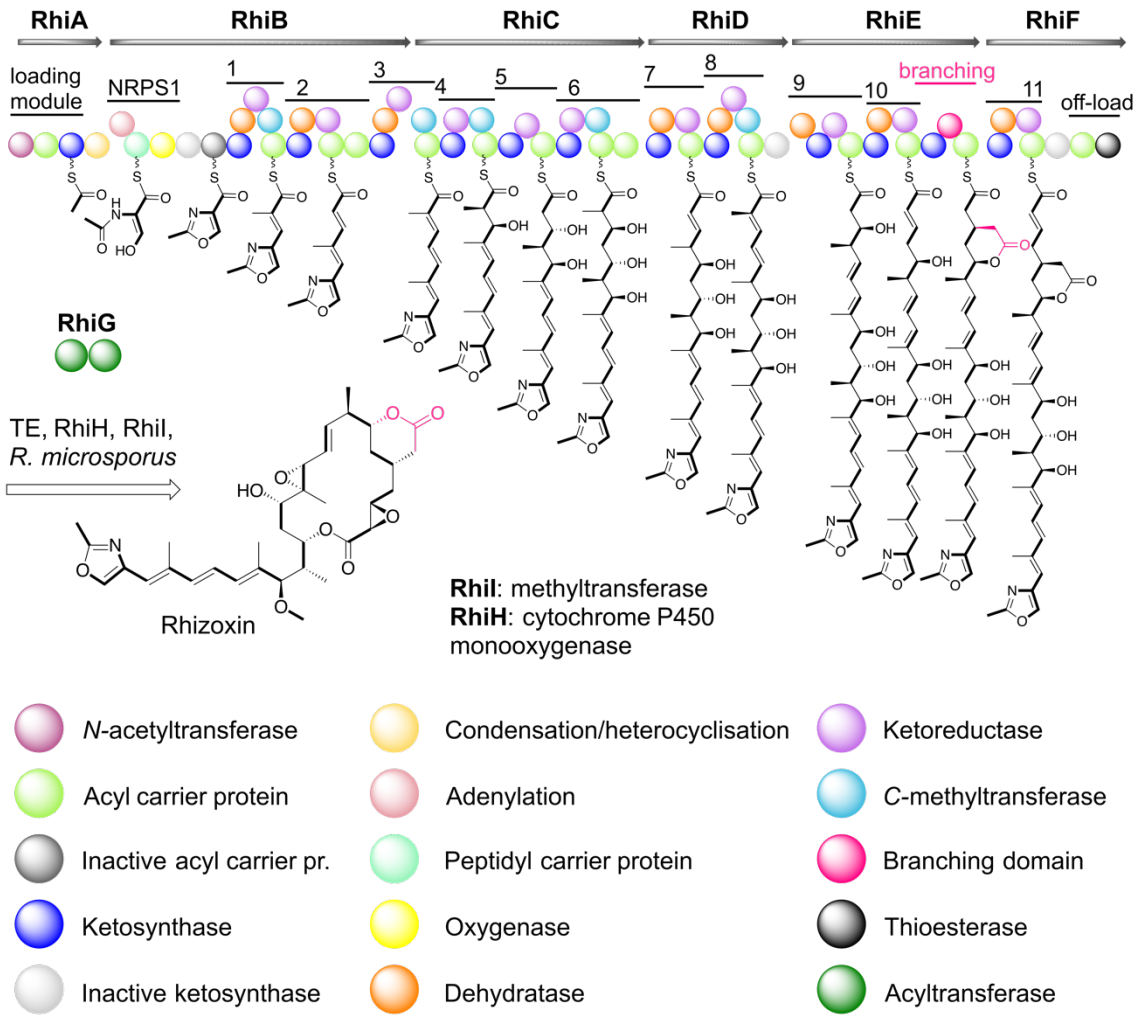


Figure 1-24 Schematic diagram of the rhizoxin synthase. *This megasynthase is encoded by six genes translating to PKS subunits RhiA to RhiF. These subunits contain a total of eleven catalytic modules, loading and off-loading modules, one NRPS module and one branching module. Three additional genes encoding a trans-acyltransferase RhiG, a methyltransferase RhiI and a cytochrome P450 monooxygenase RhiH are part of the same biosynthetic cluster and are fundamental for the formation of rhizoxin product. Branching module is responsible for inserting the C2 carbon branch in the linear backbone of the rhizoxin (Lackner, Moebius, Partida-Martinez and Hertweck, 2011; Partida-Martinez and Hertweck, 2007; Kusebauch et al., 2009).*

1.6.3 A summary of project goals

Since their discovery in the 1990s, significant efforts have been made to understand the structure-function relationships of modular polyketide synthases. Despite some previous successes in the synthetic engineering of these systems, it is becoming apparent that the lack of high-resolution structural information of the intact PKS complexes is hindering the successful generation of polyketide analogues for evaluation as new pharmaceutical leads.

This thesis describes the experiments carried out to structurally and functionally characterise the indanomycin and rhizoxin PKSs and the associated enzymes to gain a better understanding of the interfaces between the functional domains, the assembly of the full complex and the chemistry involved in the generation of the mature polyketide.

The proposed work has three primary goals:

- **To characterise the structure and function of the post-PKS putative indanomycin cyclase IdmH.**

Rational engineering of chimeric PKS lines requires a catalogue of functional domains, modules, linkers and accessory enzymes capable of catalysing the formation of structurally diverse and novel natural products. Elucidating the structure and reaction mechanism of the novel enzyme IdmH will expand a synthetic biologist's toolkit of natural product modifying enzymes.

To achieve this, multiple steps will have to be taken, such as: cloning of the *idmH* gene and construction of an expression vector, recombinant protein expression in *E. coli* and subsequent purification to obtain a homogenous sample of IdmH, structural characterisation by X-ray crystallography and functional characterisation by other biochemical methods. Chapter 3 describes the approaches taken to achieve these objectives.

- **To elucidate the structure of IdmO, the fully-reducing ninth module from the indanomycin PKS.**

Due to the sheer size and high flexibility of PKS systems it has not been possible to obtain crystal structures of any intact PKS modules, let alone the whole complexes (Weissman, 2015b). Elucidating the structures of full-length PKS modules will increase our understanding of the interfaces between the functional

domains and bring us closer to the rational engineering of these systems for the production of novel therapeutic compounds.

IdmO is the discrete module of the fourth subunit of the indanomycin PKS and contains all possible PKS catalytic domains making it an especially exciting target for structural studies. Chapter 4 describes the cloning, expression and purification of the IdmO as well as preliminary structural and functional characterisation of this full-length PKS module by electron microscopy and enzyme kinetics.

- **To investigate the three-dimensional architecture of the full-length branching module from the rhizoxin PKS.**

Branching of the linear polyketides has the potential to increase the diversity and functionality of polyketide natural products. A novel type of PKS module introducing a branching rather than the head-to-tail fusion of extender units was discovered in rhizoxin PKS. This module was successfully characterised using a plethora of biochemical techniques (Bretschneider et al., 2013); however, the structure of the full-length functional branching module comprising the flexible ACP remains an enigma. Chapter 5 describes the cryo-EM approaches taken to investigate the architecture of the full-length branching module.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Chemicals and biological reagents

Unless otherwise stated, all chemicals and reagents were purchased from Sigma-Aldrich or Fisher Scientific Ltd.

Agar, agarose, tryptone, isopropyl-β-D-1-thiogalactopyranoside (IPTG) and dithiothreitol (DTT) were purchased from Melford Laboratories Ltd. Ammonium persulfate was purchased from Amersham Biosciences Ltd. $^{15}NH_4Cl$ and D-glucose-$^{13}C_6$ were purchased from Cambridge Isotope Laboratories, Inc. Indanomycin was obtained from Insight Biotechnology. Quick Start Bradford dye reagent and Laemmli Sample Buffer were purchased from Bio-Rad Laboratories Ltd; InstantBlue protein dye was obtained from Expedeon Ltd.

All restriction enzymes were purchased from New England Biolabs (NEB). Thrombin protease was purchased from GE Healthcare; lysozyme protease was purchased from MP Biomedicals. Q5® High-Fidelity DNA Polymerase, Quick-Load® 1 kb Extend DNA ladder and broad range Colour Prestained Protein Standard were purchased from NEB. Deoxynucleotide Triphosphates (dNTPs), and 100 bp and 1 kb DNA ladders were purchased from Promega.

2.1.2 Consumables and chromatography media

Mini-PROTEAN® TGX™ precast protein gels were purchased from Bio-Rad Laboratories Ltd. Vivaspin® centrifugal concentrators were purchased from either Generon or Sartorius Stedim UK Ltd. Miniprep kits were purchased either from Promega or NEB; Qiaquick® PCR purification kit was obtained from QIAGEN.

Unless stated otherwise, all chromatography resin and prepared columns were purchased from GE Healthcare. Ni-NTA affinity resin was purchased from Qiagen; chitin resin beads were purchased from NEB; StrepTactin® beads and prepacked columns were purchased from IBA GmbH.

Crystallisation solutions were purchased either from Qiagen, Molecular Dimensions or Hampton Research; EM grids were obtained from either Agar Scientific or Quantifoil Micro Tools GmbH.

2.1.3 Growth media and antibiotics

Unless stated otherwise, all growth media were prepared by dissolving the ingredients in deionised water and autoclaving at 121 °C, 15 psi for 20 minutes. Ready media were then stored at room temperature. The exact ingredients of each growth media used in this work can be found in Table 2-1.

Table 2-1 Growth media composition and preparation notes.

Solution	Ingredients	Preparation notes
2TY medium	16 g/l Tryptone 10 g/l Yeast extract 5 g/l NaCl	As above.
10x M9 salts	80 g/l Na_2HPO_4 40 g/l KH_2PO_4 5 g/l NaCl **5 g/l NH_4Cl***	As above. ***NH_4Cl can be replaced with $^{15}NH_4Cl$.**
M9 additives	1 M $MgSO_4$ (autocl.) 1 M $CaCl_2$ (autocl.) **20% (*w/v*) glucose* (filt. or autocl.)** 5% (*w/v*) FeCl(III) (filt.) 100x MEM Vitamins (Sigma)	First, M9 salts were mixed with deionised water and autoclaved. Prior to inoculation, each litre of M9 medium was supplemented with: 2 ml $MgSO_4$, 100 µl $CaCl_2$, 20 ml 20% glucose, 800 µl FeCl(III), 10 ml MEM vitamins**. *Glucose can be replaced with ^{13}C D-glucose.**
Trace elements (100x)	5 g/l EDTA 0.83 g/l $FeCl_3$ x 6 H2O 84 mg/l $ZnCl_2$ 13 mg/l $CuCl_2$ x 2 H2O 10 mg/l $CoCl_2$ x 6 H2O 10 mg/l H_3BO_3 1.6 mg/l $MnCl_2$ x 6 H_2O	First the EDTA was dissolved in water and the pH was adjusted to 7.5. Then the other components were added and the solution was sterilised by filtration. Trace elements instead of FeCl(III) were used to make the M9 additives solution during selenomethionine labelling experiments.
Agar plates	Chosen medium 20 g/l agar	Agar was added to a chosen medium and the mixture was autoclaved. Agar was allowed to cool down to 60 °C before the addition of antibiotic(s) and pouring onto the petri dishes.

Antibiotic stocks were prepared as outlined in Table 2-2. An appropriate amount of antibiotic was weighed out, dissolved, sterile filtered using a syringe with a 0.22 µm filter and stored in 0.1-1 ml aliquots at -20 °C.

Table 2-2 Antibiotic solutions.

Antibiotic	Stock preparation	Working concentration
Kanamycin (Kan)	50 mg/ml in dH20	50 µg/ml
Ampicillin (Amp)	100 mg/ml in dH20	100 µg/ml
Chloramphenicol (Cam)	34 mg/ml in ethanol	34 µg/ml

2.1.4 Bacterial strains

E. coli strains used during this research work are outlined in Table 2-3. All transformations were performed according to the manufacturer's guidelines.

Table 2-3 *E. coli* strains used in this research.

Strain (Supplier)	Purpose	Genotype
XL10-Gold® Ultracompetent (Stratagene)	DNA manipulation	Tet^{r}Δ(*mcrA*)*183* Δ (*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1* lac Hte [F´ *proAB lacI*^{q}Z*ΔM15* Tn*10* (Tet^r) Amy Cam^r]
5-alpha (High Efficiency) (NEB)	DNA manipulation	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
BL21-Gold(DE3) (Agilent)	Protein expression	*E. coli* str. B F$^-$ *ompT hsdS$_B$*($r_B^-m_B^-$) *dmc*$^+$ Tet^r *gal* λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) *endA* Hte
Lemo21(DE3) (NEB)	Tuneable protein expression	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo*(Cam^R) *λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5* pLemo = pACYC184-*PrhaBAD-lysY*
NiCo21(DE3) (NEB)	Protein expression with improved purity	*can::CBD fhuA2 [lon] ompT gal (λ DE3) [dcm] arnA::CBD slyD::CBD glmS6Ala ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*
B834(DE3) (Merck Chemicals Ltd)	Methionine autotroph strain of BL21(DE3) for labelling	*E. coli* str. B F$^-$ *ompT hsdS$_B$*($r_B^-m_B^-$) gal *dmc met* (DE3)

2.1.5 Oligonucleotides, genes and vectors

Oligonucleotide primers were synthesised by Integrated DNA Technologies, Inc.

The *idmH* and *idmO* genes were amplified by PCR from *Streptomyces antibioticus* NRRL 8167 genomic DNA. A codon-optimised gene for the ACP_9 was synthesised by GenScript.

pET28a(+) and pCDF_duet_SFP plasmid vectors were kind gifts from Dr Ryan Seipke (Faculty of Biological Sciences, University of Leeds, UK). SUMO tag and TEV protease cleavage sites were amplified from pET_SUMO and pET26c_GO plasmid vectors, kind gifts from Dr Glyn Hemsworth and Prof. Mike McPherson (Faculty of Biological Sciences, University of Leeds, UK).

2.2 DNA methods

2.2.1 Plasmid DNA purification and quantification

Plasmid DNA preparations were performed using either the Wizard® Plus SV Minipreps DNA Purification Kit (Promega) or Monarch® Plasmid Miniprep Kit (NEB) according to the manufacturer's guidelines.

DNA concentration was determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific). 1 µl aliquot of the DNA solution was applied to the sample pedestal and the DNA concentration (ng/µl) was determined from the A_{260} reading.

2.2.2 Primer design

In order to amplify the genes of interest from the genomic DNA of *Steptomyces antibioticus,* oligonucleotide primers, designed to bind to the opposite strands of the DNA helix, were synthesised (Table 2-4). Primers used to sequence the pET coding region and the full-length IdmO are also included in the table. NEBuilder online tool was used to design primers with complementary overhangs. NEBaseChanger™ online tool was used to design mutagenesis primers.

Table 2-4 The oligonucleotide primers used molecular biology. *Capital letters represent the oligonucleotide sequence complementary to the template DNA whereas lower case letters represent the overlapping region.*

Function	Code	Direction	Primer sequence (5' to 3')
Amplification of *idmH*	ID1	Fwd	ctggtgccgcgcggcagccatATGGCTCATCAGCCTTCG
	ID2	Rev	tccaccagtcatgctagccatTCACAGGGACGCCTTCAC
Amplification of pET28 for constructing pET28_IdmO	ID17	Fwd	CACCACCACCACCACCAC
	ID6	Rev	GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAG
***idmO* nested PCR round**	ID43	Fwd	GACGAAATGTTCGCGCTG
	ID45	Rev	CTGCTCAGGACGATGGAC
Amplification of *idmO*	ID51	Fwd	ctttaagaaggagatataccATGCACATGGTCGGCGTGG
	ID52	Rev	cagtggtggtggtggtggtgTGAACGGTCCGTCCCTTCGG
Sequencing of the pET28 multiple cloning site	T7P	Fwd	TAATACGACTCACTATAGGG
	T7TERM	Rev	CTAGTTATTGCTCAGCGG
Sequencing of *idmO*	IdmOseq1	Fwd	TCGGGATGCTCCTGCTG
	IdmOseq2	Fwd	TGCCTTGGGTCTGGATC
	IdmOseq3	Fwd	TACTGGCTGGAGGACG
	IdmOseq4	Fwd	ACTGGGTTCCCGTGGAG

Sequencing of *idmO*	IdmOseq5	Fwd	ACGGGTTGGTGGACCTG
	IdmOseq6a	Fwd	GAGGCACTCTCGATGGACC
	IdmOseq6b	Fwd	ACTGTTCGGCATCTCACC
Amplification of Strep tag and TEV cl. site	ID57	Fwd	TAAGAAGGAGATATACCATGTGGAGCCATCCGCAGTTTG
	ID58	Rev	TCTTCCACGCCGACCATGTGACCCTGGAAGTACAGGTTTTCTTTCT
Amplification of pET28_IdmO for the insertion of Strep tag	ID55	Fwd	CACATGGTCGGCGTGGAAG
	ID56	Rev	CATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGG
Amplification of *acp9*	ID59	Fwd	aggagataccatgCGGCTCGCCGGCC
	ID60	Rev	gtggtggtggtggtgGGTGTCGAGGGCGAGTTCG
Amplification of pET28 for constructing pET28_ACP$_9$	ID17	Fwd	CACCACCACCACCACCAC
	ID35	Rev	CATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGG
Inserting TEV protease cleavage site between *idmO* and His6	ID51	Fwd	ctttaagaaggagatataccATGCACATGGTCGGCGTGG
	ID65	Rev	CGACTGGAAATACAGATTTTCTGAACGGTCCGTCCCTTCG
Amplification of pET28 for constructing pET28_IdmO_TEV_His	ID67	Fwd	GAAAATCTGTATTTCCAGTCGCACCACCACCACCACCAC
	ID66	Rev	GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTATC

Amplification of codon-optimised *acp9*	ID71	Fwd	AGAACAGATTGGTGGTATGGAGCGTCTGGCGG
	ID72	Rev	ATACCTAAGCTTGTCTTTAGTGGTGGTGATGATGGTGACC
Amplification of the vector backbone for constructing pET_SUMO_acp9	ID73	Fwd	AGACAAGCTTAGGTATTTATTCGGC
	ID74	Rev	ACCACCAATCTGTTCTCTGTG
***idmO* AT mutagenesis (YASH = AT1) – 2 pairs of primers were used to split Kan^R**	ID96	Fwd1	CTTTTAACAGCGATCGCGTATTTCGTC
	ID79	Rev1	gagcCATTCGCCGTTGATGGAACC
	ID95	Fwd2	GACAATTACAAACAGGAATCGAATGCAACC
	ID80	Rev2	gcataACTGACCCGCAGACGCTT
***idmO* AT mutagenesis (HASH = AT2)**	ID81	Fwd	gagcCATTCGCCGTTGATGGAACC
	ID82	Rev	gcatgACTGACCCGCAGACGCTT
***idmO* AT mutagenesis (YAFH = AT3) – 2 pairs of primers were used to split Kan^R**	ID96	Fwd1	CTTTTAACAGCGATCGCGTATTTCGTC
	ID83	Rev1	gtttCATTCGCCGTTGATGGAACC
	ID95	Fwd2	GACAATTACAAACAGGAATCGAATGCAACC
	ID84	Rev2	gcataACTGACCCGCAGACGCTT

***idmO* AT mutagenesis (VASH = AT4) – 2 pairs of primers were used to split Kan^R**	ID96	Fwd1	CTTTTAACAGCGATCGCGTATTTCGTC
	ID85	Rev1	gagcCATTCGCCGTTGATGGAACC
	ID95	Fwd2	GACAATTACAAACAGGAATCGAATGCAACC
	ID86	Rev2	gccacACTGACCCGCAGACGCTT
***idmO* AT mutagenesis (AAAH = AT5) – 2 pairs of primers were used to split Kan^R**	ID96	Fwd1	CTTTTAACAGCGATCGCGTATTTCGTC
	ID87	Rev1	cgcgCATTCGCCGTTGATGGAACC
	ID95	Fwd2	GACAATTACAAACAGGAATCGAATGCAACC
	ID88	Rev2	gccgcACTGACCCGCAGACGCTT
IdmH mutagenesis to generate IdmH-Δ96-104 (Obajdin, 2018)	97-106loop_fwd		GGCCGGCCCTTCGAGATG
	97-106loop_rev		GGAGTGCGTTCCCCGTGC
Universal sequencing primers	T7	Fwd	TAATACGACTCACTATAGGG
	T7Term	Rev	GCTAGTTATTGCTCAGCGG

2.2.3 Polymerase chain reaction and site directed mutagenesis

Unless stated otherwise, the polymerase chain reaction (PCR) was carried out using Q5 DNA Polymerase according to NEB's instructions. Annealing temperatures for PCR were estimated using NEB T_m calculator.

Site-directed mutagenesis was performed using Q5® Site-Directed Mutagenesis kit according to the NEB's guidelines using primers from Table 2-4.

2.2.4 Agarose gel electrophoresis

0.6-2% agarose gels were made by combining TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) with an appropriate amount of agarose. SYBRsafe DNA gel stain (Invitrogen) was added to cooled-down molten agarose.

DNA samples were prepared by the addition of 6x gel loading dye (Promega). One of the following DNA size ladders were used: 100 bp (Promega), 1 kb (Promega) or 1 kb Extend (NEB). The samples were resolved by applying 100V voltage for 0.5-1.5 hours. DNA was visualised under UV light and photographed.

2.2.5 Plasmid construction

E. coli XL10 Gold ultracompetent cells were used as a host for manipulating plasmid DNA. Plasmid ligation was performed either with a NEBuilder® HiFi DNA Assembly® kit or by using a ligase-independent FastCloning approach (C. Li et al., 2011). Plasmid DNA or a ligation mix was transformed into *E. coli* competent cells using the heat shock method according to the manufacturer's guidelines.

2.2.5.1 HiFi DNA Assembly® method

HiFi DNA Assembly® is analogous to the Gibson assembly method. Both methods join the insert(s) and the plasmid in a single, isothermal, three-enzyme-containing reaction in a small (20 µl) working volume. pET28_IdmH vector was constructed using this method. The vector was linearised using *Nde*I restriction enzyme. The insert was amplified by PCR to add the overlapping regions and then purified using a purification kit. The reaction was assembled and performed according to the NEB's guidelines.

2.2.5.2 Ligase-independent cloning method

Ligase-independent cloning method, as described in FastCloning protocol (C. Li et al., 2011), was used for the construction of all pET_IdmO and pET_ACP_9 plasmid vectors. Briefly, the insert and the vector were amplified using the primers designed to create the complementary regions of DNA. The fragments were amplified using 18-24 cycles and then mixed together in 1:1 ratio in the presence of *Dpn*I restriction enzyme. The reaction was incubated at 37 °C for 1 hour and 2 µl of the ligation mix were used to transform XL10 Gold ultracompetent cells.

2.2.5.3 pET28_IdmH

The gene sequence encoding the full-length *idmH* gene (ACN69984.1) was amplified by PCR using the genomic DNA from *S. antibioticus* NRRL8167 (kind gift from Dr Ryan Seipke, Faculty of Biological Sciences, University of Leeds, UK) as a template and primers ID1 and ID2 (Table 2-4). The amplified sequence contained overhang sequences which facilitated ligation to a *NdeI*-linearised pET28(a)+ vector using the HiFi DNA Assembly® kit (NEB) (section 2.2.5.1). The generated plasmid, pET28_IdmH, comprised an N-terminal hexahistidine tag and a thrombin cleavage site. The inserted *idmH* gene was confirmed by sequencing with T7 and T7term universal primers. Site-directed mutagenesis to generate IdmH-Δ96-104 was performed using wild-type pET28_IdmH as a template and primers 97-106loop_fwd and 97-106loop_rev (Obajdin, 2018). Site-directed mutagenesis PCR product was treated with KLD enzyme mix (NEB) before transformation into *E. coli* 5-α competent cells.

2.2.5.4 pET28_IdmO

The gene sequence encoding the full-length *idmO* gene (ACN69991.1) was amplified from the genomic DNA of *S. antibioticus* NRRL8167 in two rounds of nested PCR (nPCR). First, a DNA fragment surrounding the *idmO* gene was amplified in the presence of GC enhancer (NEB) using primers ID43 an ID45. Then, DNA products from the first round of nPCR were purified using Qiaquick PCR purification kit (QIAGEN) and used as a template for the second round of nPCR. The full-length *idmO* gene was amplified using primers ID51 and ID52 which were designed to extend the ends of the PCR product to include overhangs to the vector facilitating a ligase-independent cloning approach (C. Li et al., 2011). pET28 plasmid was linearised using primers ID17 and ID6 (Table 2-4). The

generated plasmid, pET28_IdmO, comprised a C-terminal hexahistidine tag. The successful *idmO* clone was confirmed by sequencing with T7 and T7term universal primers and specific IdmO sequencing primers outlined in Table 2-4.

Analogous approach was used for all subsequent modifications of pET28_IdmO construct. Primer sequences used for the amplification of *idmO*, vector backbone, Strep II tag and TEV protease cleavage site can be found in Table 2-4.

2.2.5.5 pET28_ACP$_9$ and pET_SUMO_ACP$_9$

Initially, the *acp$_9$* gene was amplified by PCR using pET28_IdmO as a template and primers ID59 and ID60. The amplified sequence contained overhangs which facilitated cloning into a primers ID17 and ID35-linearised pET28 vector according to the FastCloning method (C. Li et al., 2011).

Due to the absence of soluble expression, a codon-optimised *acp$_9$* gene was purchased from GenScript. To improve the solubility and expression levels of codon-optimised ACP$_9$, the *acp$_9$* was subcloned into The Champion™ pET SUMO expression vector (Thermo Fisher Scientific) to yield pET_SUMO_ACP$_9$ construct. For subcloning, primers ID71 and ID72 were used to amplify the *acp$_9$* sequence, and primers ID73 and ID74 were used to linearise pET_SUMO vector.

2.2.5.6 Colony screening

The resulting colonies were assessed by colony PCR using Q5 DNA polymerase using standard protocols. Any colonies showing the correct band size on an agarose gel were picked for plasmid amplification, isolation and sequencing.

2.2.6 DNA sequencing

DNA sequencing was carried out by Genewiz (formerly known as Beckman Coulter Genomics). Sequencing primers can be found in Table 2-4. Plasmid DNA (15 µl of approximately 100 ng/µl) and custom sequencing primers (10 µM, 5 µl per reaction) were provided for Express Single sequencing service and sequencing data were returned as .ab1 files. Results were analysed using Clustal Omega Multiple Sequence Alignment online tool.

2.3 Protein methods

2.3.1 Recombinant protein expression

2.3.1.1 General notes on scaling up the expression from a single colony

A single colony of *E. coli* cells containing the plasmid(s) of interest was used to inoculate 5 ml day culture containing the appropriate antibiotic(s). If the culture was to be used for a small-scale protein expression screen, the cells were grown at 37 °C with shaking at 200 rpm until the OD_{600} reached 0.6 before protein expression was induced with IPTG and the recombinant protein was allowed to express at 15 °C overnight. For large-scale protein expression, 5 ml starter day culture was incubated at 37 °C with shaking at 250 rpm for 6-8 hours after which a 500 µl aliquot was used to inoculate a 50 ml overnight starter culture. The next morning, 10 ml of the overnight culture were used to spike 0.5 l volume of medium in a 2 litre shaker flask. The culture was then grown at 37 °C with shaking of 200 rpm until the OD_{600} reached 0.4-0.6 and the protein expression was induced by adding an appropriate amount of IPTG. The induced culture was then incubated overnight at 15 °C with shaking at 200 rpm.

2.3.1.2 Native IdmH and IdmH-Δ96-104

Recombinant IdmH was expressed in *E. coli* BL21(DE3) cells harbouring the pET28_IdmH plasmid. Cells were grown in 2TY medium containing 50 µg/ml kanamycin, at 37 °C to an OD_{600} of 0.4 before lowering the temperature to 15 °C. Expression was induced with 0.4 mM IPTG at an OD_{600} of 1.0 and cells were allowed to grow overnight. The cells were harvested by centrifugation at 6,000 *g* at 4 °C for 15 min, and the pellets stored at -80 °C until use.

2.3.1.3 Single- and triple-labelled IdmH

Single labelling of IdmH with ^{15}N isotope was achieved by cultivating *E. coli* BL21(DE3) cells harbouring the pET28_IdmH plasmid in M9 minimal medium supplemented with $^{15}NH_4Cl$. Triple labelling of IdmH with ^{15}N, ^{13}C and ^{2}H isotopes was achieved by cultivating *E. coli* BL21(DE3) cells harbouring the pET28_IdmH plasmid in M9 minimal medium made up with D_2O instead of deionised H_2O and supplemented with $^{15}NH_4Cl$ and D-glucose-$^{13}C_6$. Scale-up cultures (5 ml day and 50 ml overnight) did not contain isotope compounds, just regular NH_4Cl and D-glucose-C_6. The rest of the expression was performed as for the native IdmH.

The labelling was confirmed by liquid chromatography/mass spectrometry (LC/MS) analysis.

2.3.1.4 Selenomethionine-labelled IdmH and IdmH-Δ96-104

Production of selenomethionine-labelled IdmH was achieved by cultivation of methionine auxotrophic *E. coli* B834(DE3) cells in M9 mineral medium supplemented with L-(+)-selenomethionine (Anatrace). During culture scale-up, 5 ml day, 50 ml overnight and 0.5 l expression cultures were grown with 50 µg/ml methionine. Once the OD of the 0.5 l culture reached 0.4, the culture was gently spun down and the cells were resuspended in fresh M9 minimal medium without any methionine. The culture was allowed to grow for 2 hours, during which the remaining methionine was used up. At this point, the culture was supplemented with 50 µg/ml selenomethionine, incubated for 30-60 minutes to allow the cellular intake of selenomethionine and the recombinant protein expression was induced by the addition of 0.4 mM IPTG. The rest of the expression was performed as for the native protein. The incorporation of selenomethionine was confirmed by LC/MS analysis.

2.3.1.5 Wild-type IdmO and AT mutants

Recombinant IdmO and all IdmO AT mutants were expressed in either *E. coli* BL21(DE3), *E. coli* Lemo21(DE3) or *E. coli* NiCo21(DE3) cells harbouring the appropriate IdmO plasmid as well as pCDF_duet_SFP plasmid to produce SFP transferase required to generate phosphopantetheinylated proteins. Transformed bacteria were cultured at 37 °C until OD_{600} reached 0.6 in 0.5 l volume of 2TY medium supplemented with 50 µl/ml kanamycin required for the selection of pET28_IdmO plasmid and 100 µl/ml ampicillin required for the selection of pCDF_duet_SFP plasmid. In case of *E. coli* Lemo21(DE3), the culture medium was also supplemented with chloramphenicol to maintain pLemo system and 0-2000 µM of L-rhamnose to tune the expression. Once the OD_{600} reached 0.6, the cell culture was cooled down to 15 °C, supplemented with 0.1 mM IPTG and allowed to express for 18-20 hours with shaking of 200 rpm. The cells were harvested by centrifugation at 6,000 *g* at 4 °C for 15 min and stored at -80 °C until use.

2.3.1.6 ACP_9

Recombinant codon-optimised ACP_9 was expressed in *E. coli* BL21(DE3) cells harbouring the pET_SUMO_ACP_9 as well as pCDF_duet_SFP plasmid to produce SFP transferase required to generate *holo*-ACP_9. Transformed bacteria were cultured at 37 °C until OD_{600}=0.6 in 0.5 l volume of 2TY medium supplemented with 50 µ/ml kanamycin required for the selection of pET_SUMO_ACP_9 plasmid and 100 µl/ml ampicillin required for the selection of pCDF_duet_SFP plasmid. Once the OD_{600} reached 0.6, the cell culture was cooled down to 15 °C, supplemented with 0.1 mM IPTG and allowed to express for 18-20 hours with shaking of 200 rpm. The cells were harvested by centrifugation at 6,000 *g* at 4 °C for 15 min and stored at -80 °C until use.

2.3.2 Purification of recombinant proteins

2.3.2.1 General notes on cell disruption

Appropriate amount of pellet was removed from -80 °C storage and resuspended in a suitable buffer containing EDTA-free protease inhibitors (Roche). The suspension was incubated with lysozyme and DNAse for 30-60 minutes at room temperature while on a roller to weaken the cellular membranes and reduce the viscosity. In small-scale purifications, the cells were disrupted using sonication (10 pulses for 10 seconds, 20 seconds break in between). If the cell suspension volume was larger than 25 ml, a high-pressure press (Constant Systems) was used instead. In this case, a homogenised sample was applied to the high-pressure press set at 30 kpsi in continuous running mode. The lysate was passed twice and cleared by centrifugation at 50,000 *g* at 4 °C for 45 minutes.

2.3.2.2 Native IdmH and all IdmH variants

The cell pellet was resuspended in buffer A (50 mM Tris-HCl pH 7.4, 0.5 M NaCl) containing 20 mM imidazole, 0.1 mg/ml lysozyme, 0.05 mg/ml DNAse, 2 mM $MgCl_2$ and EDTA-free protease inhibitors. After disrupting the cells with either a sonicator or a high-pressure press, the cell debris was removed by centrifugation at 50,000 *g*.

The supernatant was loaded onto a 5 ml HisTrap HP column (GE Healthcare, USA) which had been pre-equilibrated with buffer A. Following sample loading, the column was washed with 10 volumes of buffer A before the target protein was

eluted from the column with buffer A supplemented with 200 mM imidazole. The eluted recombinant protein was desalted into buffer B (10 mM Tris-HCl pH 7.2, 50 mM NaCl) using a PD-10 desalting column (GE Healthcare, USA) and the hexahistidine tag was then removed by incubation of the sample with 1 U of thrombin (GE Healthcare, USA) per mg of recombinant IdmH, overnight at 4 °C. Thrombin was removed by loading the protein mixture onto a 5 ml HiTrap Q column (GE Healthcare, USA) equilibrated with buffer B. The column was washed with 5 column volumes of buffer B before applying a linear gradient of increasing NaCl concentration to elute bound proteins. 2.5 ml fractions were collected across the gradient which were analysed by sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Fractions containing the recombinant protein were pooled, concentrated and applied onto a size-exclusion chromatography column (HiLoad 26/600 Superdex 75 pg, GE Healthcare) pre-equilibrated with buffer B. After a void volume of 40 ml 1 ml fractions were collected. These were once more analysed using SDS-PAGE and the fractions containing the most highly pure IdmH were combined to form the final sample. This was concentrated to 20 mg/ml (as determined by the A_{280} and extinction coefficient = 17,990) using a centrifugal filter concentration device (Vivaspin 20; 10000 MWCO). Finally, samples were flash frozen in liquid nitrogen and stored at -80 °C until use.

2.3.2.3 IdmO, IdmO AT mutants and ACP_9

Cell pellets were resuspended in 20 mM sodium phosphate pH 7.4, 500 mM NaCl, 45 mM imidazole, 10% glycerol, 5 mM β-mercaptoethanol (wash buffer) containing 0.1 mg/ml lysozyme, 0.05 mg/ml DNase, 2 mM $MgCl_2$ and EDTA-free protein inhibitors. Cells were disrupted with a high-pressure press, cleared, and the supernatant was loaded onto a 5 ml HisTrap column which had been pre-equilibrated with the wash buffer. Proteins were eluted with wash buffer supplemented with 200 mM imidazole.

For protein preparations involving the overnight cleavage step, elution fractions containing the recombinant IdmO or SUMO-tagged ACP_9 were pooled and desalted into 20 mM sodium phosphate pH 7.4, 100 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol buffer using dialysis tubing or PD-10 desalting columns. The tag (hexahistidine tag for IdmO or SUMO tag for ACP_9) was then removed by incubation of the sample overnight at 4 °C with either TEV or SUMO protease, respectively. IdmO cleavage products were passed through an equilibrated 5 ml

HisTrap column to remove any uncleaved protein and hexahistidine containing peptide.

For IdmO grown in *E. coli* NiCo21(DE3) cells, the eluate from HisTrap step was applied to a gravity column filled with chitin beads (NEB) and the suspension was incubated on a roller at room temperature for 30-45 minutes. Following the incubation, the unbound flow-though was collected for further purification by size-exclusion chromatography.

IdmO and ACP_9 protein samples lacking the purification tag or post-chitin column were concentrated and further purified with either a HiPrep 16/600 Superdex S200 pg column (IdmO) or HiPrep 16/600 Superdex S75 pg column (ACP_9) in 25 mM HEPES pH 7.4, 200 mM NaCl. For recombinant proteins to be used in an AT activity assay, the size-exclusion buffer was supplemented with 10% (*v/v*) glycerol. The elution fractions were analysed using SDS-PAGE and the fractions containing the most highly pure recombinant protein were combined, flash frozen in liquid nitrogen and stored at -80 °C until use.

2.3.3 Polyacrylamide gel electrophoresis

To perform reducing sodium dodecyl sulphate polyacrylamide gel electrophoresis (reducing SDS-PAGE), protein samples were mixed with either 2x loading buffer (200 mM Tris-HCl, pH 6.8, 20% (*v/v*) glycerol, 8% (*w/v*) SDS, 0.4% (*w/v*) bromophenol blue, 20% (*v/v*) β-mercaptoethanol) or 4x Laemmli Sample Buffer (Bio-Rad) and denatured. For samples containing whole cells, the denaturation step was performed for 10 minutes at 95 °C; for purified proteins, the denaturation was performed for 10 minutes at 70 °C. 10% and 15% gels were prepared as outlined in Table 2-5. IdmO samples were analysed using 4-15% gradient Mini-PROTEAN® TGX™ precast protein gels (Bio-Rad).

Broad range colour protein standard (P7712, NEB) was used to permit estimation of the size of the protein bands. Gels were run at 200 V in SDS-PAGE running buffer (25 mM Tris, 250 mM glycine, 0.1% (*w/v*) SDS, pH 8.3) for approximately one hour. The gels were stained with InstantBlue (Expedeon).

Anti-His western blot was performed using HisProbe™-HRP conjugate (Thermo Fisher Scientific) according to the manufacturer's guidelines.

Table 2-5 Polyacrylamide gel ingredients.

Ingredient	10% separating gel	15% separating gel	4% stacking gel
dH_20	3.97 ml	2.3 ml	3.97 ml
1.5 M Tris-HCl pH 8.8	2.5 ml	2.5 ml	-
0.5 M Tris-HCl pH 6.9	-	-	1.25 ml
10% (*w/v*) SDS	100 µl	100 µl	50 µl
Acrylamide/Bis (30% / 0.8% *w/v*)	3.34 ml	5 ml	650 µl
10% (*w/v*) Ammonium Persulphate	62.5 µl	62.5 µl	50 µl
TEMED	12.5 µl	12.5 µl	12.5 µl

2.3.4 Protein concentration determination

Protein concentration of IdmH, IdmO and their variants was determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific). An appropriate sample buffer was used to take a blank reading. A 3 µl aliquot of the protein solution was applied to the sample pedestal to take an A_{280} reading. The absorbance value was then used in the Beer-Lambert equation: $A = \varepsilon \times C \times l$ (ε = extinction coefficient ($M^{-1}cm^{-1}$), C = concentration (M), l = path length (cm)) (Gasteiger et al., 2005). Extinction coefficient of IdmH at 280 nm is 17,990 M^{-1} cm^{-1} and molecular mass is 17893.6 Da. Extinction coefficient of IdmO at 280 nm is 205,230 M^{-1} cm^{-1} and molecular mass is 229,499 Da.

Reliable protein concentration determination using absorbance at 280 nm relies on protein containing some tyrosine and tryptophan residues (which absorb UV light at 280 nm). Since ACP_9 protein contains only one tyrosine and zero tryptophan residues, protein concentration was estimated using an alternative Bradford method (Bradford, 1976). In this assay, the protein concentration is proportional to a colour change of the Coomassie brilliant blue dye. The dye binds to mostly basic and aromatic amino acids, which causes an absorbance maximum shift from 465 to 595 nm (Compton and Jones, 1985). A calibration curve for bovine serum albumin (BSA) is then used to deduce the concentration of protein of interest. To perform the assay, Quick Start Bradford dye reagent (Bio-Rad) was used according to the manufacturer's guidelines.

2.4 Biochemical characterisation

2.4.1 Mass spectrometry

All mass spectrometry (MS) analysis was performed by either Dr James Ault or Rachel George (Astbury Centre for Structural Molecular Biology, University of Leeds, UK).

For accurate molecular mass measurement by LC/MS, the samples were submitted at 20 µM concentration with a minimum volume of 20 µl. For peptide fingerprinting, proteins were run on 5-15% SDS-PAGE gel and stained for 5-15 minutes with InstantBlue (Expedeon). Gel bands were excised and digested with trypsin for 18 hrs. Peptides were identified by performing LC-MS/MS and searched against the UniProtKB/SwissProt database modified to include the amino acid sequence of IdmO and ACP_9.

2.4.2 Static light scattering

IdmH buffer (20mM Tris, 50 mM NaCl) pH was adjusted to 4.5, 5.0, 5.5, 6.0, 6.5, 7.0 and 7.5. 25 mg/ml IdmH stock solution was mixed with buffer in 100 µl reactions (5 µl protein and 95 µl buffer). 10 µl from each reaction were inserted into capillary tubes and placed in the OPTIM® machine (Avacta). Scans were performed from 15 to 100 °C in 1 °C increments.

2.4.3 Circular dichroism

Circular dichroism (CD) spectra (190-260 nm) were recorded using 200 µl protein solution (concentration approximately 0.2 mg/ml) in a 1 mm path-length cuvette. CD experiments were conducted at 20 °C on an APP Chirascan CD spectropolarimeter.

2.4.4 Acyltransferase activity assay

AT activity assay was carried out using an α-ketoglutarate dehydrogenase (αKGDH) coupled assay system, where the CoASH produced by IdmO acyltransferase domain is turned over in the presence of α-ketoglutarate, thiamine pyrophosphate (TPP) and NAD^+ to yield succinyl-CoA, CO_2 and NADH (Figure 2-1). The release of CoASH (and the consumption of malonyl-CoA) was then calculated from the increase in NADH fluorescence.

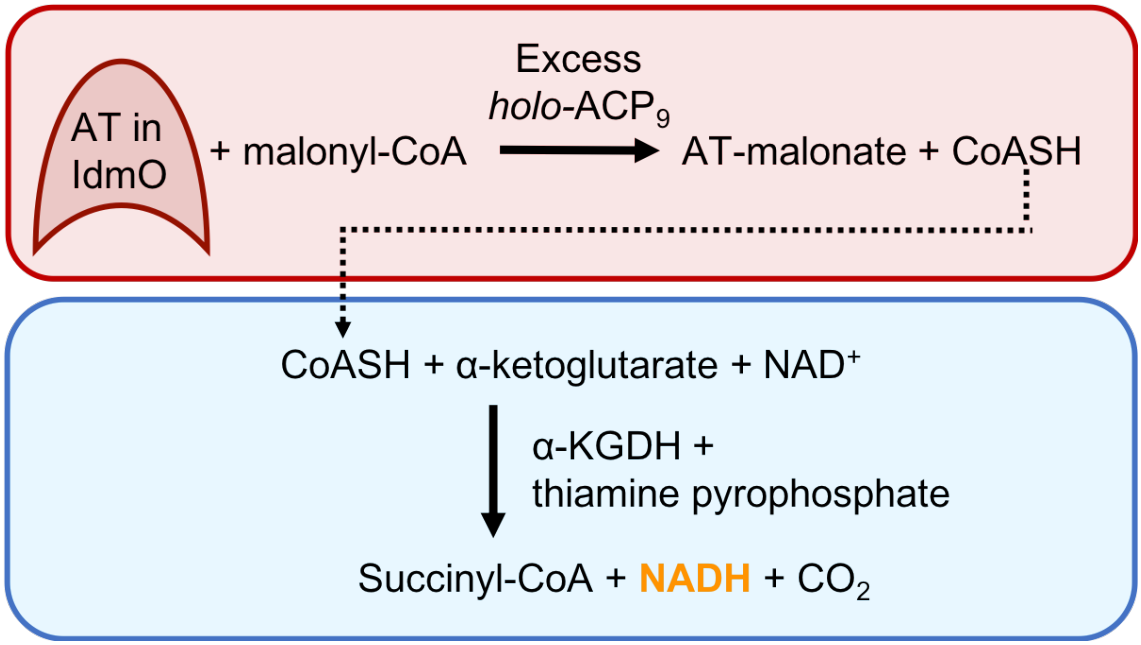


Figure 2-1 Coupling of AT domain catalysis to αKGDH activity via a sensitive fluorometric assay. *In this assay, the CoASH, released by the AT upon forming an AT-malonyl complex, is turned over by a coupled enzyme αKGDH. One of the resulting products of the coupled reaction is NADH, the production of which can be measured fluorometrically (Dunn et al., 2013).*

Assay was performed in 100 µl reaction volume in 96-well black polystyrene plates. NADH fluorescence was monitored using a FLUOstar Galaxy (BMG Labtech) microplate reader. Samples were illuminated with a tungsten light source and a 360 nm filter and fluorescence emission was monitored using a 460 nm filter. All measurements were made with a gain setting of 110. Reactions were run at room temperature for five minutes using the minimum interval between measurements.

TPP, NAD^+ and α-ketoglutaric acid stocks were prepared daily in a buffer containing 50 mM HEPES pH 7.4, 300 mM NaCl, 1 mM DTT and 1 mM EDTA. All but IdmO ingredients were mixed together and incubated for two minutes at 30 °C to allow any free coenzyme A (CoA) to be consumed. The reactions were started by adding IdmO. Final assay concentrations were as follows: 50 mM HEPES pH 7.4, 1 mM DTT, 1 mM EDTA, 0.4 mU/µl αKGDH, 0.4 mM NAD^+, 0.4 mM TPP, 2 mM α-ketoglutaric acid, 50 nM IdmO, 20 µM ACP and 10 µM malonyl-CoA. Table 2-6 outlines single reaction component volumes and concentrations.

Table 2-6 AT assay components.

Solution	Stock concentration	Volume for a 100 µl reaction	Working concentration
αKGDH	400 mU in 100 µl	10 µl	40 mU/100 µl
NAD^+	16 mM	2.5 µl	0.4 mM
TPP	16 mM	2.5 µl	0.4 mM
α-ketoglutaric acid	40 mM	5 µl	2 mM
***Holo*-ACP_9**	200 µM	10 µl	20 µM
Buffer	-	34 µl	-
Substrate	100 µM	10 µl	10 µM
IdmO (AT)	217 nM (0.05 mg/ml)	23 µl	50 nM

To estimate the kinetic parameters of IdmO, the above reaction was set up using a range of substrate concentrations (0-100 µM), and the increase in the NADH fluorescence absorbance was measured over the course of ten minutes. The raw data were plotted in Microsoft Excel and the linear portions of NADH fluorescence progression curves used to calculate the initial rates (µM NADH/min). The initial rates were then converted to specific activities (µM NADH/min/µM IdmO), and the kinetic parameters, K_M and V_{max}, were calculated by fitting the specific activities obtained at different substrate concentrations to the Michaelis-Menten equation using Origin Pro (OriginLab® Data analysis and graphing software).

2.4.5 SEC-MALLS

Size-exclusion chromatography - multi angle laser light scattering (SEC-MALLS) analysis was performed by Dr Maren Thomsen (Astbury Centre for Structural Molecular Biology, University of Leeds, UK). Estimation of IdmO molecular mass and oligomeric state was performed using a Shimadzu high pressure liquid chromatography (HPLC) system equipped with a HELEOS 1184 H28 light scattering instrument. 30 µl of 0.5-1.0 mg/ml protein was injected on a WYATT WTC-MP030S5 column and run at 1 ml/min rate.

2.5 Nuclear magnetic resonance spectroscopy

2.5.1 NMR sample preparation

^{15}N labelled protein was prepared by growing *E. coli* in M9 minimal medium using $^{15}NH_4Cl$ as the sole nitrogen source. Triple-labelled protein was obtained by preparing the M9 medium in D_2O and using ^{13}C D-glucose and $^{15}NH_4Cl$ as the sole sources of carbon and nitrogen. More details of labelled-IdmH expression and purification can be found in sections 2.3.1.3 and 2.3.2.2.

2.5.2 Experimental set up

950 MHz Bruker Ascend Aeon™ and 750 MHz Oxford NMR spectrometers equipped with cryogenically cooled triple-resonance probes were used for all NMR experiments. Spectra were recorded at 25 °C in NMR buffer (20 mM Tris pH 7.3, 50 mM NaCl, 1 mM DTT, 0.02% (*w/v*) NaN3 and 5% (*v/v*) D_2O). 2D NMR experiments were carried out using Wilmad® NMR tubes (precision, limit 600 MHz frequency) while Shigemi NMR tubes were used for triple-resonance experiments. All TROSY spectra were recorded at a field strength of 750 MHz. All but HNcaCO triple-resonance spectra were also acquired at a field strength of 750 MHz while HNcaCO spectrum was recorded at a field strength of 950 MHz.

2.5.3 Backbone nuclei resonance assignment

A standard set of 3D backbone resonance assignment spectra (HNCA, HNcoCA, HNCO, HNcaCO, HNCACB and HNcoCACB) using Bruker library pulse sequences (with Watergate water suppression) were recorded for ^{15}N,^{13}C,^{2}H-labelled IdmH at a concentration of 0.8 mM at field strengths of 750 and 950 MHz. Non-uniform sampling (NUS) was used to speed up acquisition time and obtain high-resolution spectra.

Raw spectra were pre-processed with NMRPipe and NMRDraw (Delaglio et al., 1995). Non-uniform sampling data were analysed using MddNMR (Jaravine and Orekhov, 2006; Jaravine et al., 2008). Backbone nuclei resonance assignment was performed using CCPNmr Analysis (Vranken et al., 2005), and AutoAssign (Zimmerman et al., 1997) software.

2.5.4 Chemical shift perturbation studies

^{1}H-^{15}N HSQC-TROSY spectra were collected at IdmH concentration of 0.2 mM with varying concentrations of indanomycin (0-200 µM) at 25 °C at a field strength of 750 MHz. Raw data were pre-processed using TOPSPIN (Bruker) software. NMR spectra were analysed using CCPNmr Analysis software (Vranken et al., 2005). Chemical shift perturbation was calculated from ^{1}H and ^{15}N chemical shift differences using 0.14 as a scaling factor for ^{15}N shift changes (Williamson, 2013).

2.6 X-ray crystallography

2.6.1 Crystallisation screens

Twelve commercially-available crystallisation screens were used during this work: Crystal Screen 1 & 2 (1), Index 1 & 2 (2) and Salt RX (3) from Hampton Research, Wizard 1 & 2 (4) and Wizard 3 & 4 (5) from Rigaku (Emerald), MIDAS (6), Morpheus (7) and PACT (8) from Molecular Dimensions and JCSQ Core I-IV (9-12) from Qiagen. Screens were set up in 3-drop MRC 96-well plates using a dispensing robot (NT8, Formulatrix). Crystals were imaged using SONICC (Second Order Nonlinear Imaging of Chiral Crystals) technology (Formulatrix). Optimisation screens were set up manually in 24-well sitting drop plates and imaged using a light microscope.

2.6.2 Protein crystallisation and X-ray diffraction data collection

Initial crystallization conditions for the purified IdmH and IdmH-Δ96-104 were determined by the sitting-drop vapour diffusion method with the JCSG Core Suites I-IV (Qiagen). The drops were set up at a constant protein concentration (20 mg/ml) at three different protein to crystallisation solution ratios (1:1, 2:1 and 1:2) and were equilibrated against 25 µl of reservoir solution.

Initial wild-type IdmH hits were optimised with respect to pH and precipitant concentration. The final crystals used for data collection were obtained in 0.2 M calcium acetate, 0.1 MES pH 5.5, 18.6% (*w/v*) polyethylene glycol (PEG) 8000. Crystals were cryoprotected by soaking in mother liquor solution containing 20% (*v/v*) glycerol for about 30 seconds before being flash-cooled on CryoLoops (Hampton Research) by plunging into liquid nitrogen. X-ray diffraction data were

collected on beamline I04 at the Diamond Light Source. 2400 helical line scan images were recorded with a 0.1° oscillation angle and exposure time of 0.8 s.

Crystallisation conditions for native and selenomethionine-labelled IdmH-Δ96-104 were obtained by repeating the JCSG Core Suites I-IV (Qiagen) as described above. High-quality crystals grew overnight in a range of conditions containing 3.6-4 M sodium formate and 10-15% (*v/v*) glycerol. Crystals were mounted on CryoLoops (Hampton Research) without additional cryoprotecting agents and flash-cooled in liquid nitrogen. X-ray diffraction data for selenomethionine-labelled IdmH-Δ96-104 were collected on beamline MASSIF-1 at the European Synchrotron Radiation Facility (Bowler et al., 2015; Svensson et al., 2017; Svensson et al., 2015; Nurizzo et al., 2016; Bowler et al., 2016) for which 3600 images were recorded with a 0.1° oscillation angle and exposure time of 0.133 s. X-ray diffraction data for native IdmH-Δ96-104 were collected on beamline I03 at the Diamond Light Source for which 266 helical line scan images were recorded with a 0.15° oscillation angle and exposure time of 0.15 s. All data collection statistics are listed in Table 2-7.

2.6.3 Crystal structure determination, model building and refinement

Diffraction data were indexed and integrated using XDS (Kabsch, 2010) before scaling in AIMLESS and all subsequent processing steps performed in the CCP4i2 suite (Potterton et al., 2018). Following experimental phasing, jelly-body restraints were applied in REFMAC5 in order to reduce phase bias. During later stages of model refinement, both TLS (translation-libration-screw motion) and NCS (non-crystallographic symmetry) parameters were utilised.

The structure of IdmH-Δ96-104 was determined to 2 Å resolution by Single wavelength Anomalous Dispersion (SAD) phasing using the SHELX pipeline followed by density modification using PARROT and automated model building in BUCCANEER. Iterative rounds of manual model building and refinement were then performed in Coot and REFMAC5, respectively (Sheldrick, 2010; Cowtan, 2010; Cowtan, 2012; Murshudov et al., 2011; Cowtan, 2006).

The structure of wild-type IdmH was solved by molecular replacement using a single monomer of IdmH-Δ96-104 as a search model in Phaser (McCoy et al., 2007; McCoy, 2007). Intial R_{work} and R_{free} were 0.31 and 0.34, respectively, and

decreased significantly (to 0.21 and 0.24, respectively) following iterative rounds of manual model building in Coot and refinement using REFMAC5 (Emsley and Cowtan, 2004; Murshudov et al., 2011). Water molecules were added manually in Coot for both the wild-type IdmH and the IdmH-Δ96-104 and the model was validated using MolProbity (Chen et al., 2010).

All structure-refinement statistics are listed in Table 2-7. In crystallography, different R values are used as an indication of data and model quality. Definitions and formulas of R values used in this work can be found in Table 2-8.

Table 2-7 Data collection and refinement statistics. *Values in parentheses are for the highest resolution shell.*

	Se-IdmH-Δ96-104	**Native IdmH-Δ96-104**	**Wild-type IdmH**
Data collection			
Wavelength (Å)	0.966	0.976	0.9795
Space group	*F*23	*F*23	$P12_11$
a, b, c (Å)	152.6, 152.6, 152.6	152.7, 152.7, 152.7	66.7, 103.5, 99.6
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 91.6, 90
Resolution (Å)	76.30-2.20 (2.27-2.20)	76.34-2.00 (2.05-2.00)	51.76-2.70 (2.82-2.70)
R_{merge}	0.118 (4.62)	0.046 (0.669)	0.086 (0.961)
R_{meas}	0.12 (4.697)	0.053 (0.769)	0.097 (1.098)
$R_{p.i.m}$	0.022 (0.846)	0.026 (0.373)	0.046 (0.522)
$\langle I/\sigma(I)\rangle$	20.8 (1.4)	12.7 (1.4)	11.4 (1.4)
CC(1/2)	0.999 (0.812)	0.999 (0.516)	0.998 (0.656)
Completeness (%)	100 (100)	99.9 (99.9)	99.6 (99.5)
Multiplicity	30.6 (30.6)	4.1 (4.1)	4.5 (4.3)
Refinement			
Resolution (Å)	88.10-2.20	88.15-2.00	103.52-2.70
No of reflections all/free	14989 / 726	19980 / 974	37126 / 1832

R_{work}/R_{free}	0.212 / 0.276	0.230 / 0.271	0.211 / 0.245
Total No. of atoms	3707	3704	19690
No. of ligands	0	0	0
No of water molecules	13	27	9
Wilson B factor (Å^2)	48.30	48.44	60.72
RMS deviations			
Bond lengths (Å)	0.016	0.0069	0.0118
Bond angles (°)	1.806	1.147	1.466
Ramachandran plot			
Favoured (%)	98.8	98.8	98.46
Allowed (%)	1.2	1.2	1.47
Outliers (%)	0	0	0.07
Molecules in asymmetric unit	2	2	10
PDB code	6HNL	6HNM	6HNN

5% of the data were set aside for R_{free} calculations.

Table 2-8 Definitions and formulas of R values used in this thesis.

R value	What it measures	Formula
R_{merge} / R_{sym}	Redundancy-dependent data quality	$R = \frac{\sum_{hkl}\sum_j \lvert I_{hkl,j} - \langle I_{hkl}\rangle \rvert}{\sum_{hkl}\sum_j I_{hkl,j}}$
R_{meas}	Redundancy-independent data quality	$R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n} \lvert I_{hkl,j} - \langle I_{hkl}\rangle \rvert}{\sum_{hkl}\sum_j I_{hkl,j}}$
$R_{p.i.m.}$	Precision of averaged intensities/ amplitudes	$R_{p.i.m.} = \frac{\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_{j=1}^{n} \lvert I_{hkl,j} - \langle I_{hkl}\rangle \rvert}{\sum_{hkl}\sum_j I_{hkl,j}}$
R_{work}/R_{free}	Model quality	$R = \frac{\sum_{hkl} \lvert F_{hkl}^{obs} - F_{hkl}^{calc} \rvert}{\sum_{hkl} F_{hkl}^{obs}}$

2.7 Electron microscopy

2.7.1 Negative stain EM

For negative stain EM, 3 µl of sample was adsorbed on glow-discharged Formvar/Carbon on 400 mesh copper grids (Agar Scientific) followed by staining with 1% or 2% uranyl acetate. The specimen was imaged using either a FEI T12 microscope operated at 120 keV or a FEI TF20 microscope operated at 200 keV. Images were recorded at a magnification of ×30,000 on either a Gatan UltraScan 4000 CCD camera (FEI T12) or FEI Ceta CCD camera (FEI TF20) with defocus values ranging from -0.2 to -1.0 µm. Individual particles were picked using EMAN2 software and subsequent 2D classification and model generation were performed using Relion software suite (Scheres, 2012; Scheres, 2015).

2.7.2 Cryo-EM sample preparation

Cryo-EM grids were prepared by adding 3 µl of 0.1-0.5 mg/ml protein solution to a 1.2/1.3 Quantifoil grid that had been glow discharged at 20 mAmp for 60 s before use. The grids were blotted and frozen using a Vitrobot Mark IV (Thermo Fisher Scientific) at 100% humidity and 4 °C with 6 s blot time and a force of 6. Data were collected on a FEI Titan Krios microscope operating at 300 keV and equipped with a Falcon 3EC direct electron detector.

2.7.3 Cryo-EM data acquisition

All cryo-EM data were collected automatically, using automated data acquisition software for single particle analysis workflow, EPU, at 75,000× magnification, corresponding to 1.075 Å/pix sampling. Images were collected using integrating mode, with a total electron dose of approximately 100 $e^-/Å^2$ during an exposure time of 2 s and dose-fractionation into 79 movie frames. Data were automatically collected with a defocus range from -0.75 µm to -3.0 µm.

For KSB-ACP*S data collection, a total of 4267 micrographs were collected over a single 48 hr session. For KSB data collection, a total of 2706 micrographs were collected over a single 24 hr session.

2.7.4 Image Processing

All steps of data processing were done in Relion 2.0 software suite (Scheres, 2012; Scheres, 2015). The movie stacks for each micrograph were corrected for

drift and dose-weighted using MotionCor2 (Zheng et al., 2017) and contrast transfer function (CTF) parameters were estimated using Gctf (Zhang, 2016).

2.7.5 Calculating KSB-ACP*S reconstruction

Particles were manually picked to generate initial 2D classes, which were low-pass filtered and used for automated particle picking, resulting in 1,569,223 particles. These particles were extracted in a 200^2-pixel box. 2D classification process was used to clean the data, with 826,801 particles belonging to well-defined classes and exhibiting clear secondary structure. A reconstruction obtained from this particle stack displayed smearing of the density in one dimension, which is indicative of the lack of some Fourier components due to preferential orientation. To overcome this, 3D classification into three classes was run for 25 iterations with an angular sampling of 7.5°, 150 Å diameter mask, a regularisation parameter T of 4, C2 symmetry operator and a 60 Å low-pass filtered initial “stretched” model. Based on its features, the second class, comprising of 286,101 particles, was selected for auto-refining, which yielded 4.3 Å (C1, no symmetry) and 4.0 Å (C2 symmetry) reconstructions of KSB-ACP*S. These reconstructions were subjected to Relion 2.0 post-processing, which resulted in final maps displaying 3.7 Å and 3.9 Å global resolution (for C1 and C2 symmetry, respectively).

2.7.6 Investigating the heterogeneity and potential ACP binding sites of KSB-ACP*S

To remove more particle images showing carbon edge contaminants or damaged complexes, the 826,801-particle stack was subjected to another round of 2D classification. The resulting 637,978 particle stack was used for 3D classification into three classes (25 iterations with an angular sampling of 7.5°, a regularisation parameter T of 4, no symmetry operator (C1), 150 Å diameter mask and a 60 Å low-pass filtered C2 KSB-ACP*S reconstruction). The resulting three classes displayed significant differences with the first two classes exhibiting a previously unobserved density either above or below the KS domain, and the third class showing a poorer reconstruction of incomplete complex and no evidence of an alternative conformation. Particles from the first class, comprising 225,095 particles, were refined to produce a 4.4 Å reconstruction (C1); however, smearing of the density was observed which indicates preferential orientation effects. 3D

classification into four classes with no alignment was run to alleviate the preferential orientation effects; however it was unsuccessful as all four classes exhibited similar smeared density. The second class, comprising 256,197 particles, was auto-refined to generate a 4.8 Å reconstruction (C1). 3D classification with no alignment of this particle stack revealed four distinct classes. Once again, one class showed a poor resolution map of an incomplete complex and was, therefore, discarded, while the other three classes showed varying levels of separation between the extra density and the KS domain. All three extra-density containing classes were refined separately (C1) and resulted in 5.9 Å, 7.1 Å and 6.5 Å global resolution maps (for classes 2,3 and 4, respectively).

2.7.7 Calculating KSB didomain reconstruction

Automated particle picking was performed using low-pass filtered KSB-ACP*S 2D classes, resulting in 764,119 particles. These particles were extracted in a 200^2-pixel box. Three rounds of 2D classification were used to remove the images of carbon edge and those containing multiple KSB particles which resulted in 202,563 particles belonging to well-defined classes and exhibiting clear secondary structure. A reconstruction obtained from this particle stack displayed smearing of the density in one dimension, which was overcome by performing 3D classification into two classes (performed for 25 iterations with an angular sampling of 7.5°, C2 symmetry operator, 150 Å diameter mask, a regularisation parameter T of 4 and a 60 Å low-pass filtered KSB-ACP*S model). Based on its features, the second class, comprising a total of 99,401 particles, was selected for auto-refining, which resulted in a post-processed map with a global resolution of 4.4 Å (C2 symmetry) or 5.0 Å (C1 symmetry). For accurate comparison to the KSB-ACP*S 3D classes generated using 3D classification with no alignment, an identical procedure was applied to the KSB map as well.

2.7.8 Refinement, model fitting and visualisation

All refinements were performed using fully independent data half-sets (gold-standard refinement), and resolutions are reported based on the Fourier shell correlation (FSC) = 0.143 criterion (Rosenthal and Henderson, 2003; Scheres and Chen, 2012). FSC curves were calculated with a soft mask (5-pixel soft edge). Post-processing and local resolution calculations were performed within Relion 2.0.

KSB model derived from the crystal structure (PDB 4KC5) was rigid-body fitted into the overall maps of KSB-ACP*S and KSB. All maps were visualised in UCSF Chimera (Goddard et al., 2007) and model illustrations were prepared using either PyMOL (https://www.pymol.org) or Chimera.

Chapter 3 Investigating the structure and function of a putative polyketide cyclase IdmH

3.1 Introduction

Polyketide synthases, non-ribosomal peptide synthetases and hybrids of both are megadalton-sized machines responsible for the production of complex natural products. Both systems consist of a range of domains which incorporate a number of starter and extender units to build a linear structure. The resulting linear non-ribosomal peptide or polyketide can then be further tailored by various enzymes including cyclases, oxidases, reductases and methylases to yield a diverse array of bioactive compounds (Olano et al., 2010).

The majority of this thesis focuses on a hybrid NRPS/PKS system responsible for the production of the antibiotic indanomycin **1** (Figure 3-1) (C. Li et al., 2009). The NRPS portion of the pathway is proposed to generate a pyrrole moiety, which is then extended by the sequential addition of malonyl-CoA, methylmalonyl-CoA and ethylmalonyl-CoA building blocks by eleven predicted PKS modules. This biosynthetic pathway forms a linear non-ribosomal peptide-polyketide hybrid natural product **2** (Figure 1-23).

At least two cyclisation reactions are needed to generate the tetrahydropyran and tetrahydroindane rings of indanomycin (Figure 1-23). The former could be installed by the Cyc11 domain in the terminal PKS module while the polyketide is still tethered to the PKS (C. Li et al., 2009), while the indane ring formation in **4** has been postulated to be catalysed by a candidate indane cyclase, IdmH (Rommel et al., 2011) (Figure 3-1a). It is important to note that the predicted end product of indanomycin NRPS/PKS is a polyketide intermediate lacking a dienophile at C19 (Figure 3-1), therefore an *in trans* dehydration step must be occurring at some other point in the pathway (Rommel et al., 2011).

The linear **2** or partially-matured **3** indanomycin (Figure 1-23) intermediate is thought to be released from the assembly line by a hitherto enigmatic eleventh module, which, in contrast to many other modular PKSs studied to date (Tsai et al., 2001; Scaglione et al., 2010), does not contain a typical thioesterase domain (C. Li et al., 2009) (Figure 1-22).

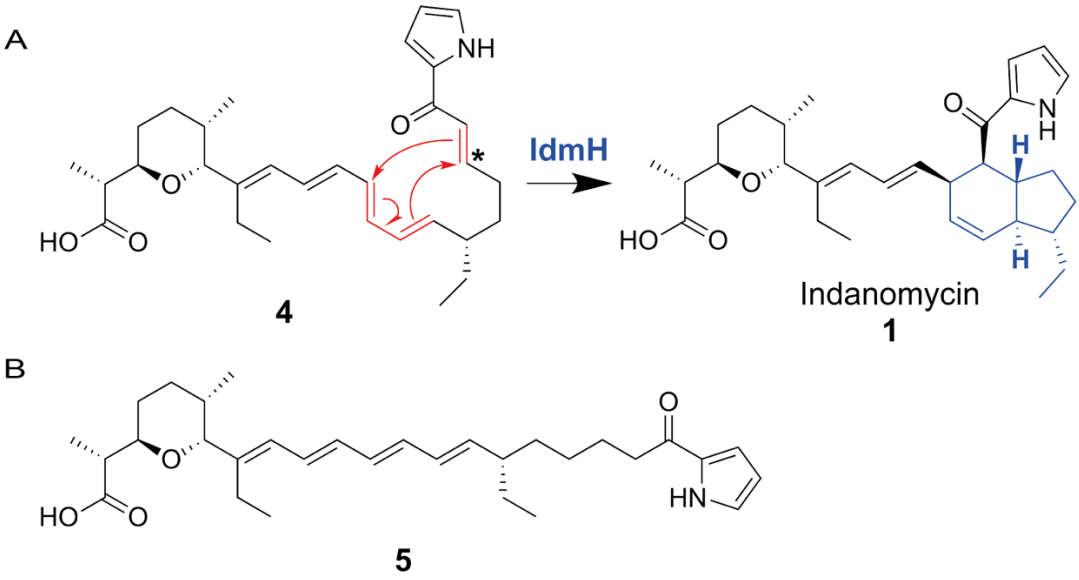


Figure 3-1 Indanomycin maturation. *(a) The putative polyketide cyclase IdmH is thought to catalyse indane ring cyclisation via [4+2] cyclisation reaction mechanism (C. Li et al., 2009).* ***4*** *is not the natural product of indanomycin NRPS/PKS. First the tetrahydropyran ring needs to be installed (for more details, refer to Figure 1-23) and second, the C19 position (denoted with an asterisk) needs to be dehydrated to produce a suitable dienophile. (b) Linear tetraene product extracted from idmH deletion mutant of S. antibioticus (Rommel et al., 2011).*

IdmH is a protein comprising 144 amino acid residues exhibiting sequence similarity to a number of natural product modifying enzymes. These include an epoxide hydroxylase MonBI from *Streptomyces cinnamonensis* (Minami et al., 2013) and a polyketide cyclase SnoaL from *Streptomyces nogalater* (Sultana et al., 2004). Both homologues exhibit an α + β barrel fold and are involved in the ring closure and aromatic ring hydroxylations in the monensin and nogalamycin biosynthetic pathways, respectively. Beyond polyketide modifying enzymes, additional proteins that display the α + β barrel fold include Δ5-3-ketosteroid isomerase (Ha et al., 2001), and scylatone dehydratase (Lundqvist et al., 1994), which demonstrates that nature has made broad use of this fold in many biosynthetic pathways.

Given the diversity of reactions catalysed by enzymes that display the α + β barrel fold, it is challenging to predict the function of IdmH based on sequence similarity alone. The Kelly lab (Rommel et al., 2011), however, produced an *idmH* deletion mutant of *S. antibioticus*, which exhibited both a significant reduction in

indanomycin yield and the production of a previously unobserved linear tetraene alternative product (Figure 3-1b). Notably, upon the *in trans* expression of IdmH in the *S. antibioticus idmH* deletion strain, the indanomycin production levels were restored back to those comparable to the wild-type (Rommel et al., 2011). IdmH is thus thought to be the critical cyclase responsible for the formation of the indane ring and hence mature indanomycin.

The formation of the indane ring is postulated to progress via a [4+2] intramolecular cyclisation, also termed a Diels-Alder reaction (Figure 3-1). The feasibility of indane ring formation via this reaction was first shown by the chemical synthesis of indanomycin (M.P. Edwards et al., 1984). The Diels-Alder reaction is widely used in synthetic organic chemistry; however, a natural Diels-Alderase that will function in aqueous conditions has been sought after for many years to facilitate a greener approach to this reaction (H.J. Kim et al., 2012). A number of putative and computationally designed Diels-Alderases have recently been reported (Linder et al., 2012; Kasahara et al., 2010; Klas et al., 2015; Serafimov et al., 2007; Serafimov et al., 2008; Auclair et al., 2000), with two such enzymes in particular, AbyU from the abyssomicin C biosynthetic pathway and Pyri4 from the pyrroindomycin biosynthetic pathway, characterised in molecular detail (Q. Zheng et al., 2016; Byrne et al., 2016).

This chapter describes the steps taken towards the elucidation of the underlying enzymatic mechanism of indanomycin maturation, namely the generation of recombinant IdmH and its truncated variant IdmH-Δ96-104, and the determination of their crystal structures.

3.2 Construction of plasmid vector harbouring the *idmH* gene

The *idmH* gene (ACN69984.1) was PCR-amplified from the genomic DNA of *Streptomyces antibioticus* NRRL8167. Primers (Table 2-4) were designed based on the gene sequence to insert DNA overhangs on both sides of the amplicon so that they would specifically anneal to a NdeI-linearised pET28a(+) vector using the Gibson Assembly method (Gibson et al., 2009) (Figure 3-2). The assembly reaction mix was used to transform *E. coli* XL10 ultracompetent competent cells.

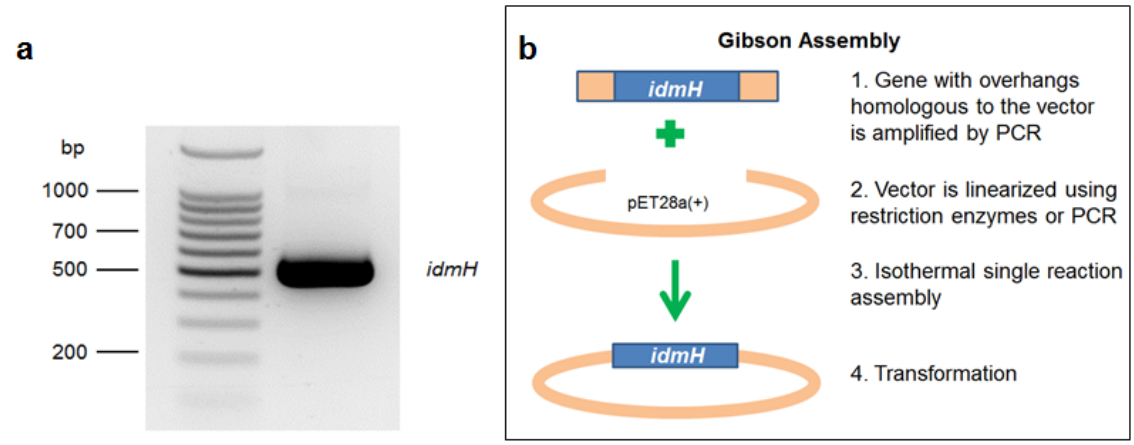


Figure 3-2 Cloning of the idmH gene. *(a) Amplification of the 478 bp idmH PCR product from S. antibioticus genomic DNA. (b) Schematic representation of idmH assembly into pET28a(+) vector. The Gibson isothermal, single reaction method involves three different enzymes: 5' exonuclease removes nucleotides from the 5' ends of double-stranded DNA molecules, then complementary single-stranded DNA overhangs anneal, DNA polymerase fills the gaps, and finally, DNA ligase seals the nicks.*

Single colonies were grown in 5 ml 2TY overnight cultures, and the resulting plasmid DNA was extracted and sequenced using the universal primers T7P and T7TERM (Table 2-4). A 25% (*v/v*) glycerol stock was made of the successful pET28a_IdmH construct.

3.3 Recombinant IdmH expression and purification

The pET28a_IdmH construct was transformed into *E. coli* BL21(DE3) cells for protein overexpression. Protein expression was performed in 2TY medium at both 15 °C and 30 °C. SDS-PAGE analysis of the whole cells revealed that expression at the lower temperature resulted in a significantly higher recombinant protein yield (Figure 3-3a). Figure 3-3b shows successful purification of IdmH by nickel affinity chromatography after expression in 2TY medium at 15 °C.

To improve the homogeneity of the purified protein, fractions containing the highest amount of IdmH were passed through a HiPrep 26/600 Superdex S75 size-exclusion chromatography column (Figure 3-4). According to the calibration data provided by GE Healthcare, a 17.9 kDa protein should elute at approximately 190 ml volume; however, IdmH eluted at around 155 ml (consistent with a 35.8 kDa protein). Therefore, these results suggested that IdmH might be forming homo-dimers.

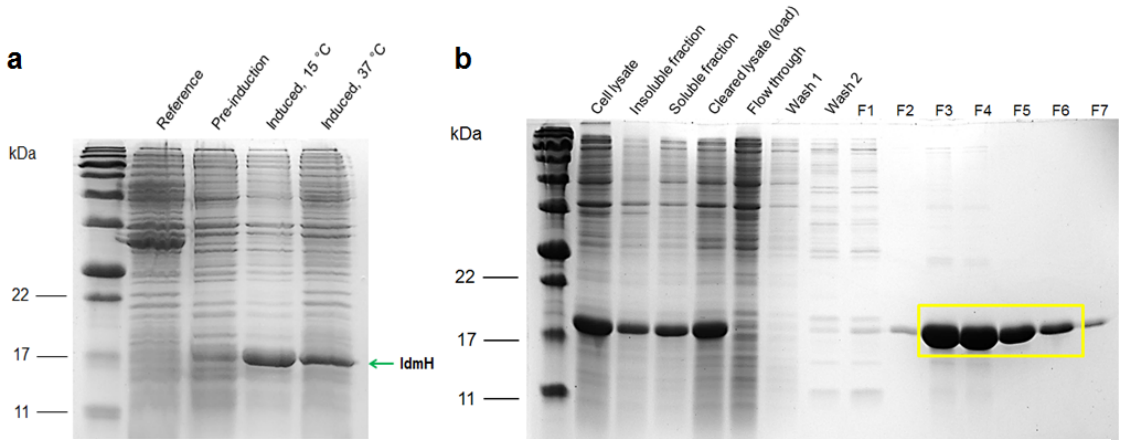


Figure 3-3 Expression and purification of IdmH. *(a) Expression of IdmH in 2TY medium at two different temperatures. The expected mass for IdmH as calculated from the amino acid sequence is 17893.6 Daltons. (b) Purification of IdmH by nickel affinity chromatography. IdmH was eluted in 1 ml fractions using 300 mM imidazole. Highlighted fractions were pooled for size-exclusion chromatography.*

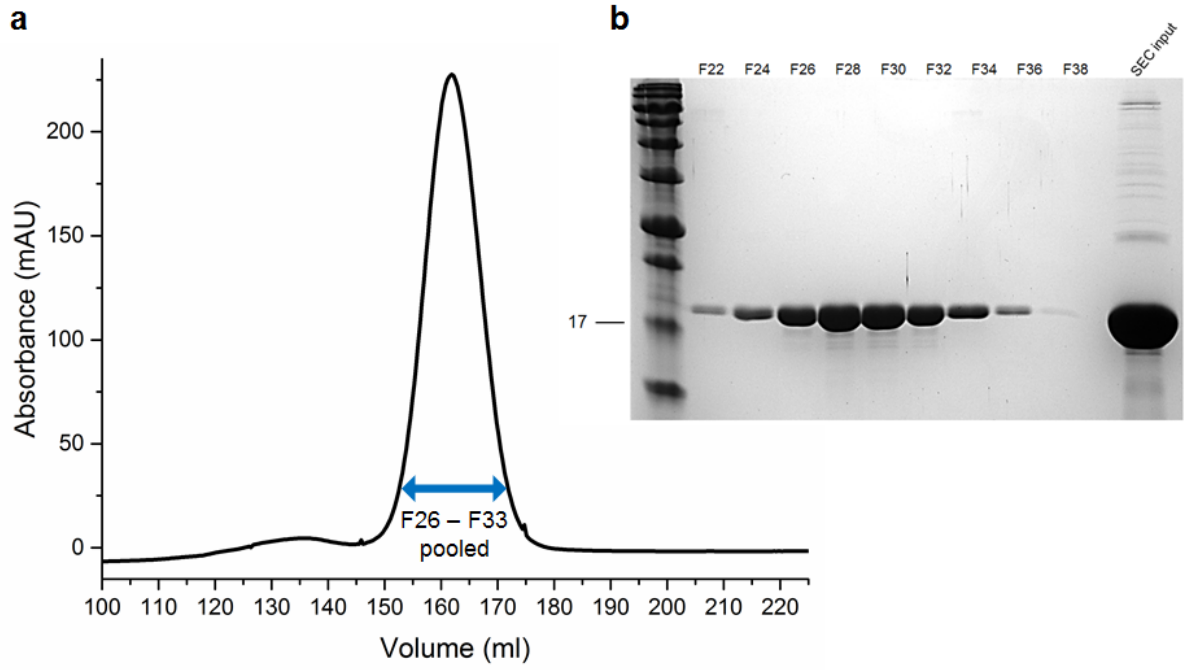


Figure 3-4 Size-exclusion chromatography of IdmH. *(a) Elution profile. Chromatography was performed using HiLoad 26/60 Superdex 200 prep grade column and a flow rate of 2 ml/min. Buffer used contained 10 mM Tris-HCl pH 7.2, 50 mM NaCl. A single peak was observed between 155 and 175 ml. (b) An SDS-PAGE gel of size-exclusion chromatography fractions.*

3.4 Accurate molecular mass and oligomeric state determination by native mass spectrometry

To confirm the molecular mass of the protein, purified IdmH was desalted into 50 mM ammonium acetate pH 7.4, and its mass was measured by native time-of-flight-electrospray-ionisation-mass-spectrometry (TOF-ESI-MS). The monomer had an observed mass of 17892.93 ± 0.13 Da, which is in excellent agreement with the calculated mass of 17893.6 Da. The mass spectrum showed peaks corresponding to monomeric, dimeric and tetrameric species indicating the possibility of enzyme oligomers (Figure 3-5).

To investigate the possibility of dimer formation via a disulphide bridge, IdmH sample was desalted into 50 mM ammonium acetate pH 7.4, 10 mM DTT and native TOF-ESI-MS analysis was repeated (Figure 3-6). The abundance of mass spectrum peaks revealed that the equilibrium shifted towards dimers; however, no tetramers were observed. This suggests that the interaction within the homo-dimer is unlikely to occur via a disulphide bridge.

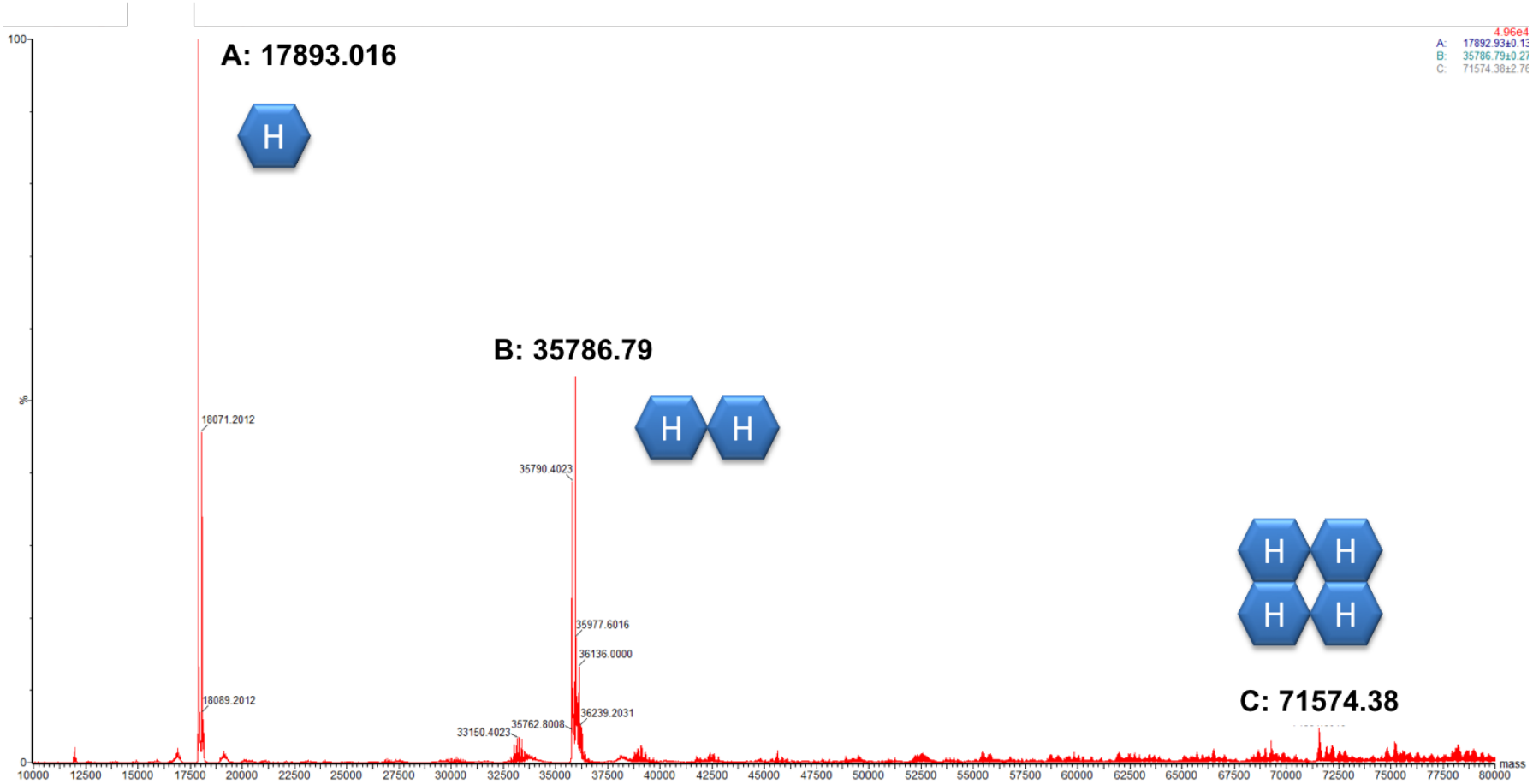


Figure 3-5 Analysis of IdmH oligomerisation state by native TOF-ESI-MS. *Protein was desalted into 50 mM ammonium acetate pH 7.4 before analysis. Here, a deconvoluted spectrum is shown with peaks for monomeric, dimeric and tetrameric IdmH clearly visible.*

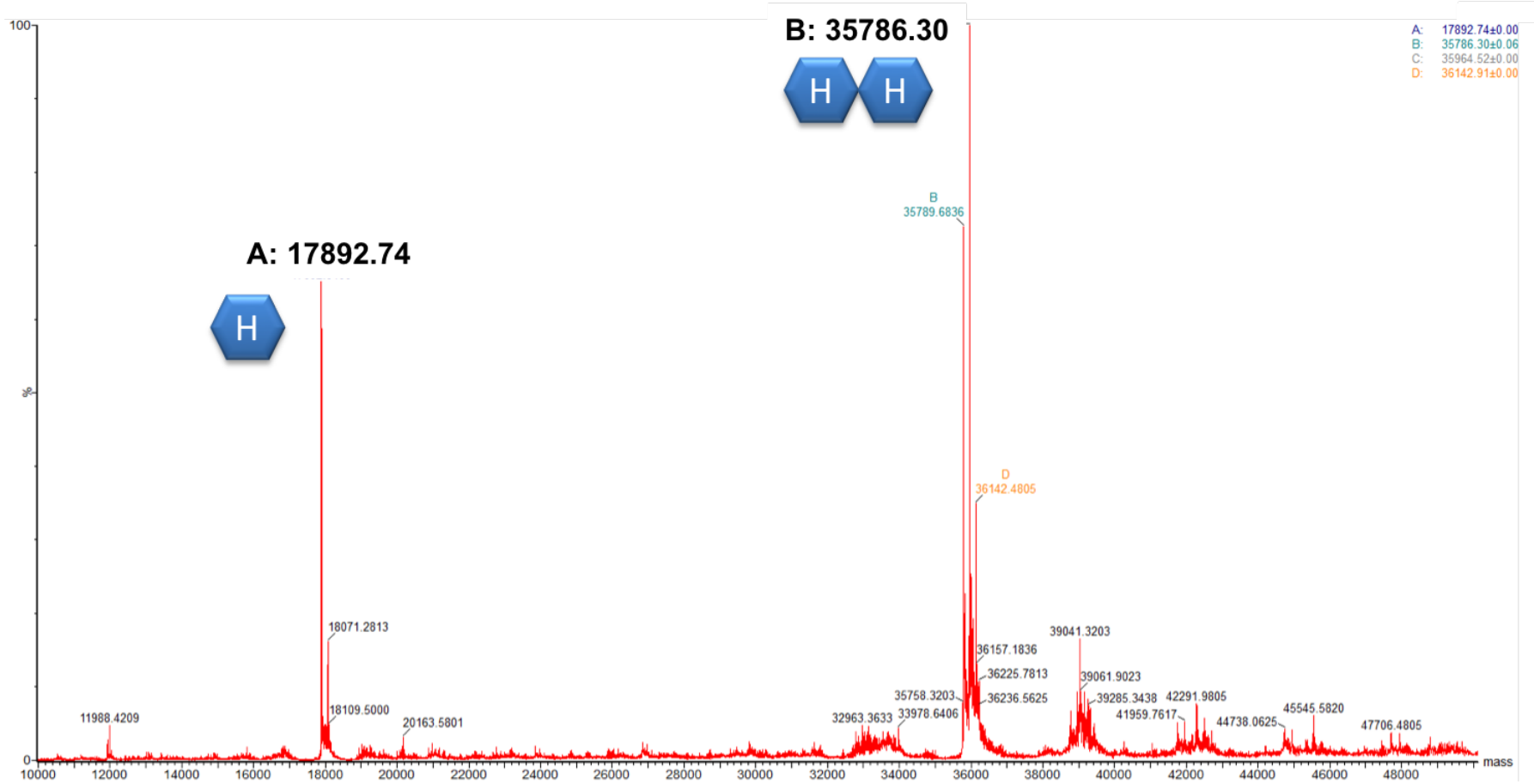


Figure 3-6 Analysis of IdmH oligomerisation state by native TOF-ESI-MS. *Protein was desalted into 50 mM ammonium acetate pH 7.4, 10 mM DTT before analysis. Here, a deconvoluted spectrum is shown with peaks for monomeric and dimeric IdmH clearly visible. The presence of dimeric IdmH peak in these reducing conditions suggests that the dimer interactions are unlikely to occur via a disulphide bridge.*

3.5 Secondary structure prediction by sequence analysis and circular dichroism spectroscopy

The amino acid sequence of IdmH was submitted to the JPred protein secondary structure prediction server (Drozdetskiy et al., 2015). Jpred uses PSI-BLAST to construct a multiple alignment for the protein sequence of interest which is then followed by the prediction of secondary structures using algorithms. Predictions of solvent accessibility and coiled-coil regions are also performed. The results of Jpred prediction, shown in Figure 3-7, suggested that IdmH has an alpha/beta fold with 27% of its amino acids present in four helices, 27% amino acids present in five sheets and the remaining 46% in random coils.

An experimental approach was also taken to estimate the secondary structure of IdmH. Circular dichroism (CD) spectroscopy was used to obtain a far UV (190 – 250 nm) spectrum of IdmH (Figure 3-8). The CD signal (measured as molar ellipticity) originates from the peptide bond excitations of the proteins and can be used to distinguish between different secondary structure elements. The resulting CD spectrum of IdmH resembled a typical α-helical shape; however, the intensity at 200 nm suggested the occurrence of random coils. β-sheet contribution to the spectrum could be seen between 210 and 220 nm.

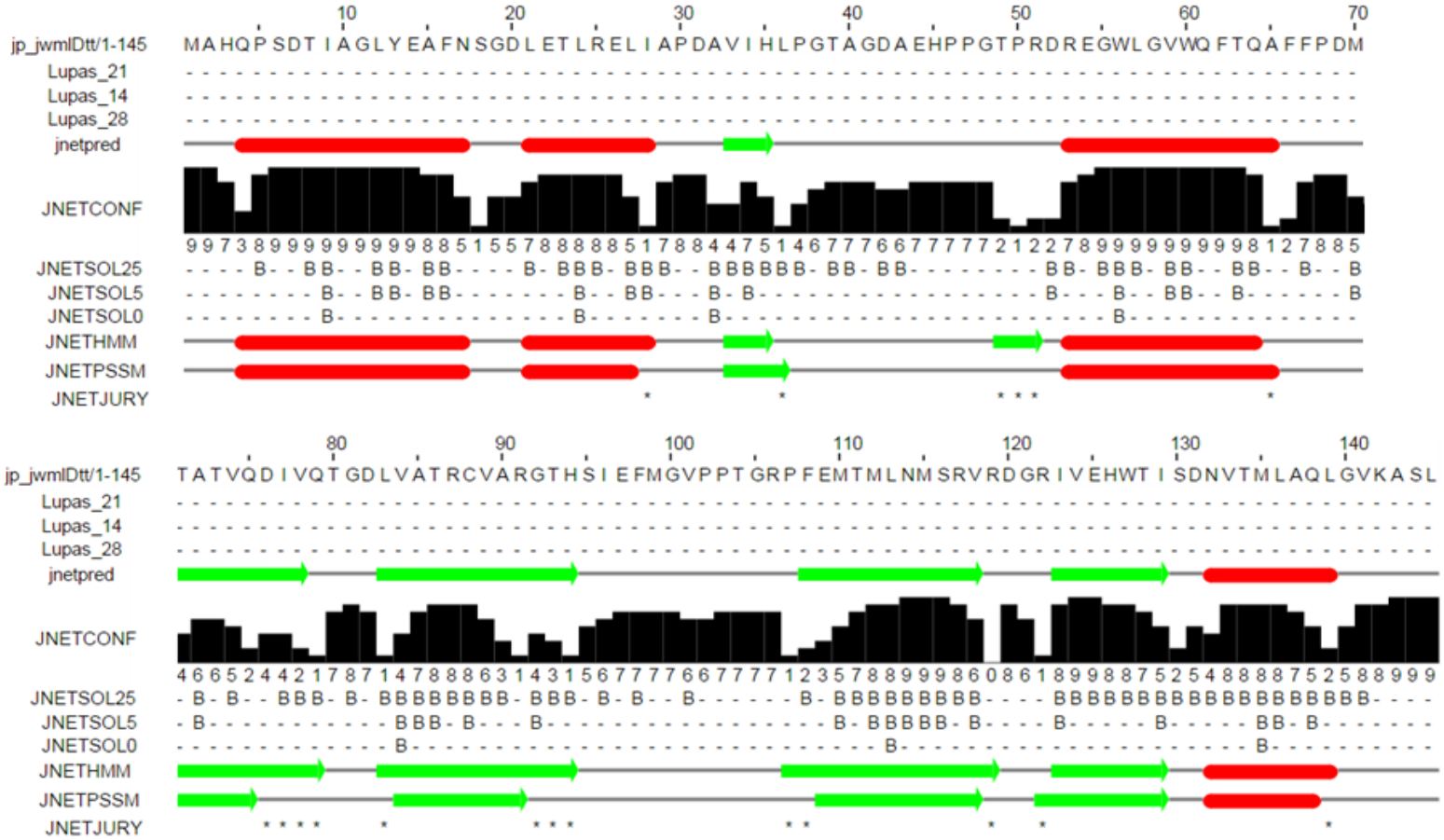


Figure 3-7 Jpred secondary structure prediction for IdmH. *Lupas_21, Lupas_14, Lupas_28 make coiled-coil predictions for the sequence; JNETSOL25, JNETSOL5, JNETSOL0 make solvent accessibility predictions (binary predictions of 25%, 5% or 0% solvent accessibility); JNetPRED is the consensus prediction - helices are marked as red tubes, and sheets as green arrows; JNetCONF is the confidence estimate for the prediction (high values mean high confidence); JNetHMM is hidden Markov Model prediction, and JNETPSSM is PSSM (position-specific scoring matrix) based prediction. An asterisk indicates that the JNETJURY was invoked to rationalise significantly different primary predictions.*

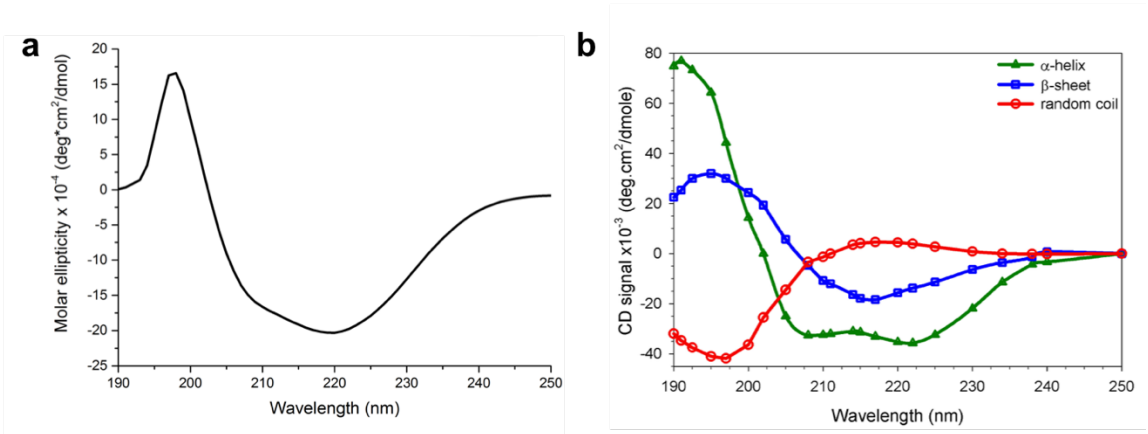


Figure 3-8 Circular dichroism analysis. *(a) CD spectrum of IdmH. Intensity at 190 nm suggests a substantial contribution to the spectrum from random coils, and asymmetrical peaks at 210 and 220 nm suggest β-sheet contribution. (b) Reference spectra for α-helical, β-sheet and random coil folds. Reprinted with permission from Ghulam Khan (personal communication with Ghulam Khan, University of Leeds, 2018).*

DichroWeb is an online tool for quantitative analysis of protein secondary structure content (Lobley et al., 2002; Whitmore and Wallace, 2004). Various methods have been developed which compare the experimental CD spectrum to those of reference proteins. Here, the CONTIN programme, based on the ridge regression algorithm of Provencher and Glockner, was used (Sreerama and Woody, 2000). Results of Jpred and Dichroweb analysis are shown in Table 3-1.

Table 3-1 Circular Dichroism data analysis using CONTIN. *Numbers represent the percentage of all residues in the protein involved in a particular secondary structure fold.*

Result	Helix	Strand	Turns and random coils
Closest matching solution from Dichroweb	10%	38%	52%
Jpred estimation	27%	27%	46%

DichroWeb analysis agrees with the Jpred prediction regarding the β-sheet and random coil content; however, the prediction for α-helices is a lot lower than Jpred (10% compared to 26%) highlighting the level of unreliability associated with secondary structure prediction tools.

3.6 Determination of optimum pH and approximate melting temperature by static light scattering

One of the uses of static light scattering (SLS) technique is to measure protein stability and aggregation. In order to determine optimal buffer pH for structural studies of IdmH, an OPTIM® machine (Avacta) was used to investigate the effects of pH on IdmH aggregation (Figure 3-9).

Concentrated IdmH was diluted into buffers of varying pH and inserted into 10 µl capillary tubes. The capillaries were heated from 15 °C to 100 °C and light scattering data were recorded in 1 °C increments.

The results of the analysis showed that at pH 4.5 significant SLS was present even at low temperatures (15-20 °C). Between pH 5.0 and 7.5, IdmH appeared relatively stable at temperatures up to 30 °C. Finally, the IdmH was shown to be the most stable at pH values higher than 7.0, and to remain folded at temperatures up to 40 °C when the pH values were kept between 7.0 and 7.5

The approximate melting temperature at different pH values can be estimated by measuring the tip of the SLS peak. Therefore, for different pHs, the melting temperature was estimated to range from 44 °C (pH 4.5) and 50 °C (pH 7.5).

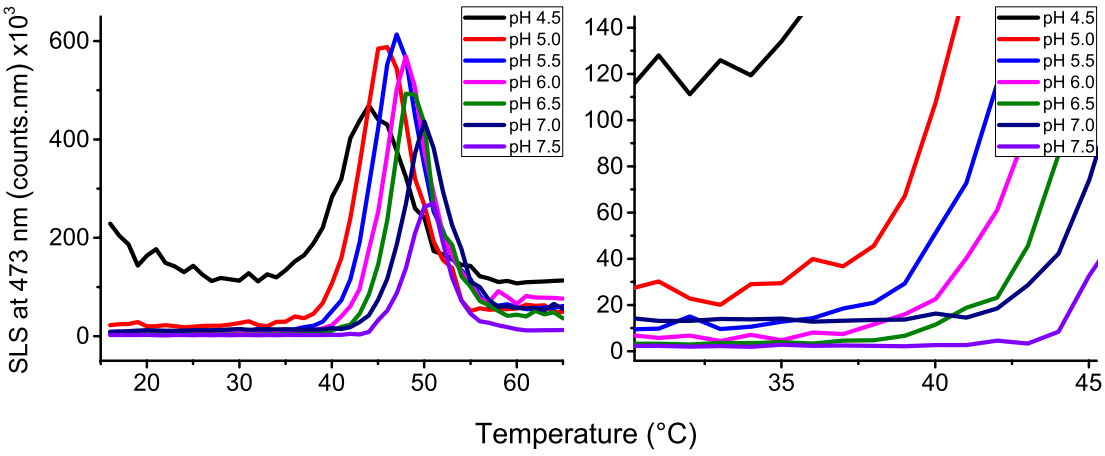


Figure 3-9 Aggregation curves of IdmH at different pHs (left). *The zoomed-in graph can be seen on the right. For this analysis, 25 mg/ml stock of IdmH was mixed with buffers pH 4.5 to 7.5 at a ratio of 1:20. The solution was then inserted into capillary tubes which were heated up to 100 °C. SLS was measured at 1 °C increments.*

3.7 Structural studies by X-ray crystallography

IdmH shows a distant (<30%) sequence similarity to a number of enzymes from other natural product biosynthetic pathways. Notable examples include a polyketide cyclase SnoaL from *Chromobacterium violaceum* (Sultana et al., 2004), putative hydrolases AclR and SnoaL2 from *Streptomyces galilaeus* and *Streptomyces nogalater*, respectively (Beinker et al., 2006) and an aklanonic acid methyl ester cyclase DauD from *Streptomyces* species strain C5 (Dickens et al., 1995). However, they come from different organisms and display considerable amino acid sequence differences to IdmH. Therefore, to investigate the function and the potential enzymatic mechanism of this putative cyclase, we sought first to attempt to determine the three-dimensional structure of IdmH by X-ray crystallography.

3.7.1 First round of commercial crystallisation screens

Eight commercially available crystallisation screens were set up: Crystal Screen 1 & 2, Index 1 & 2 and Salt RX from Hampton Research; Wizard 1 & 2 and Wizard 3 & 4 from Rigaku (Emerald) and MIDAS, Morpheus and PACT from Molecular Dimensions. Several needle-like crystals were observed in multiple crystallisation conditions after one week. Unfortunately, needles growing in only one direction are too thin in other directions to acquire a strong enough diffraction signal and, therefore, are not suitable for data collection.

Approximately five months after the plates had been set up, a larger crystal hit, exhibiting a growth in all three directions, was identified in Midas H6 condition comprising 0.1 MES pH 6.0, 10% (*v/v*) ethanol and 30% (*w/v*) poly(acrylic acid sodium salt) (PAA) 5100 (Figure 3-10a). A full dataset, collected at the Diamond synchrotron, confirmed that it was a protein crystal. Unfortunately, further data processing steps, required to determine the space group and calculate the unit cell parameters, failed. A few possible reasons for this outcome are crystal peculiarities such as pseudo translation or twinning. An optimisation screen was set up around the Midas H6 condition to obtain more uniform crystals; however, this resulted in liquid-like quasi-crystal formation.

Since the hexahistidine tag and a linker adds almost 2 kDa to the mass of IdmH, it was hypothesised that removing this flexible part of the protein might improve the crystal formation. Therefore, the tag was cleaved off using the protease

thrombin and all eight commercially available screens were repeated. This time a rod-like protein hit was identified in Wizard 3&4 G7 condition comprising 0.1M HEPES pH 7.0, 20 mM $MgCl_2$ and 20% (*w/v*) PAA 5100 (Figure 3-10b).

Crystal growth in both conditions appeared to be encouraged by the same precipitant – PAA polymer with an average molecular mass of 5100. Therefore, crystal optimisation plates were set up to screen various commercially available buffer components at pH values 5, 6, 7 and 8 while the PAA concentration was kept similar to the initial hits – at 20%, 25% and 30% (*v/v*). However, no two-dimensional crystals (only needles) were observed in the optimisation screens.

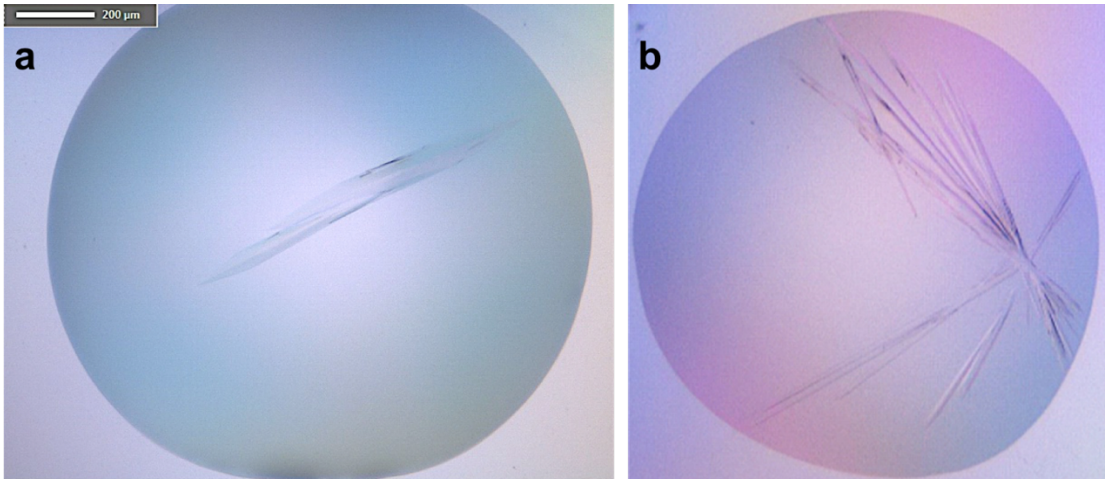


Figure 3-10 IdmH crystal hits identified from the first round of crystallisation screens. *(a) His-tagged IdmH crystal formed in Midas H6 condition (0.1 M MES pH 6, 10% (v/v) ethanol, 30% (w/v) PAA 5100). (b) Tag-less IdmH crystal formed in Wizard 3&4 G7 condition (0.1 M HEPES pH 7, 20 mM MgCl2, 20% (w/v) PAA 5100).*

3.7.2 Second round of commercial crystallisation screens

Since the first round of commercial screens did not result in any promising hits, another set of four commercially available crystallisation screens, designed by Qiagen and the Joint Centre for Structural genomics (JCSG), were set up. The most promising crystals grew in JCSQ Core I condition C5 comprising 0.2 M calcium acetate, 0.1 M MES pH 6.0 and 20% (*w/v*) PEG 8000 (Figure 3-11a). An optimisation screen was set up around the initial conditions and larger, uniform rod-shaped crystals were obtained under condition comprising 0.2 M calcium acetate, 0.1 M MES pH 5.5 and 18.6% (*w/v*) PEG 8000 (Figure 3-11b).

X-ray diffraction data collected to 2.7 Å suggested that these crystals belonged to a monoclinic point group, but attempts to determine the structure by molecular replacement were unsuccessful, presumably due to the lack of a suitable search model and a large number of molecules in the asymmetric unit as estimated using MATTPROB (Kantardjieff and Rupp, 2003).

IdmH was labelled with seleno-methionine to attempt determining the phases using an experimental approach. C5 condition optimisation screens were repeated; however, no crystals suitable for data collection were obtained.

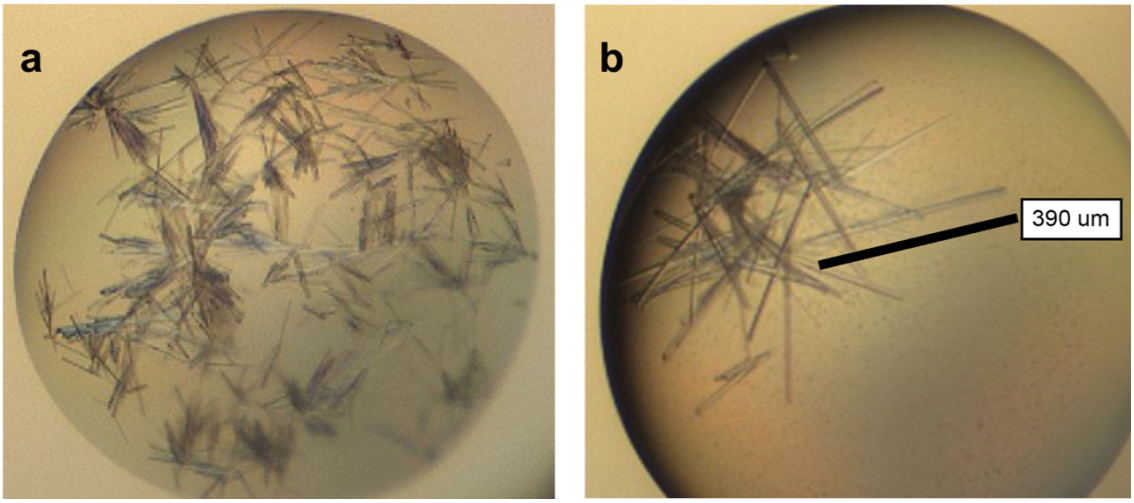


Figure 3-11 IdmH crystals obtained from the second round of crystallisation screens. *(a) IdmH crystals formed in JCSG I C5 condition comprising 0.1 M calcium acetate, 0.1 M MES pH 6, 20% (v/v) PEG 8000. (b) Optimised crystals grown in 0.2 M calcium acetate, 0.1 M MES pH 5.5, 18.6% (v/v) PEG 8000.*

3.7.3 IdmH mutant generation to improve crystallisation

Obtaining high-quality diffracting crystals remains the main bottle-neck in many crystallography projects. One notion of improving the formation of protein crystals involves the reduction of protein surface disorder (also known as surface entropy) which would in result favour the formation of successful intramolecular contacts within the crystal lattice (Goldschmidt et al., 2007). This can be done on two different scales. First, predicted flexible surface loops can be truncated. This method is, however, risky as the structural and functional role of the truncated loop might be underestimated and the resulting truncated polypeptide variant might be rendered insoluble or non-functional. A second, less invasive, technique is to exchange single or multiple flexible surface-exposed residues with those

exhibiting lower conformational entropy. The surface entropy reduction prediction server (Goldschmidt et al., 2007) can be used to predict such residues and, based on an algorithm, design a number of mutations to change the protein of interest's entropy profile.

To improve crystallisation of IdmH, both approaches outlined above were attempted. Four surface entropy reduction and three predicted loop truncation IdmH constructs, outlined in Table 3-2, were generated by a final year undergraduate student, Jana Obajdin (Obajdin, 2018).

Table 3-2 IdmH variants produced and the outcome of crystallisation experiments.

Variant	Justification	Outcome
E29A	Reducing surface entropy	No effect
E47A	Reducing surface entropy	No effect
E57A	Reducing surface entropy	No effect
E100A	Reducing surface entropy	No effect
Residue 46-50 truncation	Decreasing flexibility	Insoluble
Residue 96-104 truncation	Decreasing flexibility	New crystal form
Residue 139-144 truncation	Decreasing flexibility	Insoluble

All seven IdmH variants were overexpressed, purified and screened using the four JCSQ Core I-IV crystallisation screens. In one case, for a truncated IdmH missing residues 96 to 104 (IdmH-Δ96-104), a new crystallisation condition was identified which resulted in large, easily reproducible crystals in the cubic space group F23 (Figure 3-12b).

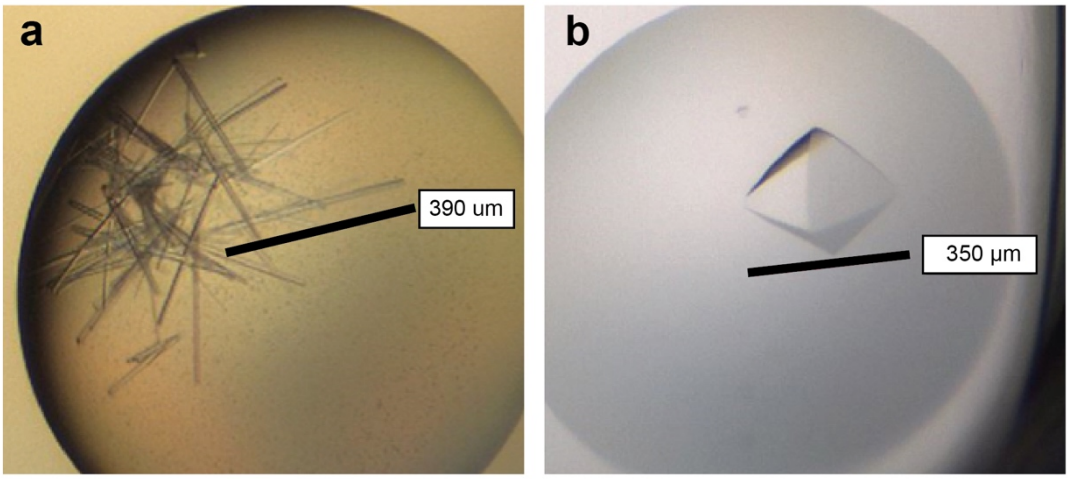


Figure 3-12 Comparison of crystal morphology between IdmH and IdmH-Δ96-104. *(a) Wild-type IdmH crystallised as long rods in the space group P1212. (b) IdmH-Δ96-104 resulted in large tetrahedral crystals with a space group F23.*

3.7.4 Labelling of IdmH with selenium and acquiring single-wavelength anomalous diffraction data

The single-wavelength anomalous diffraction (SAD) method is a dominant approach to solve the structures of proteins exhibiting low homology to other proteins with known structures already present in the Protein Data Bank (PDB) (Berman, 2000). In fact, 73% of such new structures were solved using SAD by 2013 (Terwilliger et al., 2016). Briefly, this method involves X-ray data collection at a certain wavelength, which is at or near the absorption edge of an element, such as selenium or sulfur, present in the macromolecule. The resulting anomalous scattering from the atoms in this substructure allows the localisation of these atoms and subsequent estimation of the crystallographic phases for the entire macromolecule (Sheldrick, 2010; Terwilliger et al., 2016).

In order to facilitate this method for solving the structure of IdmH, IdmH-Δ96-104 variant was labelled with selenium atoms. The incorporation of selenium was achieved by overexpressing IdmH in methionine auxotrophic *E. coli* B834(DE3) cells and supplementing the minimal medium with selenomethionine. Selenium-labelled IdmH-Δ96-104 was purified as described previously and the incorporation of selenium atoms was analysed by LC/MS (Figure 3-13).

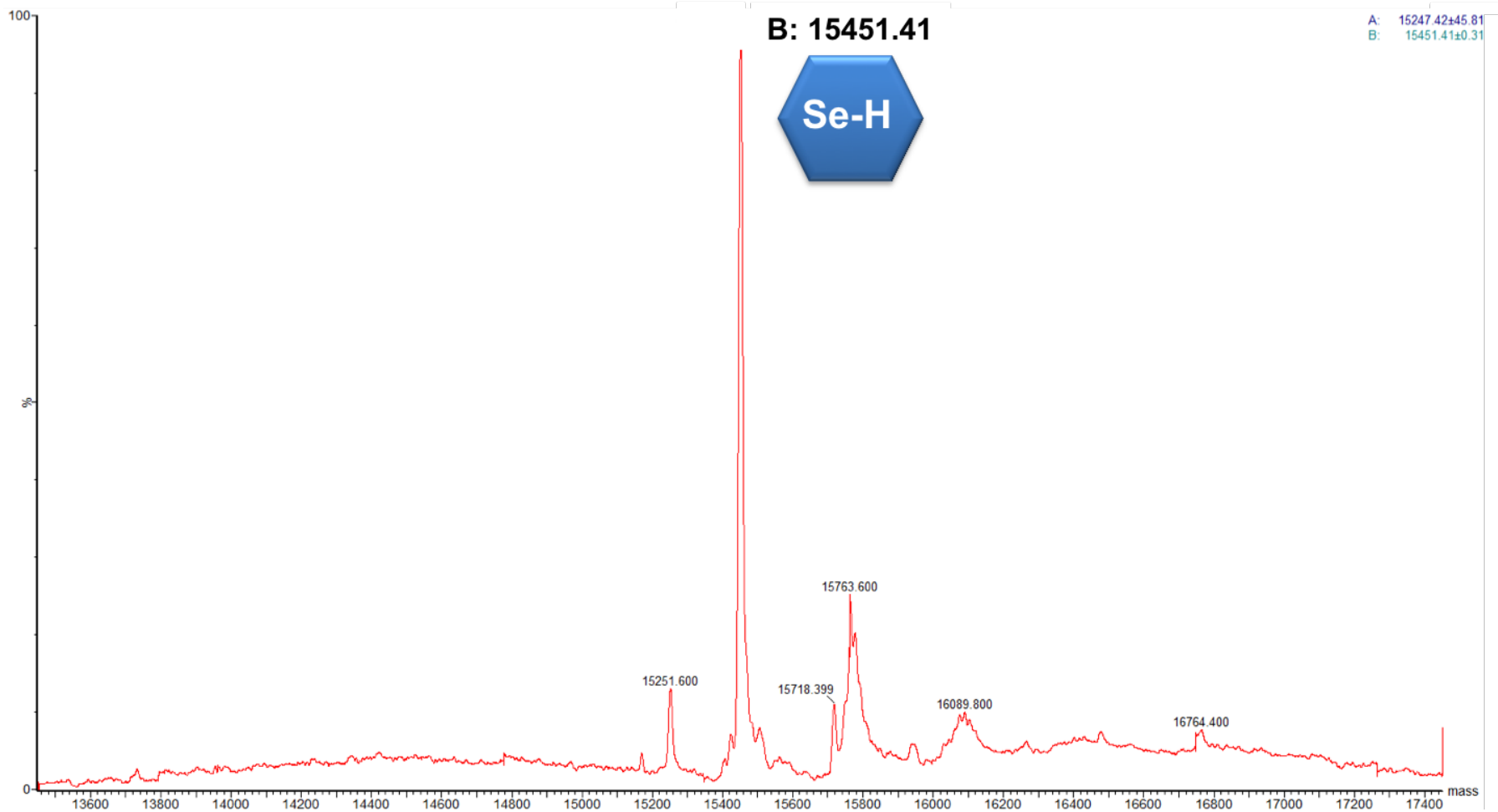


Figure 3-13 LC/MS analysis of selenomethionine-labelled IdmH-Δ96-104. *Here, a deconvoluted spectrum is shown with the largest peak (B) corresponding to the monomeric IdmH-Δ96-104. Peak B shows the observed mass of 15451.41, which is in excellent agreement with the calculated mass of 15451.42 Da, confirming the successful selenomethionine-labelling of IdmH-Δ96-104.*

The molecular mass of native IdmH-Δ96-104 was calculated to be 15170.05 Da, while the molecular mass of the fully selenomethionine-labelled IdmH-Δ96-104 (when all six methionines are replaced with selenomethionines) was calculated to be 15451.42 Da. According to the LC/MS analysis, the selenomet-IdmH-Δ96-104 had an observed mass of 15451.41 Da (Figure 3-13), which is in excellent agreement with the calculated mass of 15451.42 Da confirming the successful labelling.

Crystals of selenium-labelled IdmH-Δ96-104 were obtained using the same crystallographic conditions as for the native IdmH-Δ96-104. X-ray diffraction data for selenium-labelled IdmH-Δ96-104 were collected on beamline MASSIF-1 at the European Synchrotron Radiation Facility (Bowler et al., 2015; Svensson et al., 2017; Svensson et al., 2015; Nurizzo et al., 2016; Bowler et al., 2016).

3.7.5 Structural determination of IdmH-Δ96-104 and native IdmH

3.7.5.1 Electron density maps and quality of the models

The crystal structure of IdmH-Δ96-104 was determined to 2 Å resolution by SAD phasing using the SHELX pipeline (Sheldrick, 2010). The asymmetric unit was shown to comprise one copy of the IdmH dimer with the entire polypeptide chain, except for the N- and C-terminal regions (residues 1-6 and 135-139), well defined in electron-density (Figure 3-14).

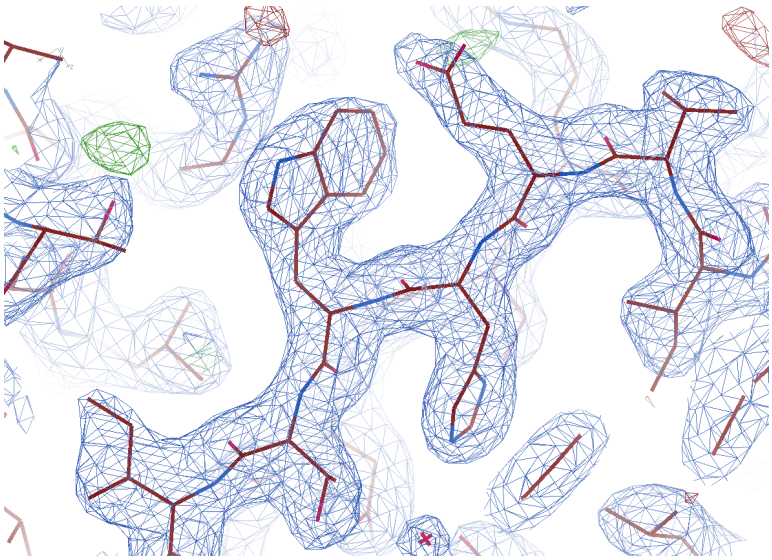


Figure 3-14 Representative $2F_o - F_c$ electron density map (in blue) from IdmH-Δ96-104, contoured at the 1 rmsd level.

The resulting IdmH-Δ96-104 models were refined to a R_{work} of 0.21 and a R_{free} of 0.28 for the selenium-labelled IdmH and a R_{work} of 0.23 and a R_{free} of 0.27 for the native IdmH-Δ96-104 (all data collection and refinement statistics can be found

in Table 2-7). With the structure of IdmH-Δ96-104 in hand, the crystal structure of wild-type IdmH (Figure 3-15) was determined by molecular replacement using a single subunit of IdmH-Δ96-104 as a search model. The asymmetric unit was revealed to contain a total of ten subunits of IdmH, for all of which the electron-density maps were well-defined except for the N-terminal residues 1-5 and C-terminal residues 144-148 (Figure 3-16).

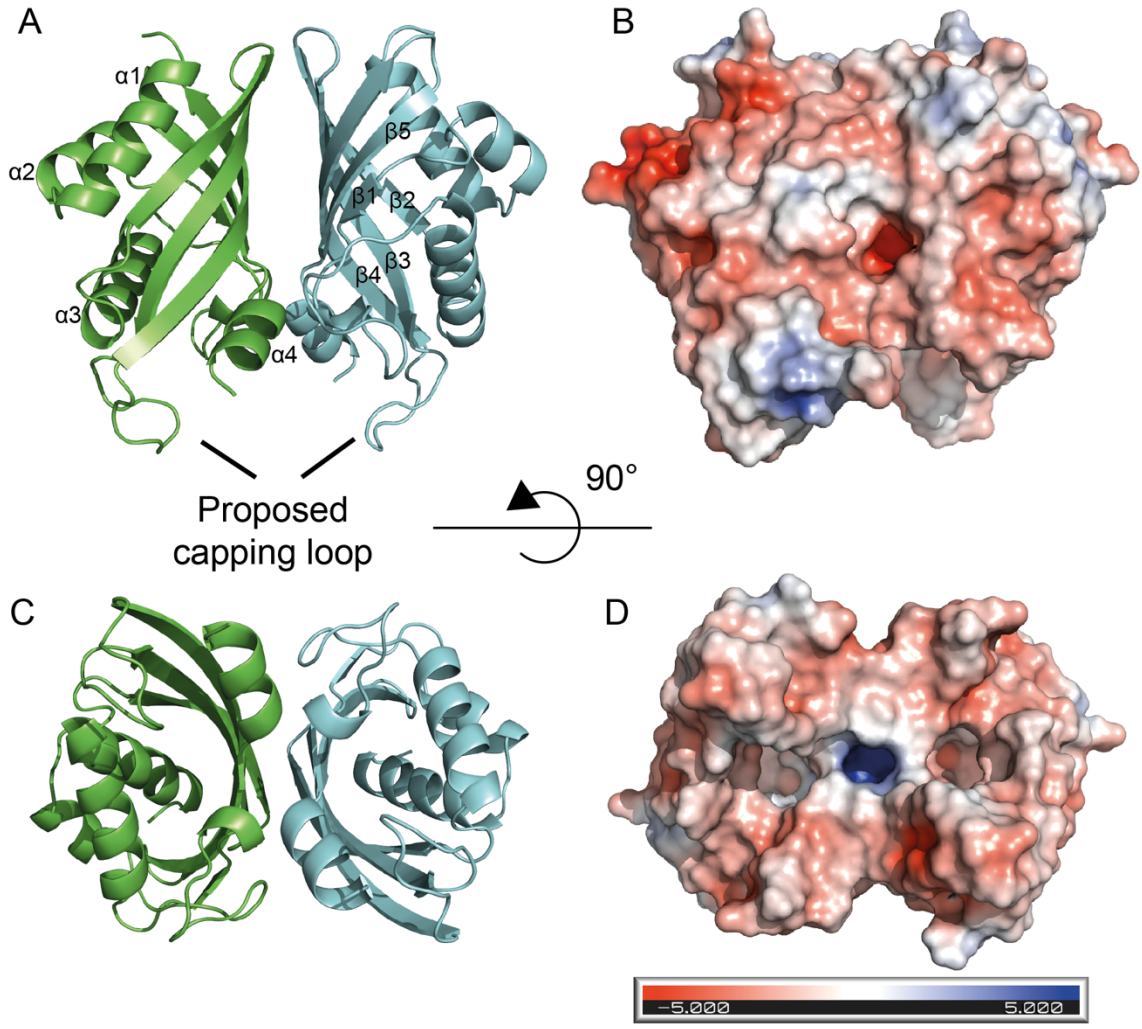


Figure 3-15 Structure of IdmH. *(a) and (c) Ribbon diagram showing IdmH dimer architecture. Individual monomers are coloured green and blue. (b) and (d) Surface representation of IdmH showing the polarity of IdmH surface and the hydrophobic cavity. Residues were coloured according to their polarity using the APBS Electrostatics plugin as part of the PyMOL suite.*

The final wild-type IdmH model was refined at 2.7 Å resolution with R_{work} = 0.21 and R_{free} = 0.25 (Table 2-7). Both native IdmH-Δ96-104 and wild-type IdmH

models were checked with MolProbity (V.B. Chen et al., 2010) revealing high-quality models with no Ramachandran outliers. MolProbity scores for wild-type IdmH and native IdmH-Δ96-104 were 1.00 and 0.91, respectively.

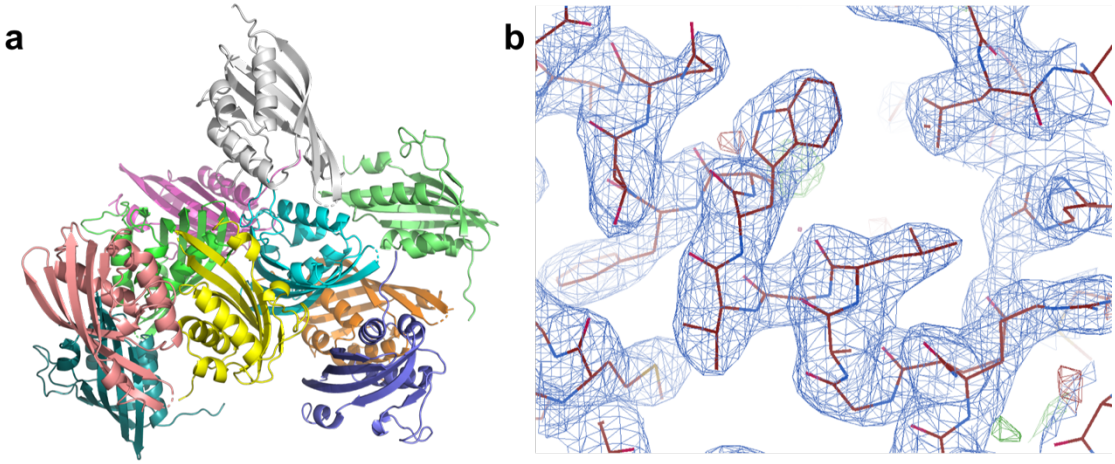


Figure 3-16 Solving the structure of wild-type IdmH. *(a) The asymmetric unit of IdmH is comprised of ten copies of monomeric IdmH. The large asymmetric unit combined with low sequence homology explains why initial phasing attempts using homologous replacement were fruitless. (b) Representative $2F_o - F_c$ electron density map (in blue) from wild-type IdmH, contoured at the 1 rmsd level.*

3.7.5.2 Fold of the subunit

The structure of IdmH reveals a homodimer with individual protomers comprised of four antiparallel β-strands and four α-helices, and belonging to the α+β class of structures (Hubbard et al., 1999). The core of the subunit is composed of four β-strands (β2–β5) that are continuous in sequence. An additional β-strand (β1), considerably shorter in length, runs parallel to strand β5, giving rise to a five-stranded mixed β-sheet (Figure 3-15). The β-sheet of each subunit is curved and forms the central part of a distorted α + β barrel. The barrel is completed by three α-helices (α1-3). The dimer assembly is driven by hydrophobic interactions between the β-sheets from the two subunits generating the dimer shown in Figure 3-15. Two loops are located near the entrance of the hydrophobic pocket and limit the access to the proposed active site from the outer solvent.

Interestingly, one of the two loops, comprising residues 96-105 and partially truncated in the IdmH-Δ96-104 variant, appears to occupy two distinct conformations within each of the five dimers in the asymmetric unit (Figure 3-17). Phe101 residues from the two chains exhibit a particularly interesting

conformational variability with the two rotamers being positioned perpendicular to each other. Other residues within the loop, such as Met102, were also observed to occupy two distinct rotamers. The observed conformational variability suggests that this loop is likely to be highly dynamic in solution. Studies of other putative Diels-Alderases have shown that the loop covering the active site becomes more ordered upon substrate binding and this loop could also play such a role here (Byrne et al., 2016; Q. Zheng et al., 2016).

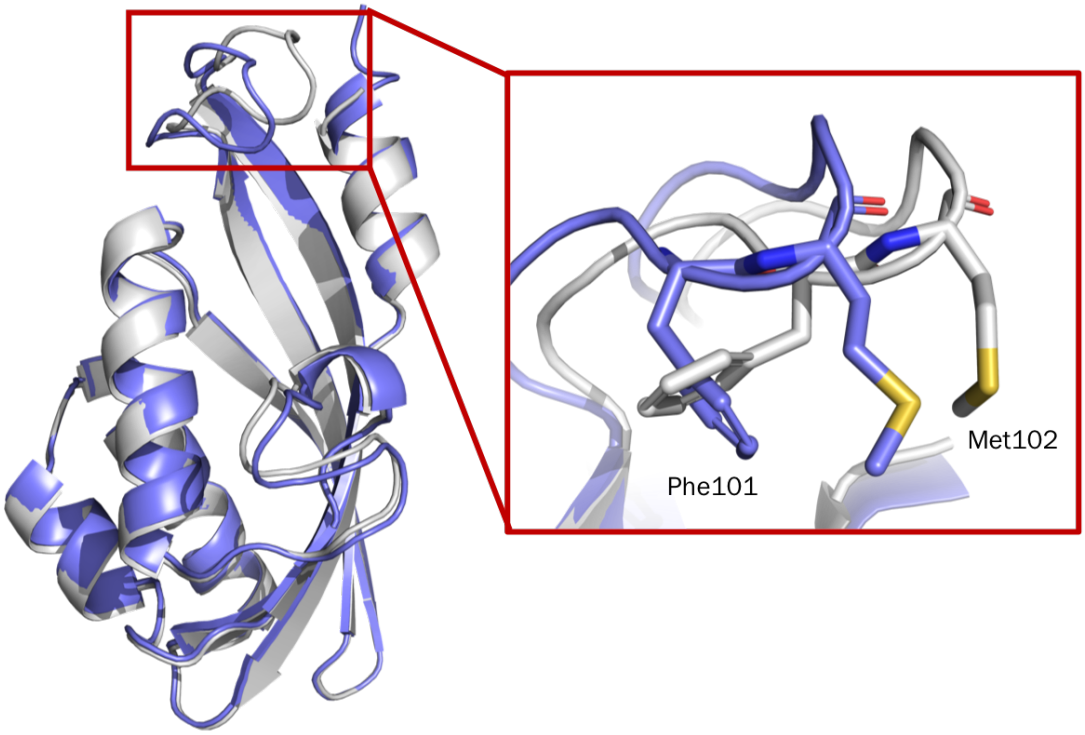
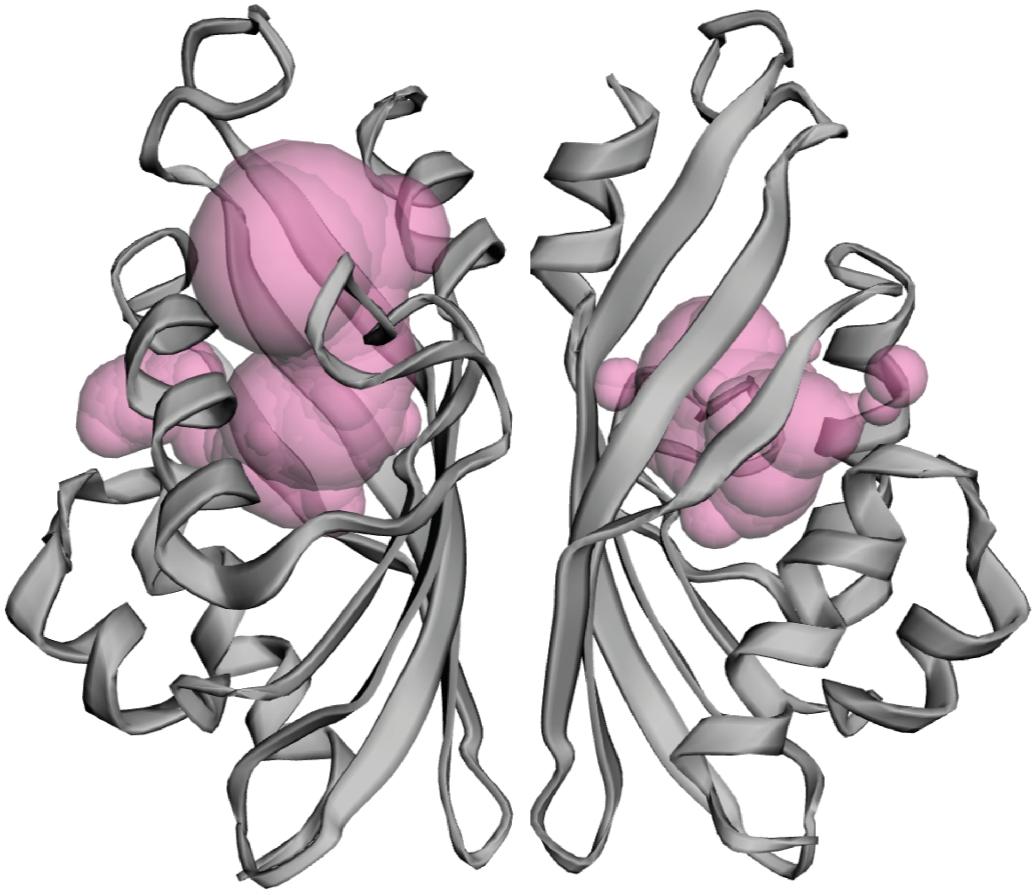


Figure 3-17 Conformational differences in the loop comprising residues 96-105. *This loop was truncated in IdmH-Δ96-104 to aid crystallisation while in the wild-type IdmH structure it was observed in two distinct conformations between chains A (blue) and B (grey). The inset shows a magnified view of the loop and two amino acids Phe101 and Met102 are shown as sticks.*

3.7.5.3 Proposed active site

At the centre of each protomer is a deep hydrophobic pocket penetrating towards the core of the barrel. Solvent-accessible volume (using a probe sphere of 1.4 Å) for each pocket was calculated using the CASTp server (Tian et al., 2018) (Figure 3-18). Surprisingly, the surface-accessible volume of chain A was calculated to be significantly bigger (409.554 $Å^3$) than that of chain B (234.257 $Å^3$).

As can be seen from Figure 3-18, the opening of the chain B pocket was not included in the solvent-accessible volume calculations. This is due to CASTp excluding shallow depressions from the calculations. To distinguish a pocket from a shallow depression, CASTp compares the cross-section of the pocket mouth opening to that of the actual pocket, and if the former is wider than the later, the area is regarded as a shallow depression. Therefore, it is possible, that a conformational change near the pocket opening (such as that shown in Figure 3-17) led to its increased cross-section and, as a result, the exclusion of this part of the pocket from the solvent-accessible volume calculations as it is deemed a shallow depression.



Wild-type IdmH Chain A
Solvent-accessible surface area: 464 Å^2
Solvent-accessible volume: 410 Å^3

Wild-type IdmH Chain B
Solvent-accessible surface area: 301 Å^2
Solvent-accessible volume: 234 Å^3

Figure 3-18 The solvent-accessible surface area and volume (displayed as pink spheres) of hydrophobic pockets from IdmH chains A and B, as computed by CASTp 3.0 (Tian et al., 2018). *1.4 Å diameter probe sphere was used.*

The reasons for these subtle differences in the pocket entrance area are hard to pin down; however, it is possible that the two monomers exist in slightly different conformations and only one chain can bind the substrate at one given time.

Each hydrophobic pocket is lined by the side chains of residues Tyr16, Phe19, Leu39, Trp59, Val62, Trp63, Met115, His129, Ser133, Asn135 (Figure 3-19). We propose that this constitutes the active site of the enzyme.

It is not surprising that this hydrophobic pocket has so many aromatic residues. It is thought that many natural product maturation enzymes act as natural product templates. Therefore, one could envisage that the aromatic residues in IdmH proposed active site could form stacking interactions and align the indanomycin precursor molecule **4** (Figure 1-23) in such an orientation which results in catalysis of the pericyclic, indane ring-forming reaction (Carey and Sundberg, 2013).

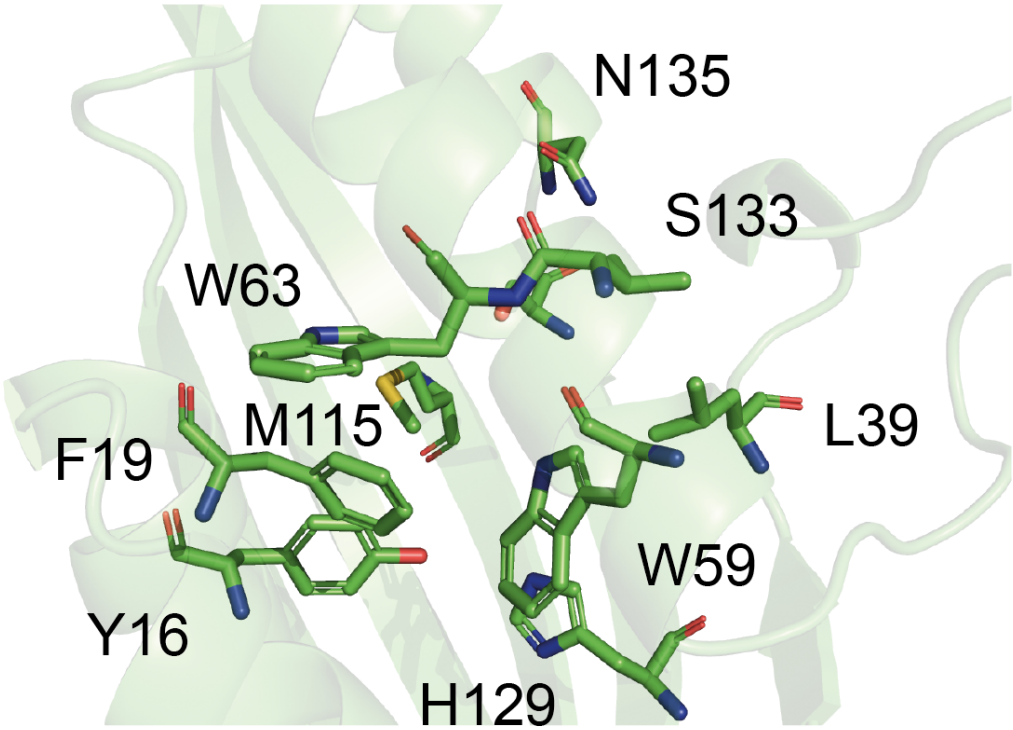


Figure 3-19 Closer look into a proposed active site of IdmH. *Key residues (Tyr16, Phe19, Leu39, Trp59, Val62, Trp63, Met115, His129, Ser133, Asn135) are shown in sticks.*

Furthermore, there are potential hydrogen-bond donors (such as Ser133 and Thr131) that could interact with the indanomycin precursor carbonyl group (C21=O) (Figure 3-1). Such an interaction could help with the proposed Diels-

Alder reaction, making the carbonyl have a stronger electron-withdrawing effect on the dienophile. It is also interesting to observe that at the deep end of the hydrophobic pocket, there are Trp59 (and Phe19/Tyr16), which are similar to Trp124 (and Phe41) in AbyU Diels-Alderase reported previously (Byrne et al., 2016).

3.7.5.4 Sequence homologues of IdmH

Next, IdmH was compared to its closest sequence homologues by performing a BLASTp search on the PDB database. The results of this search are shown in Figure 3-20. IdmH's closest sequence homologue (30% sequence identity) is a putative polyketide cyclase *Chromobacterium violaceum*, (PDB code 4LGQ). The next two closest homologues are AcIR and SnoaL2 putative hydrolases from *Streptomyces galilaeus* and *Streptomyces nogalater*, respectively (PDB codes 2GEY and 2GEX, respectively) (Beinker et al., 2006). Similarly to IdmH, all three structures also form homo-dimers. Only two of the nine reported homologue structures (polyketide cyclase SnoaL,1sjw, and epoxide hydrolase, MonBI, 3wmd) have reported functional data, while the other seven have been deposited to the PDB as stand-alone structural models and therefore only have putative activities suggested (Sultana et al., 2004; Minami et al., 2013).

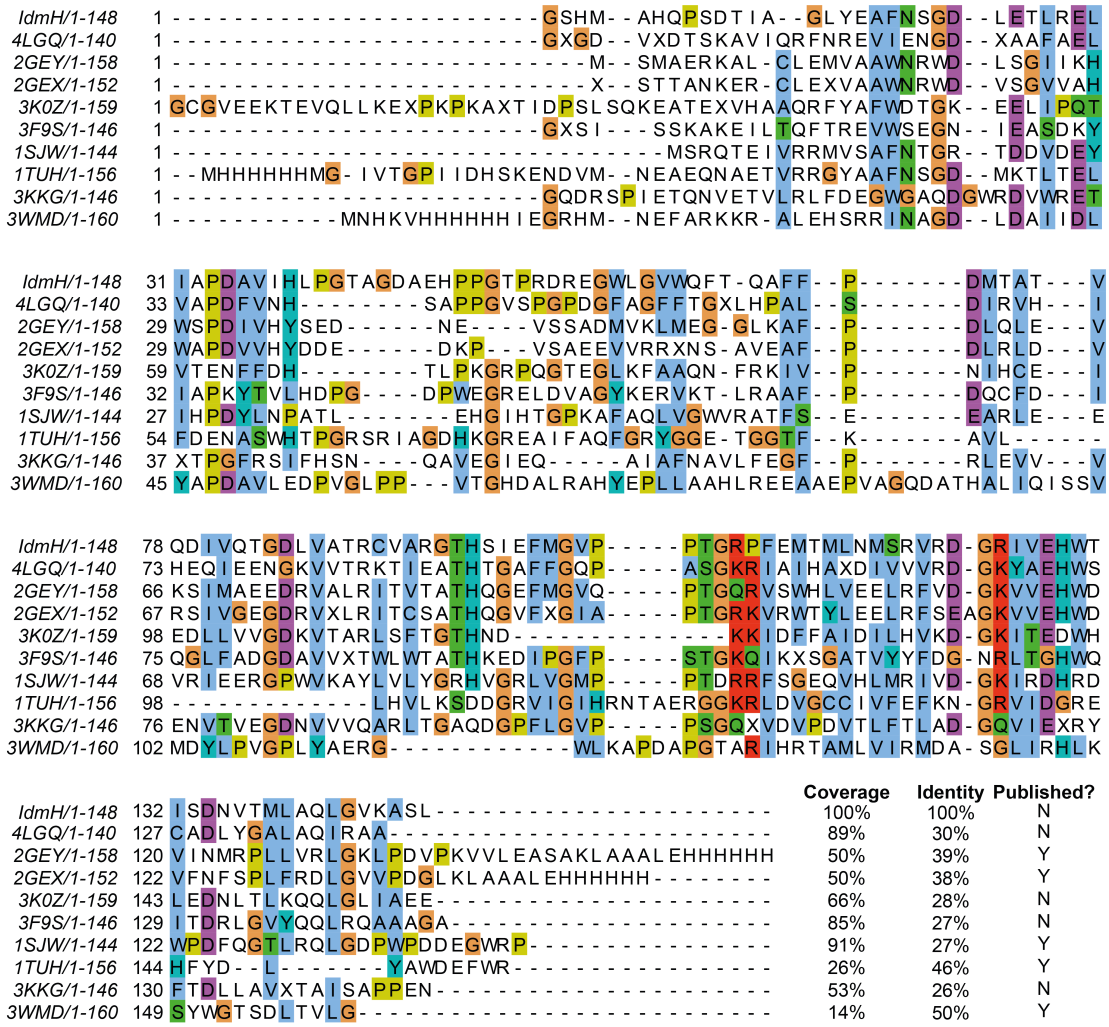


Figure 3-20 Sequence alignment of IdmH to its homologs. *The sequence of IdmH was searched against PDB database. Conserved residues are highlighted in colour using Jalview software and Clustal scheme. The associated query coverage and sequence identities with IdmH are also included. Five out of nine structural homologues have been published in peer-review journals.*

Following BLASTp search, the structure of IdmH was superimposed onto its sequence homologues using PyMOL software, and the resulting comparisons are shown in Figure 3-21. The α + β barrels of IdmH and each of its homologues were aligned relatively well; however, there are many differences between the superimposed models in the additional loops and helices. The polyketide cyclase SnoaL from nogalamycin biosynthetic pathway exhibited the highest structural similarity to IdmH as measured by the Root Mean Square of the fit (RMS=1.564), while epoxide hydrolase MonBI was shown to be the least structurally homologous to IdmH (RMS=19.865).

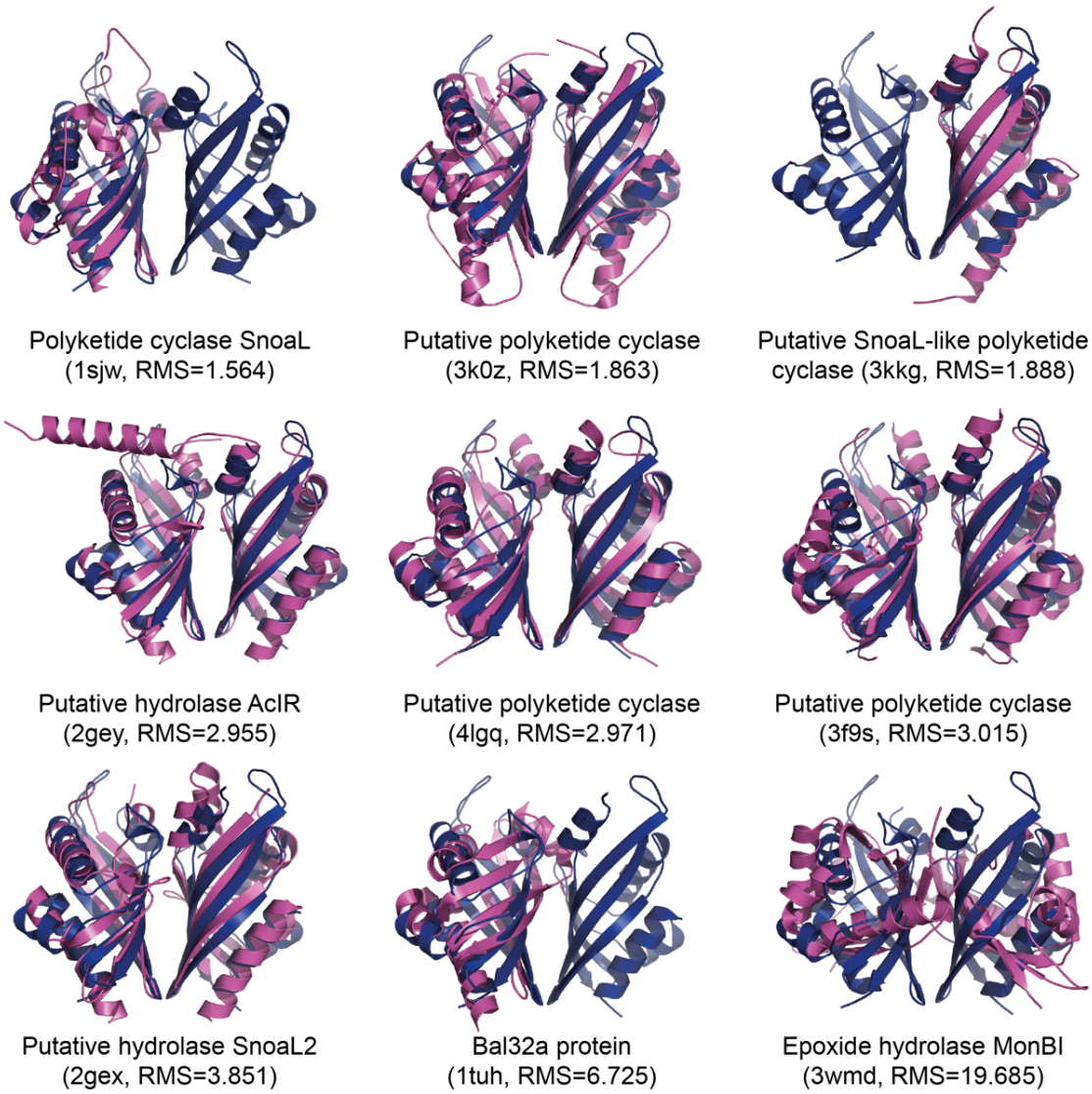


Figure 3-21 Comparison of wild-type IdmH (blue) to its closest sequence homologues (purple). *IdmH model was aligned with each PDB entry using PyMOL software. RMS stands for Root Mean Square and is reported as an estimation of the imperfection of the fit.*

3.7.5.5 Comparison to SnoaL

Both crystallographic analysis and gel filtration chromatography experiments suggested that SnoaL forms tetramers in solution (Sultana et al., 2004), while the size-exclusion chromatography analysis of IdmH showed that it forms homo-dimers in solution.

SnoaL catalyses an intramolecular aldol condensation by utilising the invariant residue Asp121 as an acid/base catalyst during the reaction. Other essential residues for SnoaL's activity include Asn33, Gln105 and two histidines, His107 and His119. In contrast to the SnoaL, IdmH active site does not contain charged

residues at the same place as SnoaL's catalytic Asp121 and His107, and has more hydrophobic residues (Met115, Ile37) instead (Figure 3-22).

Both active sites include many aromatic residues, and some of them are conserved between the two proteins (IdmH/Phe19 with SnoaL/Phe15 and IdmH/His129 with SnoaL/His119). Sequence-homologous active site tryptophan residues (IdmH/Trp63 and SnoaL/Trp54), however, appear to be in entirely different locations and an additional tryptophan (Trp59) and tyrosine (Tyr16) residues in IdmH seem to be unique to IdmH (SnoaL contains Tyr31 and Met11 at these locations). The absence of charged residues in the IdmH active site as well as the presence of an additional aromatic amino acid suggest IdmH is not performing the same reaction as SnoaL.

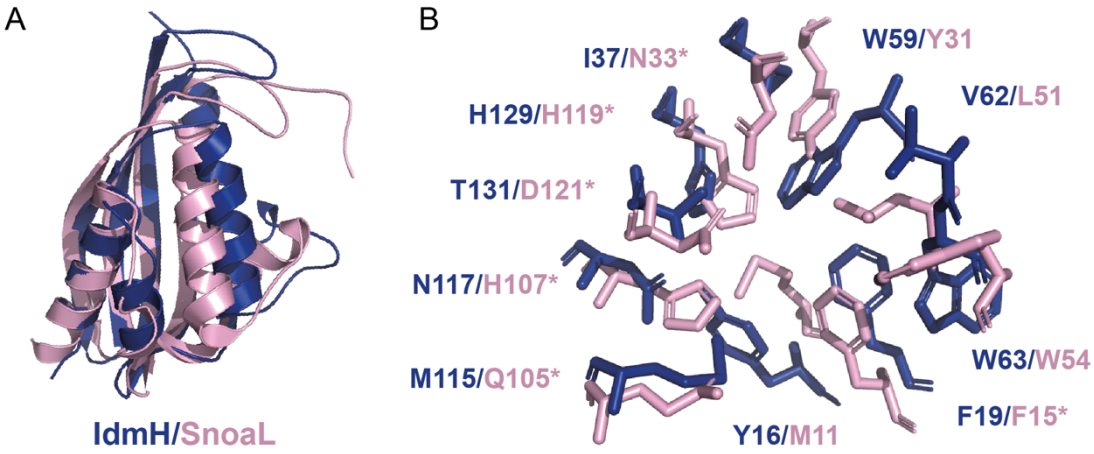


Figure 3-22 Comparison of IdmH (blue) to its closest sequence homologue SnoaL (pink). *(a) IdmH chain A was aligned with SnoaL (PDB code 1sjw) chain A using PyMOL software. (b) SnoaL active site residues (pink) compared to their counterparts in IdmH (blue). While some residues appear to be conserved between two proteins (H129/H119, W63/W54, F19/F19), charged catalytic residues (H107, Asp121) are unique to SnoaL. Another difference between the two active sites arises from different aromatic residue choice and positioning (W59/Y31 and W63/W54), and Tyr16 appears to be completely unique to IdmH. * denotes SnoaL catalytic residues.*

3.7.5.6 Structural homologues of IdmH

Sequence-based search alone is not capable of identifying structural homologues with very low sequence conservation. To do this, the structure of wild-type IdmH was searched against the PDB using the DALI server (Holm and Laakso, 2016). Among the top ten structures most closely related to IdmH are polyketide cyclase

SnoaL (Sultana et al., 2004), several putative polyketide cyclases (PDB codes 4LGQ, 3F9S and 3KKG; not published) and putative polyketide hydrolase AcIR (Beinker et al., 2006), all of which were identified by a protein sequence search described in section 3.7.5.4.

Other proteins with high level of structural conservation (but less than 20% sequence conservation) to IdmH include polyketide cyclase AknH (Kallio et al., 2006), several thermophilic limonene-1,2-epoxide hydrolases (Ferrandi et al., 2015), LEH-like epoxide hydrolase (Johansson et al., 2005), many ketosteroid isomerases (Ha et al., 2001; Schwans et al., 2011; Yun et al., 2003; Fafarman et al., 2012), engineered Digoxigenin binder (Tinberg et al., 2013), meroterpenoid cyclase Trt14 (Mori et al., 2017), nuclear transport factor 2 (Chaillan-Huntington et al., 2001), the association domain of Ca2+/calmodulin-dependent kinase II (Bhattacharyya et al., 2016) and numerous hypothetical or uncharacterised proteins and other putative polyketide cyclases.

Due to the low sequence conservation to IdmH as well as the wealth of catalytic activities displayed by the structural homologues it is difficult to gain any catalytic insights from making sequence- and structural-based comparisons alone.

3.7.6 Co-crystallisation and crystal soaking of IdmH-Δ96-104 and indanomycin

Since sequence and structural alignment did not provide information about the reaction catalysed by IdmH we sought to more precisely map the enzyme active site. This requires precise structural information on the location and binding of the indanomycin precursor **4** (Figure 1-23) in the active site. Being unable to easily obtain **4**, we therefore attempted to determine a three-dimensional model of IdmH in a complex with its putative product indanomycin **1**. This information would give further evidence that IdmH is involved in indanomycin cyclisation and also help depict its possible mechanism of action.

Co-crystallisation and soaking are two conventional approaches used to determine a protein structure with the ligand of interest. The co-crystallisation approach was tried first, whereby purified IdmH was mixed with indanomycin solubilised in dimethyl sulfoxide (DMSO) and the crystal trays were set up as before. 0.1 to 0.5 mM indanomycin concentrations were used which resulted in the final concentration of DMSO in the drop being either 0.5% or 2.5%,

respectively. Several problems were identified using this approach: first, indanomycin is very insoluble when kept at low DMSO concentrations for a prolonged period and itself formed crystals; second, IdmH does not crystallise in the presence of DMSO and hence no protein crystal growth was observed.

Since co-crystallisation was fruitless, crystal soaking was attempted next. This method involves the transfer of crystals into the soaking solution containing the small compound of interest for a period of time before flash-cooling and data collection. IdmH-Δ96-104 crystals were grown as described previously (section 3.7.3) and soaked in mother liquor solution containing 0.01-10 mM indanomycin for different periods of time (15 s, 60 s, 5 min, 30 min, 2 hr, overnight). High ligand concentrations were selected to attain high ligand-protein occupancy.

Crystals soaked in 10 mM indanomycin solution for a couple of seconds and 5 mM solution for 60-120 seconds lost the ability to diffract altogether, which is indicative of severe damage to the crystal lattice and could be the result of high solvent concentration. One crystal soaked in 5 mM indanomycin for 20 seconds diffracted to 2.8 Å but was very prone to radiation damage which prevented the collection of a full data set.

No damage to the crystal lattice was observed for crystals soaked in 0.05-1.0 mM indanomycin (corresponding to 0.5% to 10% DMSO concentration, respectively). As a result, over thirty datasets were collected and processed. Additional density in the hydrophobic cavity of IdmH was observed in a number of cases; however, the ligand occupancy was not sufficient enough to fit and refine indanomycin with high confidence (Figure 3-23).

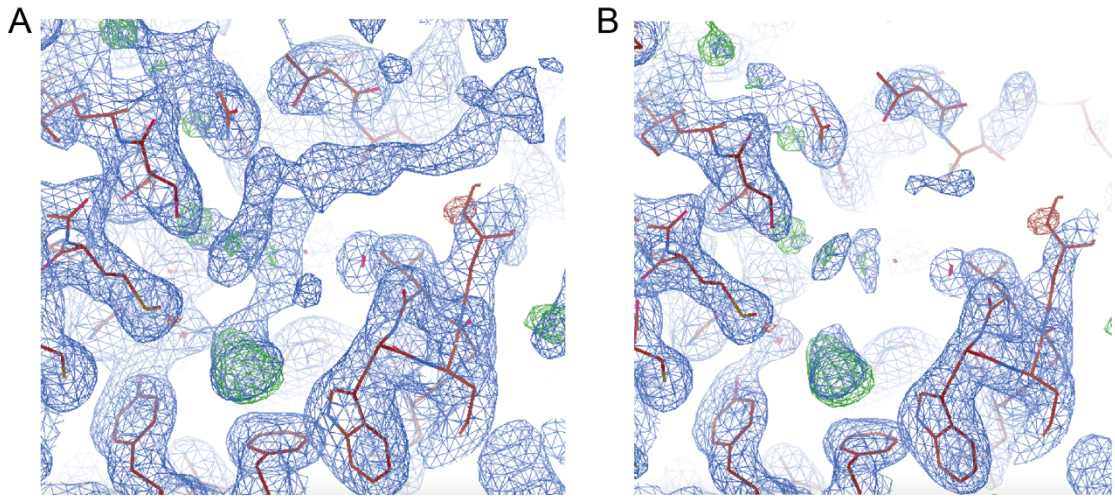


Figure 3-23 Electron-density map, derived from crystals soaked in 1 mM indanomycin solution overnight and resolved to 2.2 Å resolution. *Omit $F_o - F_c$ electron density (in green) is contoured at the 3 rmsd level. (a) $2F_o - F_c$ electron-density map (in blue), contoured at the 0.5 rmsd level, reveals continuous additional density in the cavity. (b) $2F_o - F_c$ electron-density map (in blue), contoured at the 1.0 rmsd level, shows discontinuous additional density in the cavity.*

3.8 Investigating IdmH enzymatic activity by nuclear magnetic resonance spectroscopy

Due to the complex stereochemistry of indanomycin precursor **4**, we were unable to obtain a synthetically-synthesised substrate to study the enzymatic activity of IdmH using classical enzymology methods. Since every enzyme possesses binding affinity to both its substrate and the product, we decided to investigate the binding of the product indanomycin **1** to IdmH by 2D NMR spectroscopy. This technique has the ability to show whether indanomycin binds to IdmH, and can be used to determine which residues are involved in this interaction by looking at the chemical shift perturbations and mapping them to the assigned backbone resonances.

The majority of the work described in this section was performed by the candidate while the NMR facility manager Dr Arnout Kalverda assisted the candidate with finalising the backbone resonance assignments and the calculation of chemical shift perturbation values.

3.8.1 Expression and purification of ^{15}N labelled IdmH

To study a protein by NMR spectroscopy, it needs to be labelled isotopically, at a minimum this must be with ^{15}N and for a full analysis with ^{13}C and ^{2}H as well. To achieve the single labelling, IdmH was expressed in M9 minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source (methods section 2.3.1.3). Protein purification was performed identically to the protocol used previously (methods section 2.3.2.2).

3.8.2 Collection of transverse relaxation-optimised spectroscopy spectra

Acquiring a 2D spectrum allows evaluation of whether the target protein is suitable for structural studies by NMR spectroscopy. The ^{1}H-^{15}N transverse relaxation-optimised spectroscopy (TROSY) spectrum of IdmH was collected using a sample with a concentration of 0.9 mM at 25 °C at a field strength of 600 MHz (Figure 3-24). In this spectrum each peak corresponds to a proton directly bonded to nitrogen, therefore in this experiment, each amide in the protein should have a representative peak.

The dispersion of chemical shifts in the spectrum indicated that the protein was folded and stable at the chosen buffer conditions. However, some areas of the spectrum were less resolved and showed peak broadening and merging into one another. In NMR spectroscopy, line widths generally increase with molecular mass, and since the IdmH dimer has a mass of approximately 36 kDa, the observed line broadening is not surprising. The number and intensity distribution of the peaks indicated that there is little protein dynamics in the protein backbone that could hinder structural analysis by NMR spectroscopy, but that there may be some longer loops which are more flexible than the IdmH core.

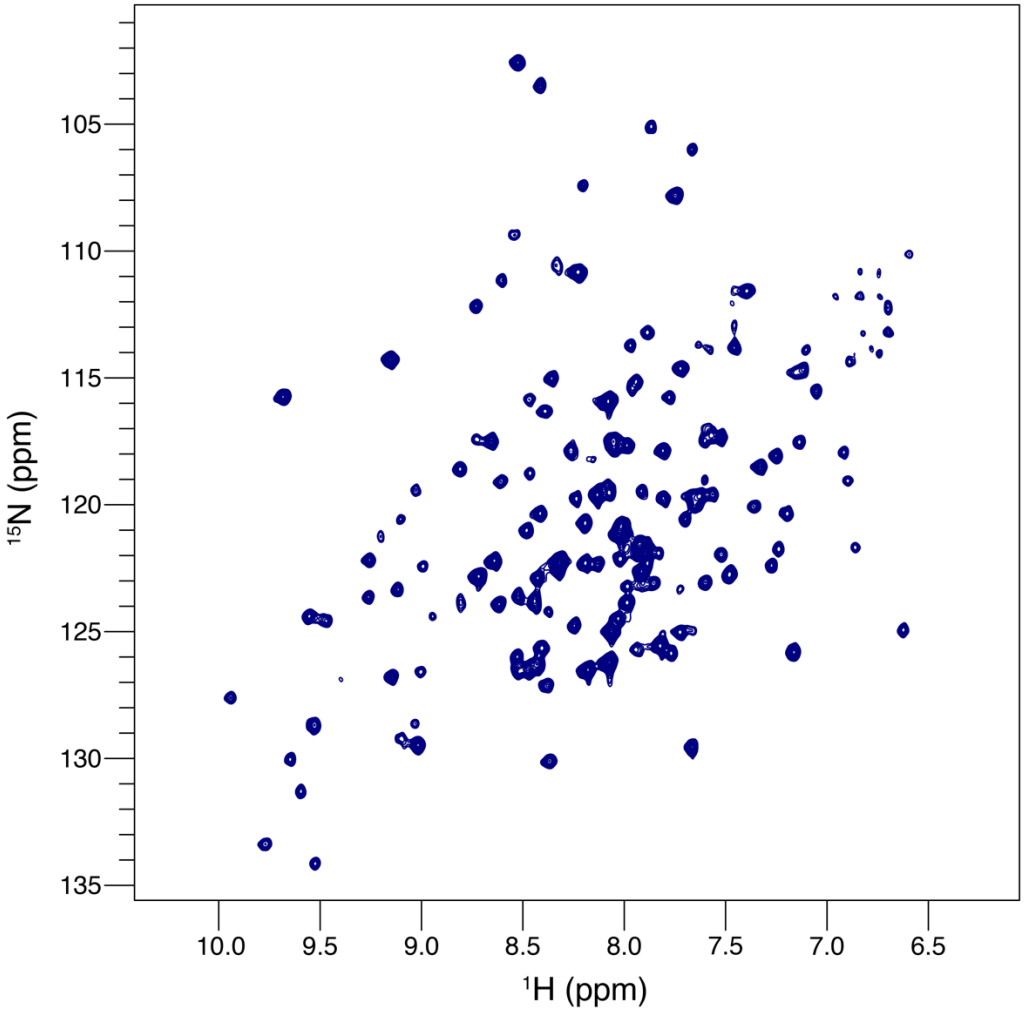


Figure 3-24 ^{1}H-^{15}N TROSY spectrum of IdmH. *The spectrum was recorded in 20 mM Tris pH 7.3, 50 mM NaCl, 1 mM DTT, 0.02% (w/v) NaN_3 at a concentration of 0.9 mM at 25 °C at 600 MHz. In this spectrum each peak corresponds to each amide in the protein.*

3.8.3 IdmH backbone nuclei resonance assignment

To study the effects of indanomycin binding on the structure of IdmH, the backbone resonance peaks need to be assigned. This is done by acquiring three-dimensional NMR spectra for which a protein sample needs to be labelled with three isotopes (^{2}H, ^{13}C and ^{15}N). To achieve this, IdmH was expressed in the M9 minimal medium in D_2O containing $^{15}NH_4Cl$ as the sole nitrogen source and $^{13}C_6H_{12}O_6$ as the single carbon source (methods section 2.3.1.3). The resulting recombinant IdmH was purified as described previously (methods section 2.3.2.2).

Triple-resonance spectra of IdmH (HNCA, HNcoCA, HNCO, HNcaCO, HNCACB and HNcoCACB) were collected at a protein concentration of 0.8 mM at 25 °C at a field strength of 750 MHz (except for HNcaCO spectrum which was acquired at 950 MHz). Non-uniform sampling (NUS) and targeted acquisition techniques were used to reduce the acquisition time (Rovnyak et al., 2004; Jaravine and Orekhov, 2006). Reconstruction of NUS data was performed using multi-dimensional decomposition and compressed censing tools in MddNMR software (Orekhov and Jaravine, 2011). This was followed by regular Fourier Transform (FT) processing of the complete data, which was also performed in MddNMR.

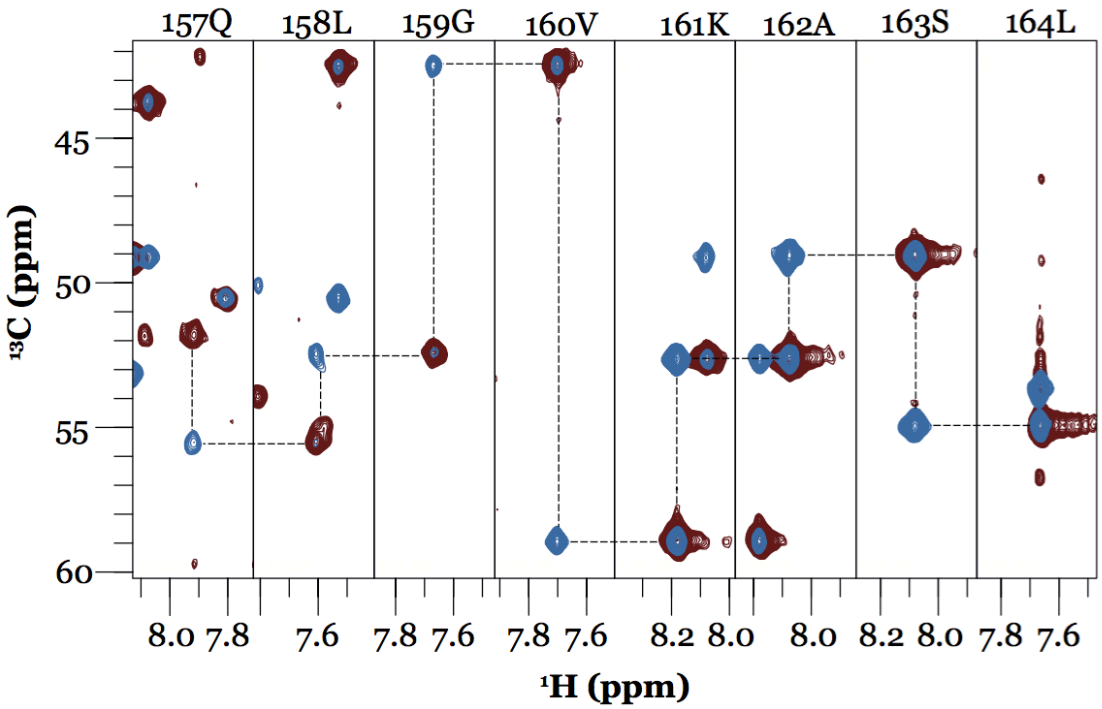


Figure 3-25 Sequential assignment of the last eight residues of the backbone of IdmH using strip plots from HNCA (blue) and HNcoCA (red) overlaid spectra.

Processed spectra were then imported into CCPNmr Analysis (Skinner et al., 2016) for peak picking and assignment of anonymous spin systems (Vranken et al., 2005). Peak lists were submitted to AutoAssign, a constraint-based expert system for automating the analysis of backbone resonance assignments (Zimmerman et al., 1997). A representative sequential backbone resonance strip plot is shown in Figure 3-25. This method resulted in the assignment of approximately 88% of IdmH backbone resonances (Figure 3-26).

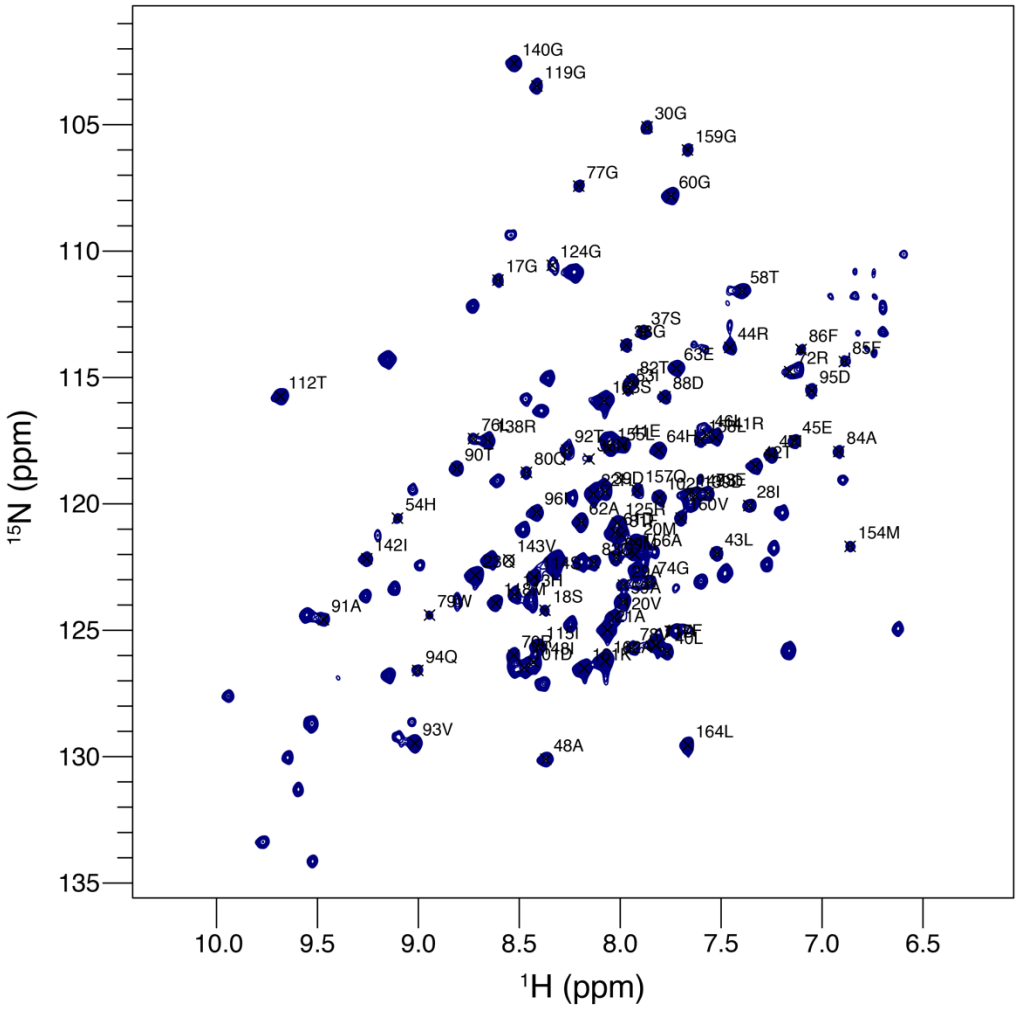


Figure 3-26 Assigned ^{1}H-^{15}N HSQC-TROSY spectrum of IdmH. *~88% of the backbone residues were assigned using AutoAssign (Zimmerman et al., 1997) followed by manual assignment in CCPNmr Analysis (Skinner et al., 2016).*

3.8.4 Binding of indanomycin to IdmH

To investigate binding, ^{15}N IdmH was mixed with a stoichiometric amount of ligand (indanomycin), and a ^{1}H-^{15}N HSQC-TROSY spectrum was acquired (Figure 3-27). Significant chemical shift perturbations (CSPs) (with no duplicate peaks for bound and unbound states) were observed for the majority of the backbone residues, suggesting a binding event (Figure 3-27).

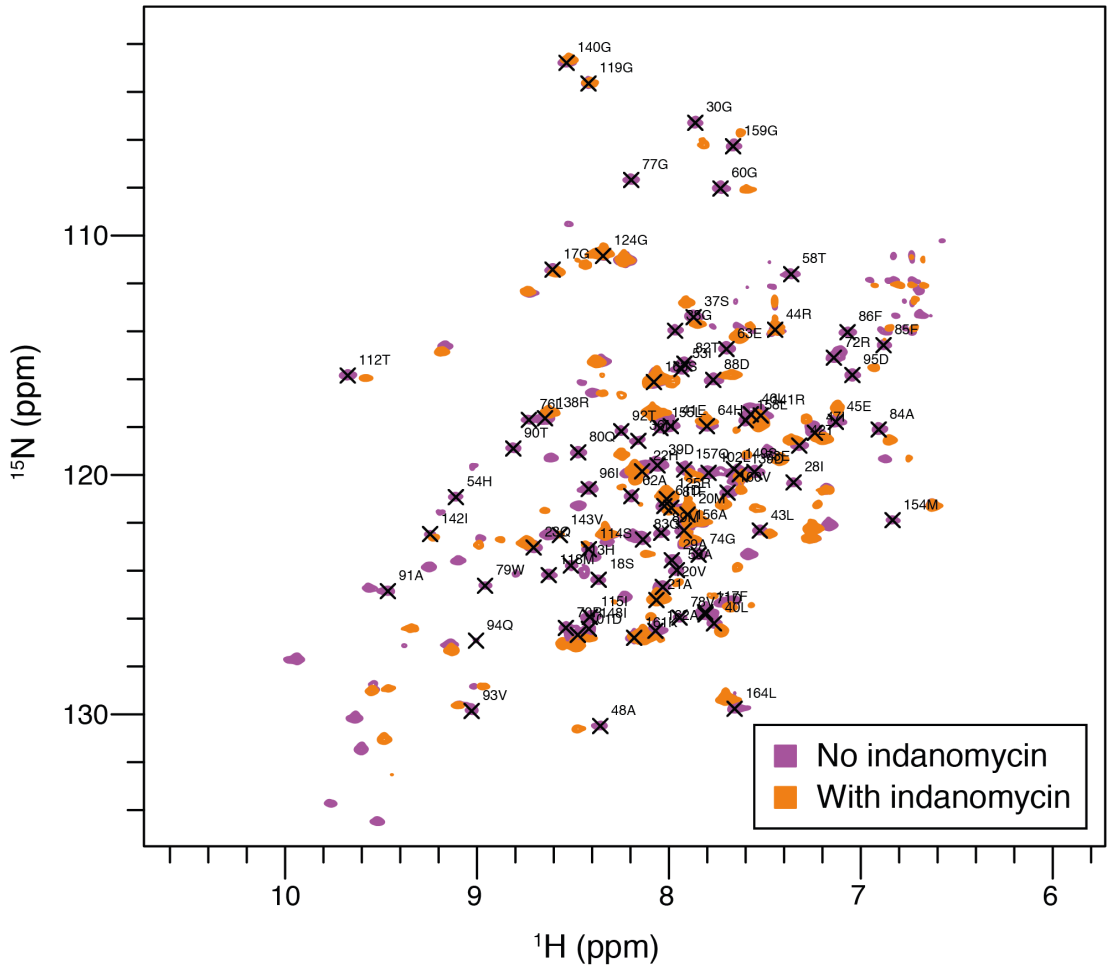


Figure 3-27 ^{1}H–^{15}N HSQC-TROSY spectra of isotopically-labelled IdmH alone (purple) and in conjunction with a stoichiometric amount of ligand indanomycin (orange). *Each peak in the spectrum represents a bonded N-H pair, with its two coordinates corresponding to the chemical shifts of each of the H and N atoms. Upon the addition of ligand indanomycin (orange spectrum), significant changes in the chemical shifts of some peaks were observed, and these peaks are likely to lie on the binding surface where the binding perturbed their chemical shifts.*

To investigate the binding affinity further, a titration experiment was performed with sub-stoichiometric protein to ligand ratios as outlined in Table 3-3. ^{1}H, ^{15}N or both shifts were observed for a vast majority of the assigned residues. Examples of specific residue chemical shift perturbations during the titration are shown in Figure 3-28.

Table 3-3 Protein and ligand concentrations used to investigate indanomycin binding to IdmH by NMR spectroscopy.

Protein to ligand ratio	IdmH conc.	Indanomycin conc.
1 : 0	200 µM	0 µM
4 : 1	200 µM	50 µM
2 : 1	200 µM	100 µM
4 : 3	200 µM	150 µM
1 : 1	200 µM	200 µM

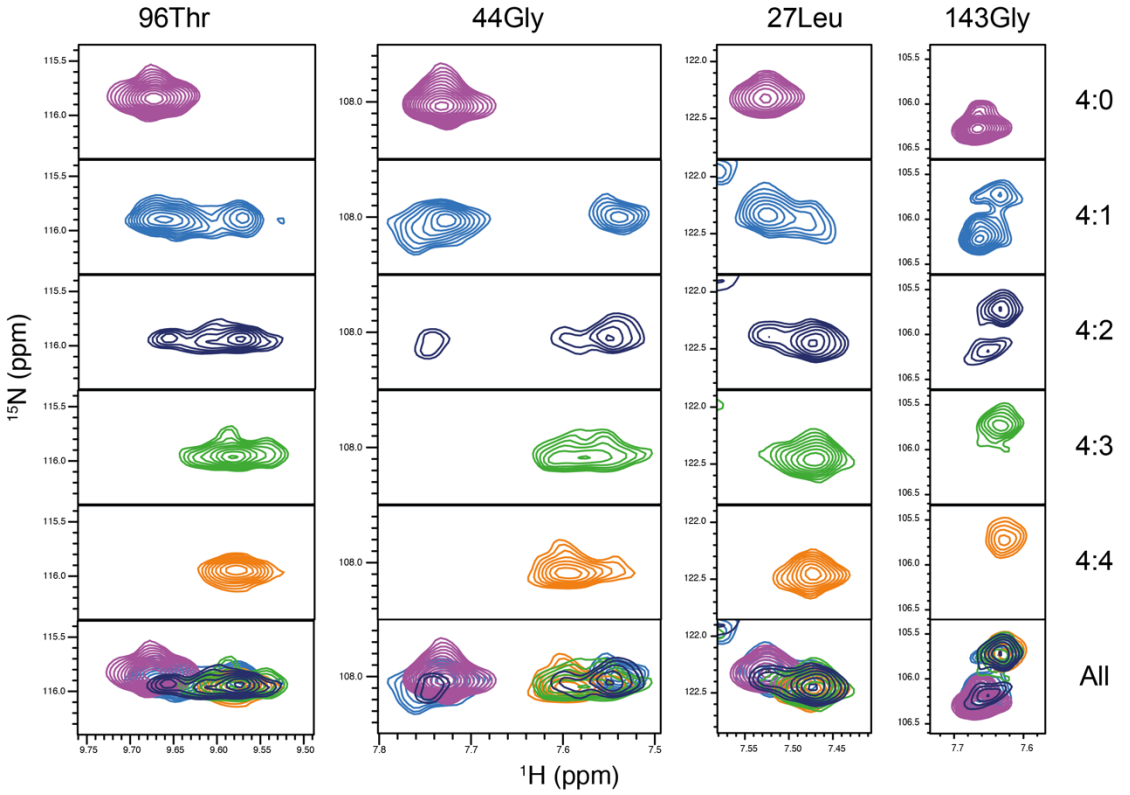


Figure 3-28 CSPs for individual residues going from free (purple) to bound (orange) state as recorded in ^{1}H–^{15}N HSQC-TROSY spectra. *Peak broadening is observed when in equilibrium between free and bound (4:2 and 4:3 ratios of protein: ligand, dark blue and green, respectively) states. Peaks sharpen up again close to saturation (orange). In all cases, the free peaks decrease in intensity as the bound peaks increase, suggesting a slow exchange.*

Since shifts can be influenced by through-bond, through-space and allosteric interactions, it is difficult to determine whether CSPs in IdmH are a result of direct binding or conformational change. However, CSPs can be used to localise the protein-ligand interactions. To do this, the geometrical distance moved by each peak was calculated for all assigned residues (Figure 3-29a) and chemical shifts

larger than the standard deviation of all shifts were mapped on IdmH diagram using yellow, orange and red colours (Figure 3-29b).

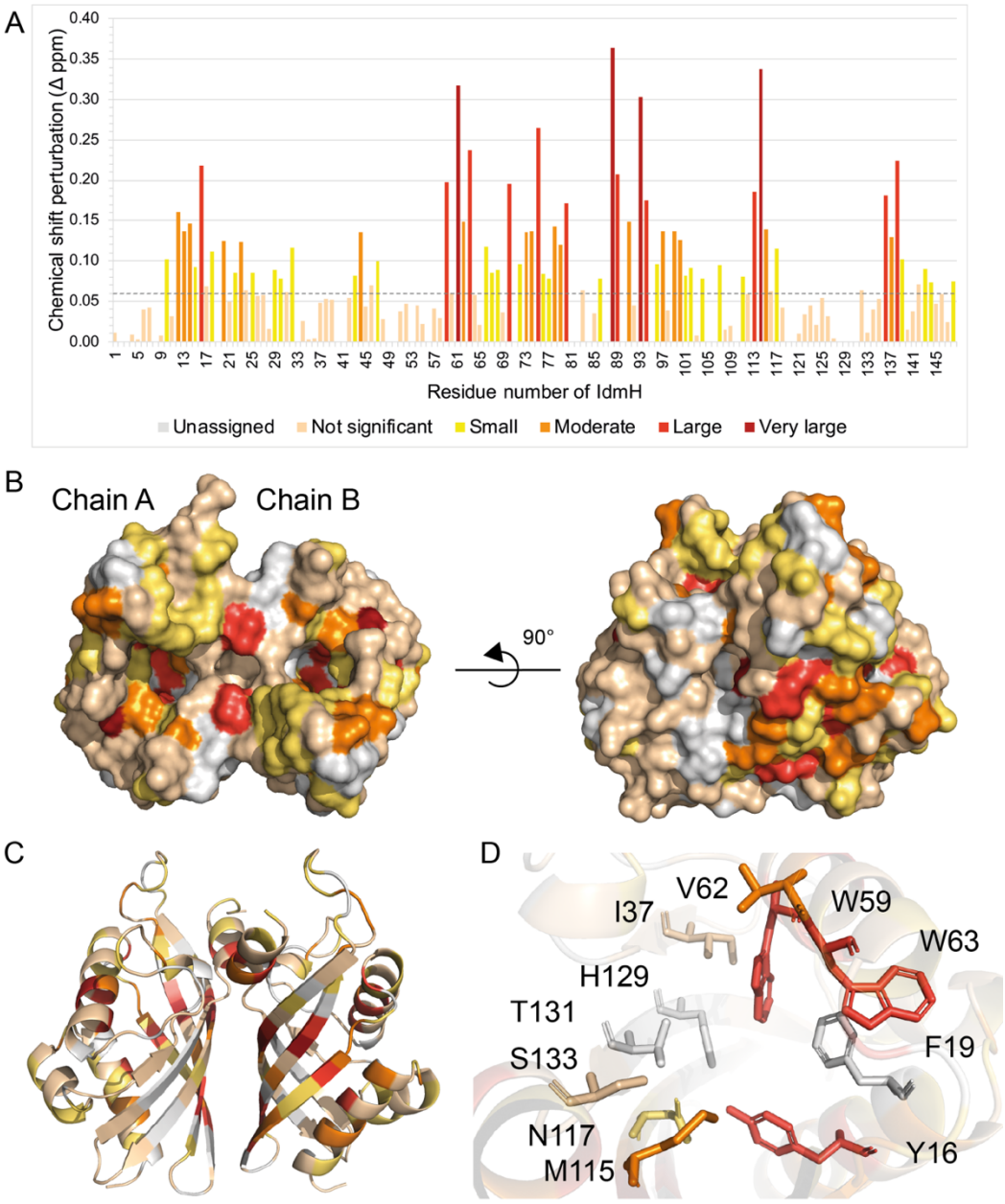


Figure 3-29 Mapping chemical shift perturbations of IdmH upon the addition of a ligand (indanomycin). *(a) A geometrical distance moved by each assigned peak was calculated using a minimum chemical shift procedure (Williamson, 2013) and coloured from yellow to dark red in increasing value. Chemical shifts which were smaller than the standard deviation of all shifts (0.072 ppm) were coloured in pale cream. Surface (b) and ribbon (c) representation of IdmH with peak shifts mapped using the same colour scheme as in (a). (d) Active site cavity of chain B with the proposed catalytic residues coloured according to the same colour scheme as in (a). CSP of only two resonances, S133 and I37, was estimated below the level of significance while all other resonances were either unassigned or exhibited a significant CSP.*

Some resonances in the bound spectrum moved a substantial distance from their free resonance counterparts making it impossible to match the two accurately. For this reason, a minimum chemical shift procedure was employed (Williamson, 2013; Farmer et al., 1996; Lüttgen et al., 2002). This method links each assigned free resonance to the signal in the bound spectrum that has moved the least from the position of the free. Therefore, each resonance gets assigned a CSP, and while the true change might be larger than the one assigned, it will never be smaller.

The results of chemical shift mapping suggest that the majority of the CSPs are localised at the top of the hydrophobic cavity of IdmH (α-helices 3-4) and on the β-sheet inside the pocket (β-strands 2-4) (Figure 3-15 and Figure 3-29c), which is consistent with the predicted location of the IdmH active site.

A closer look at the active site residues (Figure 3-29d) revealed that the majority of the resonances exhibited a moderate (Val62, Met115) to large (Trp59, Trp63, Tyr16) shift. Two resonances, Ser133 and Ile37, exhibited a shift below the level of significance (<0.72 ppm) and three remaining resonances, Thr131, Phe19 and His129, did not have their backbone resonances assigned and therefore their CSP could not be calculated.

3.9 *In-silico* studies

The work described in this section was performed by our collaborator Dr Marc Van der Kamp (University of Bristol).

3.9.1 Docking the product and simulating the protein-product complex

Autodock Vina software was used to dock indanomycin in the wild-type IdmH active sites (chain A and B). Four significantly different binding poses were found (two in chain A and two in chain B). For the binding poses in chain B, the conformer for Asn117 was adjusted to provide a slight difference in the hydrogen-bonding environment. This led to the following four options:

A - indane ring in chain A pocket, but no specific hydrogen bonding;

B - indane ring in chain A pocket, hydrogen bond from Ser133 to the substrate carbonyl group (C21=O);

C - indane ring in chain B pocket, hydrogen bond from Ser133 to the substrate carbonyl group (C21=O), but in a different orientation than pose B (Figure 3-31);

D - indane ring in chain B pocket, hydrogen bond from Thr131 to the substrate carbonyl group (with H-bond network with Asn117 and Tyr16).

Subsequently, the *Enlighten* protocols (van der Kamp, 2018) were used to simulate the protein-product complex. The PREP protocol prepares the protein-ligand system for simulation (with simulation restricted to a 20 Å sphere centred on the product). The STRUCT protocol performs a brief structural optimisation (using simulated annealing and energy minimisation), and the DYNAM protocol then performs a molecular dynamics simulation.

Poses C and D are shown in Figure 3-30 (after brief structural optimisation) with the main difference in binding pose arising from the side-chain orientation of Asn117.

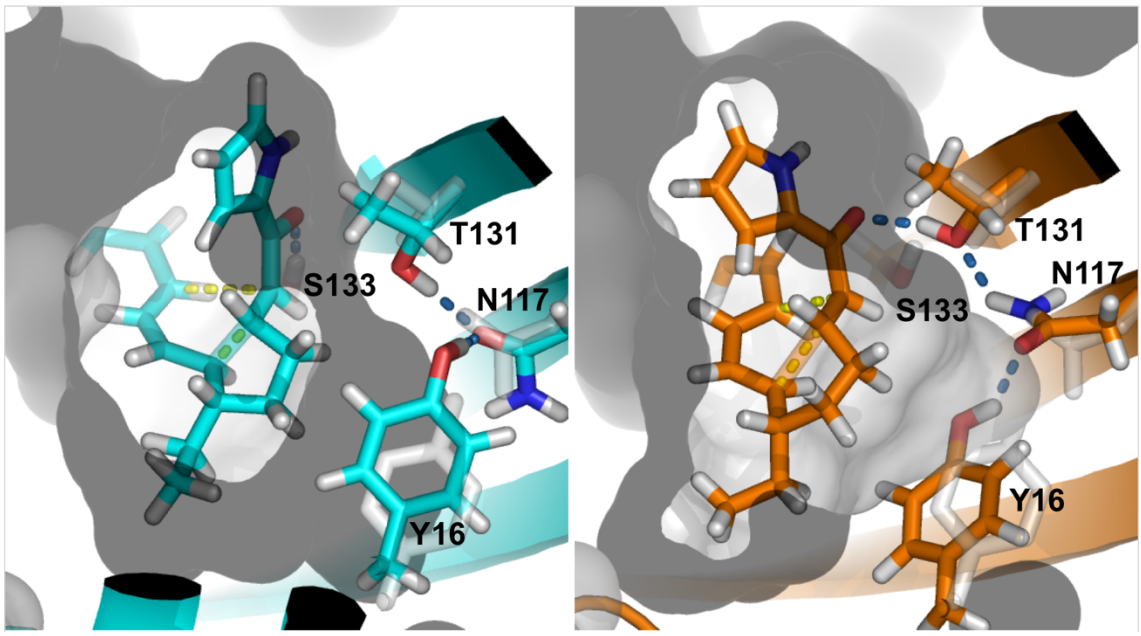


Figure 3-30 Representative transition state structures of indanomycin docked in the wild-type IdmH chain B active site. *Left – pose C, indanomycin C21=O hydrogen bonds to Ser133. Right – pose D, indanomycin C21=O hydrogen bonds to Thr131. The residue positions from the original crystal structure are shown in light grey.*

3.9.2 QM/MM simulations of the backwards reaction

Eleven quantum mechanics/molecular mechanics (QM/MM) reaction simulations were run for all four possible enzyme-substrate complexes, using the same protocol as described in Byrne et al. (Byrne et al., 2016). Representative (approximate) transition state structures of these four are shown in Figure 3-31, together with the free energy profile obtained from all eleven simulations together. Based on the energy barriers, options C and D (Figure 3-30) were found to be the most likely.

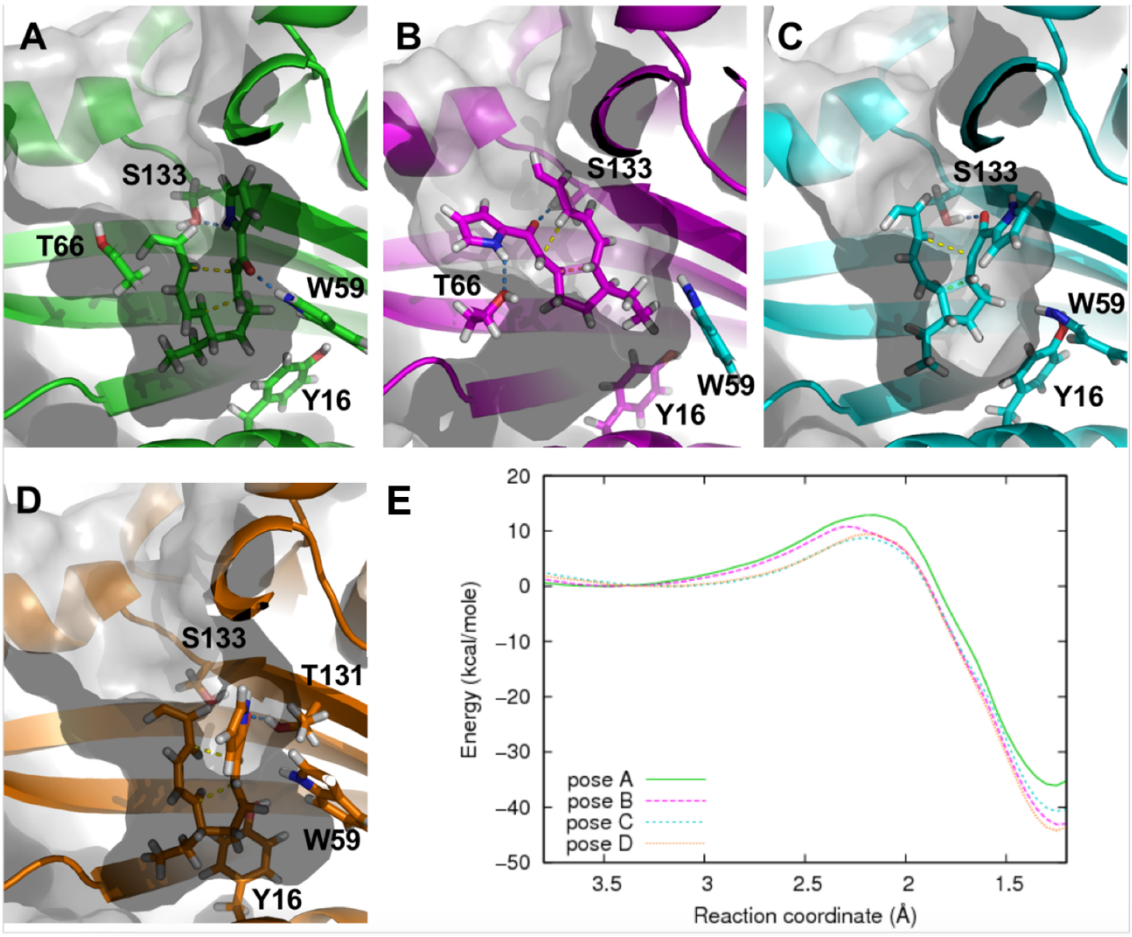


Figure 3-31 The results of QM/MM simulations. *Panels A-D show representative (approximate) transition state structures of the four possible enzyme-product complexes together with their energy profiles shown in panel E. The ring system (that is either cyclised or not before IdmH forming the indane ring) is not shown for clarity, but was present in all simulations.*

The low barriers can be rationalised based on the electron-withdrawing nature of the hydrogen bond (from Ser133 or Thr131, depending on the pose). Between the two, pose D was deemed to be the most likely, based on a good fit with the

cavity and some possible catalytic stabilization by Trp59. The reaction barrier for option A was found to be the highest, which could be expected in terms of the lack of a (good) hydrogen bonding interaction in this pose; during the QM/MM simulations, a hydrogen bond with Trp59 can be formed (see Figure 3-31), but this interaction is more transient and thus does not offer the same benefit. Molecular dynamics and QM/MM simulation of option B led to a poorer fit with the active site cavity (Figure 3-31).

The chemical shift perturbation mapping results (Figure 3-29) further support the observations from the QM/MM simulations. Pose D involves a hydrogen bonding network between Thr131, Asn117 and Tyr16 as well as catalytic stabilisation from Trp59. According to the NMR titration experiments, these residues (except for Thr131 which has not been assigned) are involved in protein-ligand interactions as measured by significant CSPs (0.22 ppm for Tyr16, 0.12 ppm for Asn117 and 0.20 ppm for Trp59). It is also worth noting that Ser133, involved in a hydrogen bonding network as part of pose C, displayed a CSP (0.012 ppm) below the level of significance (0.07 ppm) making the pose C less likely than pose D.

In summary, CSP mapping, *in-silico* docking and QM/MM simulations were used to propose the likely reactive IdmH-substrate complex. Together with the hydrogen bond, IdmH is likely to promote the correct conformation for the indanomycin precursor **4** which results in the catalysis of the indane ring cyclisation reaction.

3.10 Summary

This chapter describes the steps taken to investigate the structure and function of a putative polyketide cyclase IdmH from the indanomycin biosynthetic pathway.

The gene encoding IdmH was cloned from the genomic DNA of *S. antibioticus* to construct a pET28_IdmH vector (Figure 3-2). Soluble expression of IdmH was achieved in *E. coli,* and IdmH was purified to homogeneity by nickel affinity and gel filtration chromatography steps (Figure 3-4). Purity was confirmed by SDS-PAGE (Figure 3-4) and mass spectrometry (Figure 3-5) analysis.

Native MS analysis together with gel filtration revealed that IdmH forms homo-dimers in solution (Figure 3-5). Treatment of the dimeric IdmH with DTT, a

reducing agent, did not result in the shift towards monomeric species in the MS spectrum (Figure 3-6), which suggested that the interaction between the monomers was not occuring via a disulphide bridge.

Circular dichroism spectroscopy and static light scattering techniques were employed to estimate the fold of IdmH and its stability at different pH conditions. The recorded CD spectrum suggested the presence of both α-helices and β-sheets at similar ratios (Figure 3-8) while SLS showed that IdmH is the most stable at pH above 7.0 and temperatures below 40 °C (Figure 3-9).

To determine the atomic model of wild-type IdmH, twelve commercially available crystallisation screens were set up, which eventually led to an optimised single rod-shaped crystal (Figure 3-11). A dataset of this crystal was collected but molecular replacement approaches to solve its structure were not successful. The protein was labelled with selenomethionine for experimental phasing, but no 3D crystals grew in this case.

To improve crystallisation, seven surface entropy reduction IdmH mutants were created (Obajdin, 2018), one of which (IdmH-Δ96-104) exhibited high propensity to form large crystals in a cubic space group F23 (Figure 3-12). A crystal structure of this IdmH variant was solved using selenium SAD and was subsequently used to solve the wild-type IdmH structure by homologous replacement.

Wild-type IdmH dimer is made out of two α+β barrels with a large hydrophobic cavity located in the core of each subunit. This hydrophobic cavity is lined with aromatic residues and most probably constitutes the active site of the enzyme (Figure 3-15, Figure 3-18).

To investigate whether indanomycin, the final product of this NRPS/PKS, can bind to IdmH, NMR spectroscopy titration experiments were performed. Expression of single and triple-labelled IdmH allowed the collection of 2D and 3D NMR spectra which resulted in the successful assignment of approximately 88% of backbone resonances (Figure 3-26). Ligand binding experiments showed that at 1:1 protein to ligand ratio, significant chemical shift perturbations took place (Figure 3-27), which is indicative of a binding event between IdmH and indanomycin. To investigate the binding interface between IdmH and indanomycin, a titration experiment was performed with sub-stoichiometric protein to ligand ratios and chemical shift perturbations were mapped onto the structure of wild-type IdmH (Figure 3-29). The results of the CSP mapping

revealed that the majority of the residues in the hydrophobic cavity underwent a significant chemical shift perturbation providing further evidence that the cavity comprises the active site of IdmH and binds indanomycin.

To investigate the binding at a molecular level, *in-silico* methods were applied. Indanomycin was docked in the active sites of chains A and B and the most probable binding modes were taken forward for QM/MM simulations (Figure 3-30, Figure 3-31). The results suggested that the most likely binding mode was pose D, which involves a hydrogen bonding network between Thr131, Asn117, Tyr16 and C21=O of indanomycin (Figure 3-31). These observations were supported by the CSP mapping since both Tyr16 and Asn117 exhibited a large CSP upon titration with indanomycin.

In summary, this chapter reports the first structure of the putative indanomycin cyclase IdmH and its variant IdmH-Δ96-104. The crystal structures revealed that an IdmH monomer folds into an α+β barrel with a large hydrophobic cavity located in the middle. We propose this to be the active sites of the enzyme. Chemical shift perturbation mapping, *in-silico* docking and QM/MM simulations were used to propose the likely reactive IdmH-substrate complex. The results of these analyses are consistent with IdmH acting as a Diels-Alderase to catalyse the formation of the indane ring in indanomycin.

Chapter 4 Structural and kinetic characterisation of IdmO, a fully-reducing polyketide module from indanomycin synthase

4.1 Introduction

The modular polyketide synthases are complex multi-enzyme proteins responsible for the production of a diverse range of natural products of high therapeutic value (Cummings et al., 2014). Understanding how these complexes are structured and how different modules interact with each other is essential in order to rationally engineer these systems for the production of novel natural products with desirable pharmaceutical properties.

Indanomycin is produced by a hybrid non-ribosomal peptide synthetase-polyketide synthase (C. Li et al., 2009). Findings from pikromycin PKS (Dutta et al., 2014; Whicher et al., 2014), together with recent advantages in cryo-electron microscopy (Bai et al., 2015; Cheng, 2015), have encouraged us to apply cryo-EM methods to study the indanomycin megasynthase.

IdmO, the fourth subunit of the indanomycin megasynthase, is a particularly attractive target for structural studies by cryo-EM (Figure 4-1) for a number of reasons. First, the *idmO* gene has a length of 6 kb which can be amplified by PCR. Second, the gene translates into a 230 kDa protein which can be recombinantly expressed *E. coli,* yet is big enough for cryo-EM. Third, and most importantly, IdmO is the sole member of the fourth subunit of this assembly line and contains all six possible catalytic domains, making it the first, to our knowledge, fully-reducing type I mPKS module being studied structurally to date.

This chapter documents the cloning, expression and purification methods employed to produce recombinant IdmO. Enzymatic and structural characterisation of IdmO is also covered.

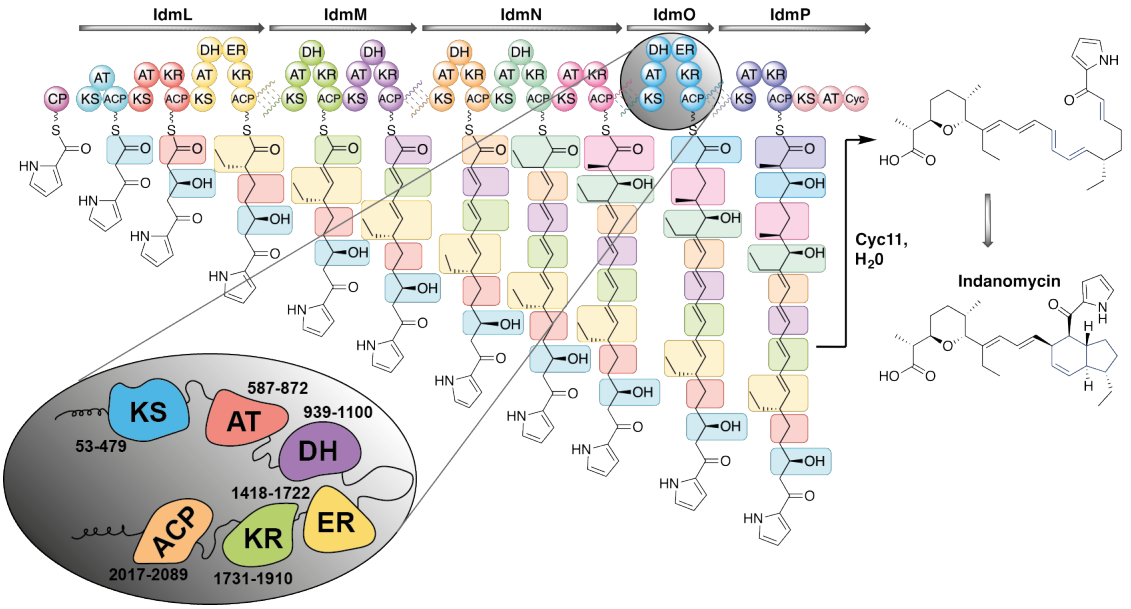


Figure 4-1 Indanomycin synthase from S. antibioticus. *Each module is responsible for a single chain elongation step. Post-PKS enzymes Cyc11 and IdmH are postulated to be involved in cyclising the final product. IdmO, the ninth module of this PKS is shown in greater detail together with the amino acid numbers for each of its six domains (KS – ketosynthase, AT – acyltransferase, DH – dehydratase, ER – enoyl reductase, KR – ketoreductase, ACP – acyl carrier protein).*

4.2 Cloning, expression and purification of IdmO

4.2.1 Cloning of the *idmO* gene

The amplification of the indanomycin biosynthetic genes from *S. antibioticus* genomic DNA is problematic due to several reasons: high GC content (>70%), the presence of repetitive sequences, and large size of the genes. Initial attempts using classical PCR method did not result in a successful amplification of the right size DNA fragment and showed high level of non-specific amplification. To overcome this, a nested polymerase chain reaction (nPCR) approach was employed (Figure 4-2) (Haff, 1994).

Another problem associated with GC-rich DNA is the formation of stem-loop secondary structures within the template. This can lead to polymerase jumping during PCR amplification, resulting in smaller than expected DNA products. A well-known strategy to relieve these secondary structures is adding organic additives, such as DMSO, betaine, polyethylene glycol, glycerol and formamide, to the PCR mixture (Kang et al., 2005; Hardjasa et al., 2010).

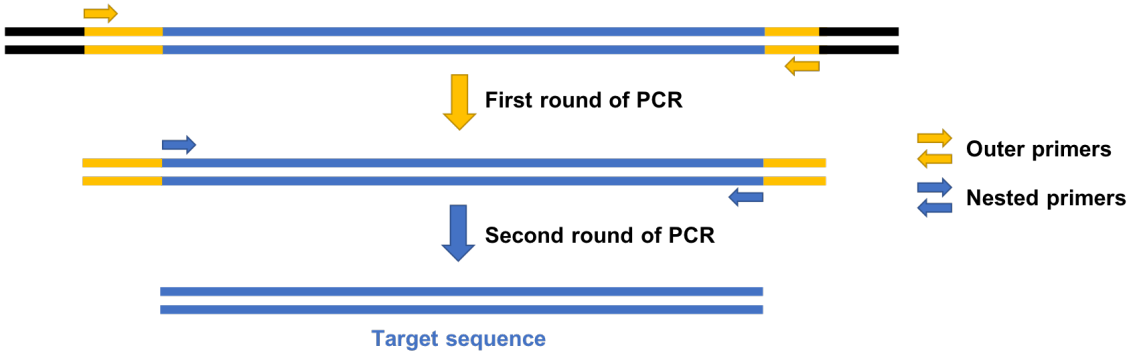


Figure 4-2 Nested polymerase chain reaction outline. *This method involves two sets of primers (outer and inner, or "nested"), used in two successive runs of PCR. The first run amplifies a larger sequence spanning the sequence of interest, and then the product from the first round is used as a template in the second round of PCR.*

Nested PCR for the region spanning *idmO* was performed using an increasing percentage of DMSO, as well as with NEB GC enhancer (Figure 4-3a). The "cleanest" product was obtained in the GC enhancer-containing PCR reaction; therefore, the product of this reaction was used as a template in the second round of nPCR (Figure 4-3b).

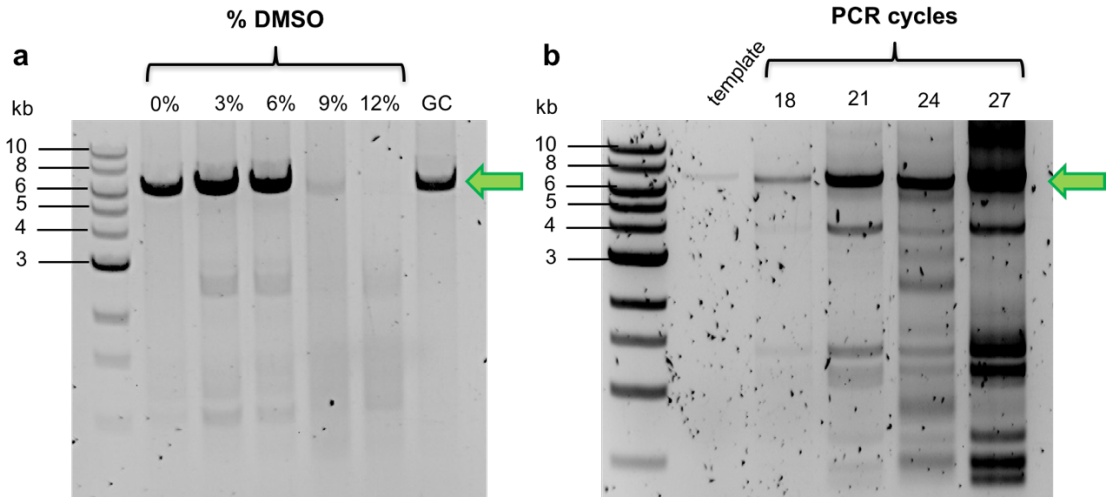


Figure 4-3 Amplification of the idmO gene. *(a) First round of nPCR of idmO. A sequence spanning the idmO gene was amplified in the presence of DMSO solvent and GC enhancer. The latter resulted in the cleanest PCR product, which was chosen to be used for the second round of nPCR. (b) Second round of nPCR using GC enhancer PCR products from the first reaction as a template. Non-specific amplification was shown to increase substantially in later rounds of PCR; therefore, PCR product from 21 cycles of PCR was chosen for the cloning reaction.*

Primers for the amplification of the *idmO* gene during the second round of nPCR were designed to add a C-terminal hexahistidine tag as well as overhangs on both ends of the amplicon to make it complementary to the vector. The product from 21 cycles of the second round of PCR was mixed with linearised vector, and Gibson assembly protocol was executed (Gibson et al., 2009). Unfortunately, the Gibson assembly approach was unsuccessful, and the FastCloning strategy was attempted instead (C. Li et al., 2011).

In this method, unpurified PCR products containing a desired gene insert are combined with linearised vector in the presence of *Dpn*I and in the absence of additional enzymes, incubated for one hour and transformed into competent cells. This approach yielded over fifty *idmO* colonies from a single transformation into *E. coli* XL10 ultracompetent cells. The resulting colonies were screened by colony PCR and sequencing with T7P and T7TERM universal primers (Table 2-4). Promising clones were fully sequenced using custom designed primers (Table 2-4), and a successful clone, termed pET28_IdmO, was used for further studies.

4.2.2 Expression and purification of IdmO

For the initial expression screen, pET28_IdmO construct was expressed in 50 ml cultures in 2TY medium in both *E. coli* BL21(DE3) and Lemo21(DE3) strains. For Lemo21(DE3), various amounts of L-rhamnose were added to the cultures to tune the expression (Wagner et al., 2008). Adding L-rhamnose increases the level of lysozyme (lysY), the natural inhibitor of T7 RNA polymerase, which in turn modulates the expression.

50 ml cultures were grown at 15 °C overnight, and the recombinant IdmO yield was assessed by performing a small-scale immobilised-metal affinity chromatography (IMAC) purification (His SpinTrap, GE). Figure 4-4 shows the SDS-PAGE analysis of the harvested cell and Figure 4-5 of the elution samples.

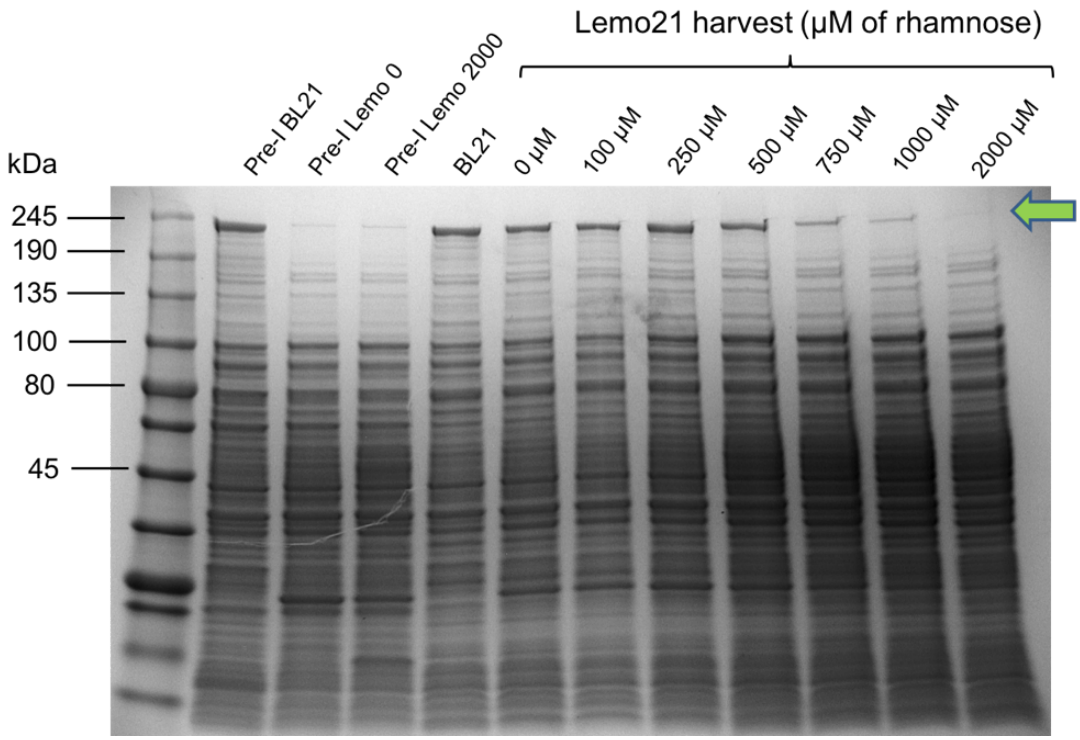


Figure 4-4 Expression of IdmO in E. coli BL21(DE3) and Lemo21(DE3) cells. *Varying the concentration of L-rhamnose "tunes" the expression levels by quantitative inhibition of T7 RNA Polymerase. Whole cells were mixed with SDS-PAGE loading dye, boiled and loaded on the SDS-PAGE gel. The calculated mass of IdmO is approximately 230 kDa. The green arrow indicates the position of IdmO-containing bands.*

As can be seen from Figure 4-4 second lane, an *E. coli* BL21(DE3) culture before IPTG induction (termed "Pre-I BL21") already contains a band for recombinant IdmO, indicating "leaky" expression. Since cultures are grown at 37 °C before induction, this can become problematic as recombinant expression of large proteins at this temperature can lead to the formation of inclusion bodies. *E. coli* Lemo21(DE3) pre-induced samples, on the other hand, do not have a prominent band at ~240 kDa, indicating a tighter control of expression. Tuneable expression of the Lemo21(DE3) strain was attempted by adding L-rhamnose to the expression culture at levels from zero to 2000 µM.

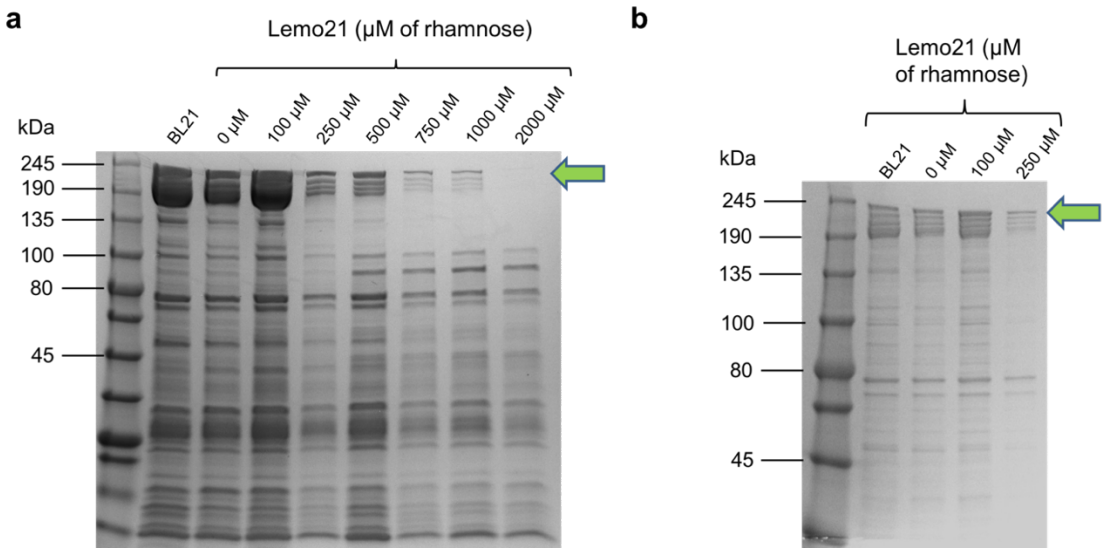


Figure 4-5 Small-scale SpinTrap purification results of IdmO expression. *(a) SpinTrap eluate samples containing soluble IdmO (indicated by the green arrow). For E. coli Lemo21(DE3) cultures, the expression was tuned by adding L-rhamnose to the media, which modulates the level of lysozyme (lysY), the natural inhibitor of T7 RNA polymerase. (b) Elution fractions from (a) were diluted three times. In both panels, it can be seen that E. coli Lemo21(DE3) 100 µM rhamnose condition resulted in most recombinant IdmO, while E. coli BL21(DE3) was a close second.*

Figure 4-5 shows that the most soluble recombinant IdmO was produced in the culture condition containing 100 µM L-rhamnose. Therefore, a large-scale IdmO expression (10 x 0.5 l) was performed in *E. coli* Lemo21(DE3) cells supplemented with 100 µM L-rhamnose. To minimise shear damage to IdmO, the cells were handled very gently, performing lysis with a cell disruptor rather than by sonication (Stathopulos et al., 2008) and carrying out all purification steps at 4 °C. An IMAC metal affinity chromatography eluate was further purified using a HiLoad 26/60 Superdex 200 Prep Grade size-exclusion column.

The size-exclusion elution profile contained four distinct peaks (Figure 4-6). The elution volume of the first peak was consistent with the void volume of the column and most likely comprised protein aggregates. The elution volume of the second peak was consistent with a ~230 kDa (globular) protein. Peaks three and four were most likely comprised of smaller than 230 kDa contaminant proteins and truncated variants of IdmO. A sample from each stage of the purification procedure was analysed by reducing SDS-PAGE and can be seen in Figure 4-6b.

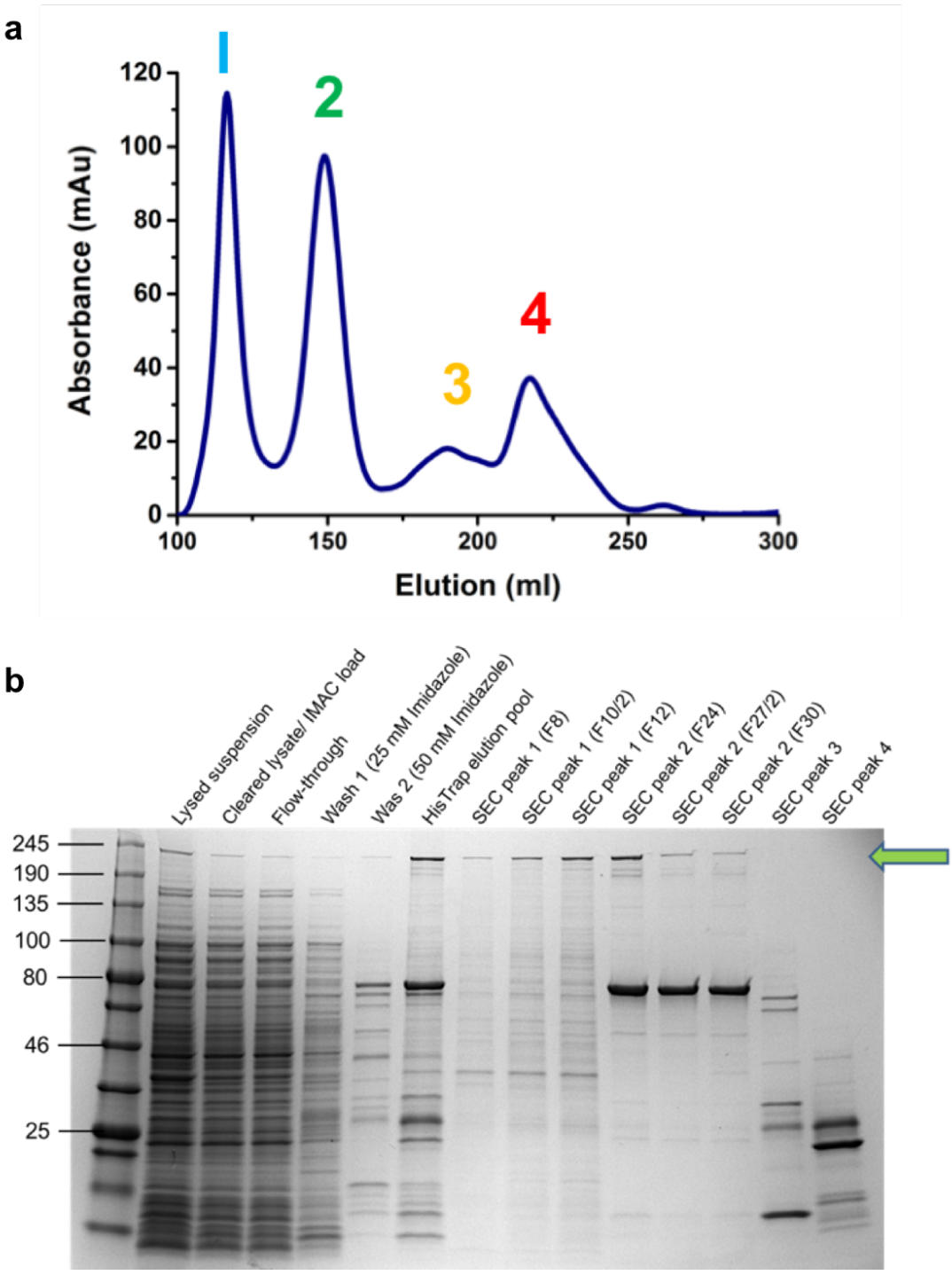


Figure 4-6 Large-scale purification of IdmO. *(a) Size-exclusion chromatography curve of IdmO. Separation was performed in 20 mM Tris pH 7.4, 50 mM NaCl at a flow rate of 2 ml/min. The protein eluted in four peaks, the first of which is the "void" volume peak containing aggregated and misfolded material. The elution volume of the second peak was found to be consistent with the 230 kDa IdmO, while the third and four peaks were most likely comprised of smaller protein contaminants. (b) SDS-PAGE analysis of samples obtained throughout the purification. The green arrow indicates the position if IdmO.*

As can be seen from Figure 4-6b, first two peaks both contained IdmO, while the second peak also contained an additional band of approximately 70 kDa. At first, it was thought that this band represented IdmO degradation products. To determine if that is the case, IdmO-containing fractions from peak 2 were concentrated and re-purified by size-exclusion chromatography one more time (Figure 4-7). The protein eluted at the same volume as the first time, indicating little or no protein degradation over time.

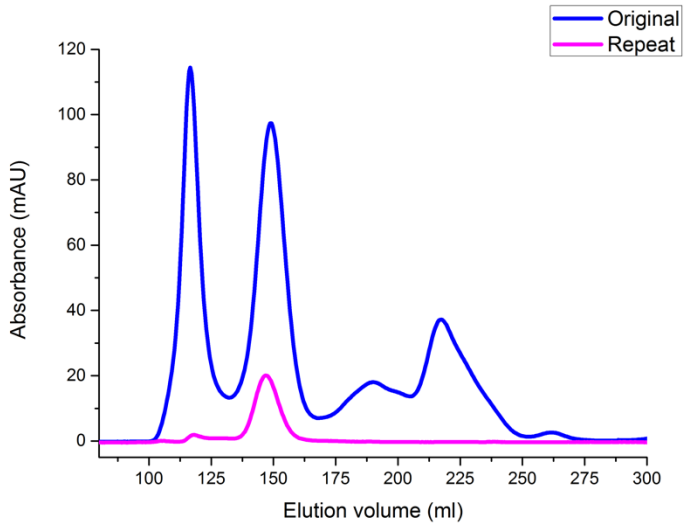


Figure 4-7 Repeating size-exclusion chromatography of IdmO. *Fractions from peak 2 from the first purification were concentrated and purified using the same size-exclusion chromatography column again. The peak eluted at the same volume indicating that peak 2 contains stable protein.*

To determine whether the 70 kDa band contained a His-tagged protein, an anti-His western blot was performed (Figure 4-8). The choice of the blot detection system (X-ray film rather than digital), the amount of protein loaded and the low percentage of the gradient gel resulted in the development of many non-specific bands in the blot (such as the ladder proteins). This is due to the anti-His antibody binding to the surface histidines of the non-his tagged proteins. Nevertheless, the absence of the 70 kDa band in the blot confirmed that this protein did not contain a hexahistidine tag, at least at the C-terminal end. Large spots at the top of the blot suggested that the top band was most likely His-tagged IdmO.

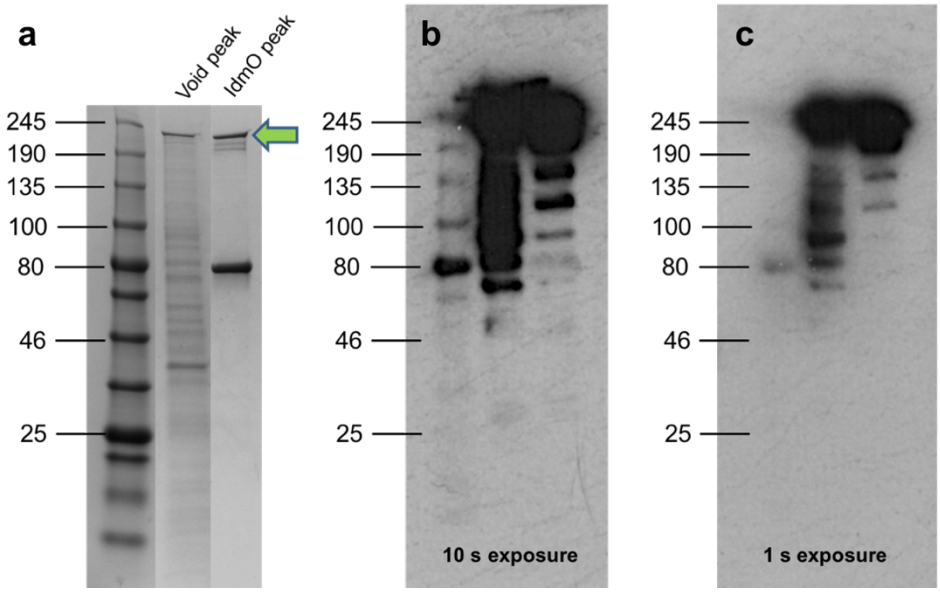


Figure 4-8 Anti-his Western blot of protein samples from first and second size-exclusion chromatography peaks of IdmO purification. *(a) SDS-PAGE gel. (b) Blot developed for 10 seconds. (c) Blot developed for 1 second. In both (b) and (c) panels, intense spots can be seen in the void peak lane (most likely corresponding to IdmO degradation products) and at the top of IdmO peak lane (corresponding to the full-length IdmO). No band can be seen ~70 kDa indicating that this protein does not contain a hexahistidine tag, at least on its C-terminal end.*

4.2.3 Identification of the 70 kDa contaminant by mass spectrometry

A protein sample from the second size-exclusion chromatography peak was submitted for accurate mass measurement by LC/MS. Only one species was observed with a mass of 74.3 kDa (Figure 4-9). IdmO, with an expected mass of 228 kDa, was not detected, most probably due to size limitations of the LC/MS system.

The same sample, in the form of a 1D gel, was also submitted for protein identification analysis by peptide mass fingerprinting (Webster and Oxley, 2012). The top band in the SDS-PAGE gel was found, as expected, to be IdmO (Table 4-1) while the contaminating band was identified to be a bifunctional polymyxin resistance protein ArnA (Table 4-2).

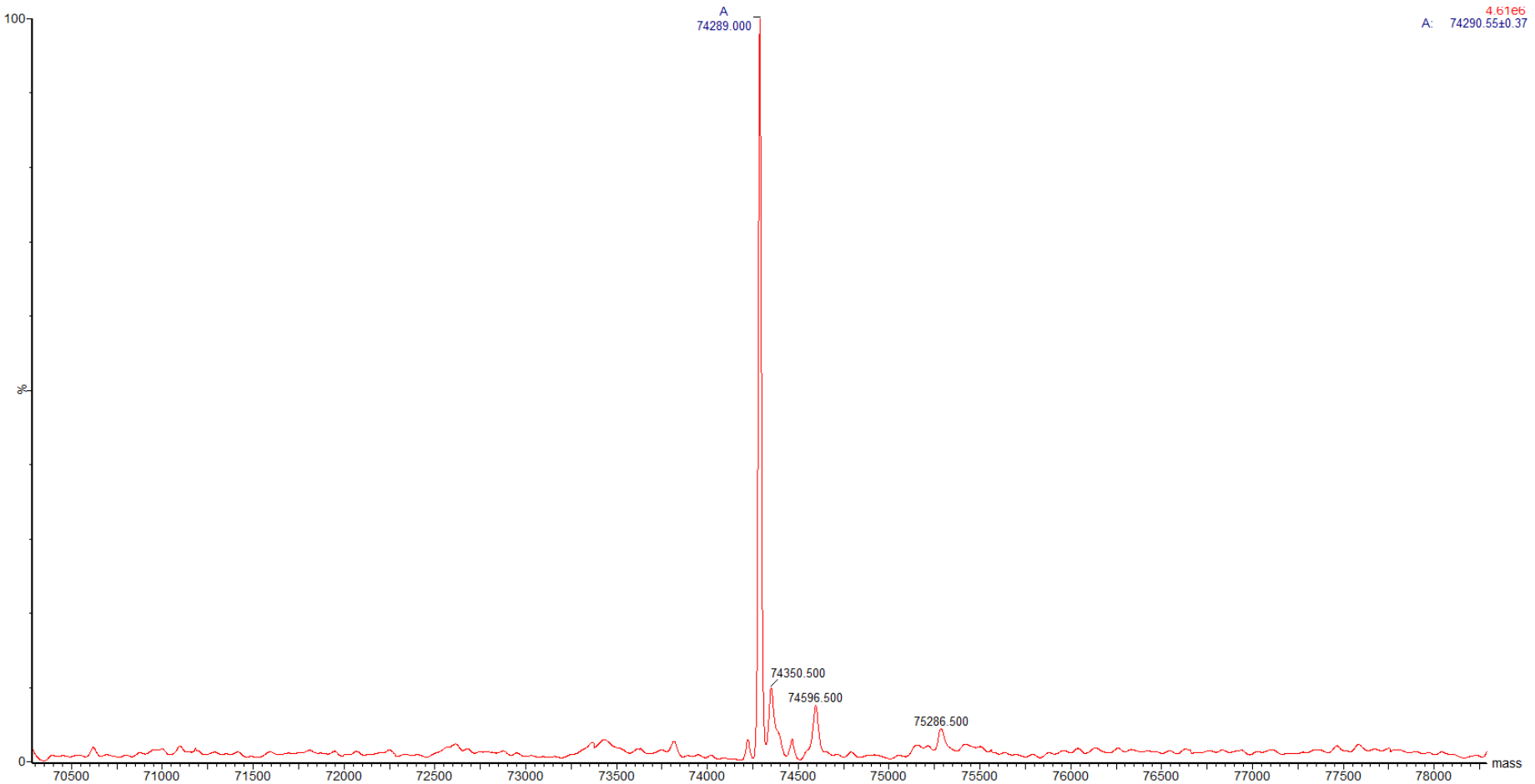


Figure 4-9. Analysis of IdmO size-exclusion peak contents by LC/MS. *Here, a deconvoluted spectrum is shown with a peak for a protein with a measured mass of 74290.55.*

Table 4-1 Peptide mass fingerprinting results of IdmO band in the SDS-PAGE gel.

Protein Group	Accession	Coverage (%)	# Peptides	Avg. Mass	Description
1	sp|1605-145_IdmO|1605-145_IdmO	30	52	228853	sp|1605-145_IdmO|1605-145_IdmO
2	sp|P00761|TRYP_PIG	16	4	24409	Trypsin OS=Sus scrofa PE=1 SV=1

Table 4-2 Peptide mass fingerprinting results of the 70 kDa contaminant band in the SDS-PAGE gel.

Protein Group	Accession	Coverage (%)	# Peptides	Avg. Mass	Description
1	sp|B1X8W8|ARNA_ECODH	76	74	74289	Bifunctional polymyxin resistance protein ArnA OS=*Escherichia coli* (strain K12 / DH10B) GN=arnA PE=3 SV=1
1	sp|P77398|ARNA_ECOLI	76	74	74289	Bifunctional polymyxin resistance protein ArnA OS=*Escherichia coli* (strain K12) GN=arnA PE=1 SV=1
1	sp|C4ZU97|ARNA_ECOBW	76	74	74289	Bifunctional polymyxin resistance protein ArnA OS=*Escherichia coli* (strain K12 / MC4100 / BW2952) GN=arnA PE=3 SV=1
1	sp|B1IXT2|ARNA_ECOLC	76	74	74289	Bifunctional polymyxin resistance protein ArnA OS=*Escherichia coli* (strain ATCC 8739 / DSM 1576 / Crooks) GN=arnA PE=3 SV=1
6	sp|P00761|TRYP_PIG	29	6	24409	Trypsin OS=Sus scrofa PE=1 SV=1

Accession: the accession number of the protein as in FASTA database. **Coverage**: the percentage of the protein sequence covered by supporting peptides. **# Peptides**: the number of high-confidence supporting peptides. **Avg. Mass**: protein mass calculated using average mass. **Description**: protein's header information as in FASTA database.

ArnA is a 74.3 kDa enzyme involved in lipopolysaccharide biosynthesis and resistance of Gram-negative bacteria to certain cationic antimicrobials, such as polymyxin (Breazeale et al., 2002). ArnA possesses several histidine clusters on the surface of its active hexamer, which is the reason why this protein is a recurring contaminant during IMAC (Robichon et al., 2011; Gatzeva-Topalova et al., 2005). ArnA forms a trimer of dimers with the total mass of the complex reaching 445 kDa, which explains why ArnA co-elutes together with the 460 kDa IdmO dimer.

Strategies to prevent co-purification of ArnA include optimisation of purification steps, changing the purification tag and using an *E. coli* Nico21(DE3) strain in which common *E. coli* metal binding proteins (SlyD, ArnA, Can and AceE) are tagged to enable removal by chitin affinity chromatography (Robichon et al., 2011).

4.2.4 Negative stain EM of IdmO and ArnA mixture

To see whether the IdmO could be visualised by electron microscopy, a negatively stained grid of IdmO, contaminated with ArnA as described in section 4.2.3, was produced. Individual particles could be seen in the micrograph (Figure 4-10a); therefore, a small dataset of micrographs was collected. Individual particles were auto-picked and averaged into 2D classes (Figure 4-10b). Unfortunately, the most prominent classes showed a strong resemblance to the crystal structure of ArnA (Figure 4-10b and c) (Gatzeva-Topalova et al., 2005) indicating a need for better sample purification strategy before structural characterisation of IdmO by electron microscopy could be attempted.

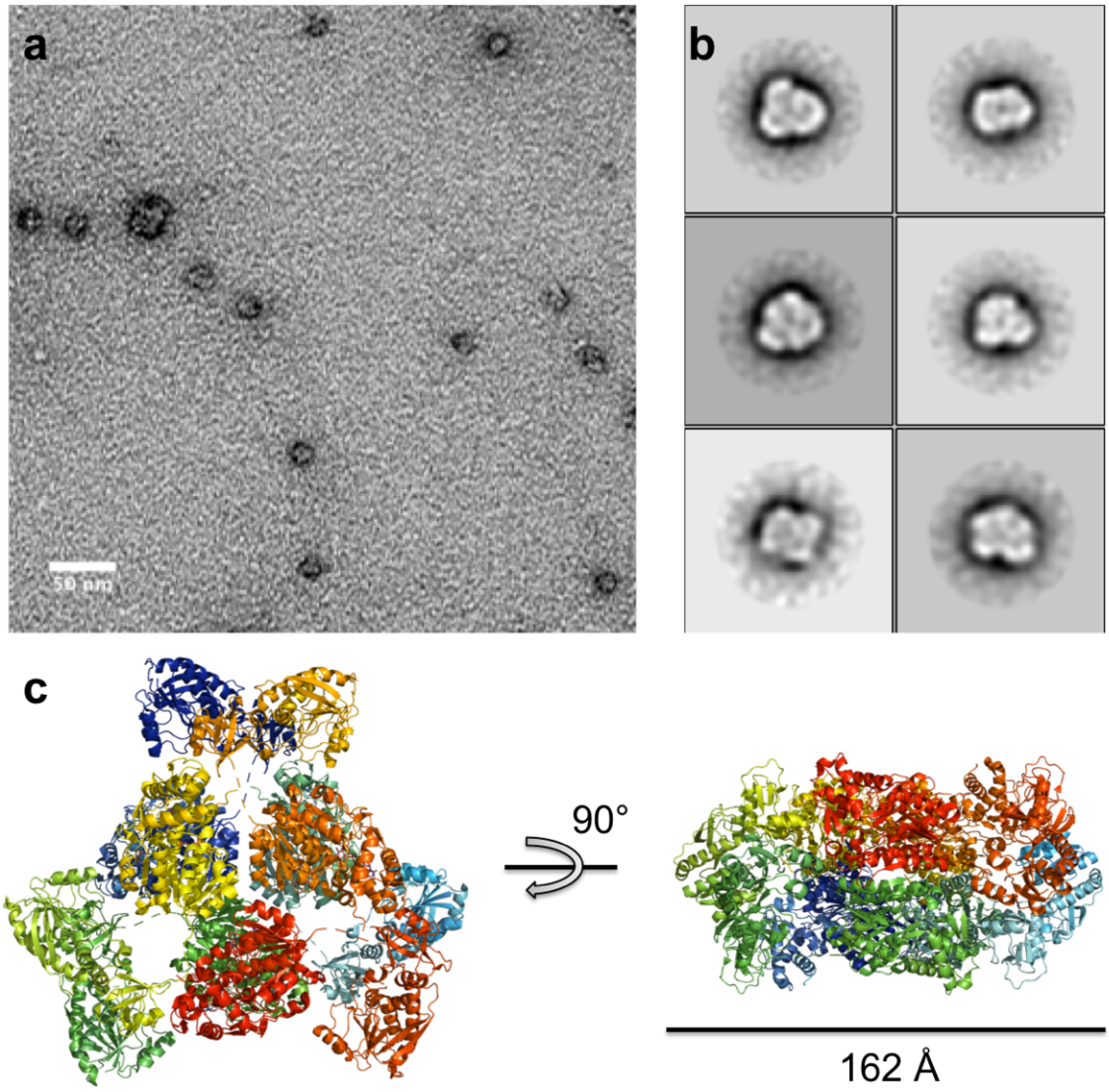


Figure 4-10 Negative stain EM of IdmO contaminated with ArnA. *(a) Negative stain micrograph was taken at 30,000 magnification. The protein sample was applied to glow-discharged amorphous carbon grids and stained with 1% uranyl acetate. Scale bar represents 50 nm. (b) 2D class averages which resulted from the collection of 50 micrographs, auto-picking of particles and 2D classification. (c) Crystal structure of ArnA (Gatzeva-Topalova et al., 2005) displaying strong resemblance to 2D classes in panel (b). (c) was made in PyMOL using PDB code 1Z7E.*

4.3 Purification process optimisation

Negative stain EM analysis indicated that an alternative purification strategy needs to be employed for structural studies of IdmO to be feasible. This subsection contains an outline of attempted approaches.

4.3.1 Tuning the concentration of imidazole in the elution buffer

It was hypothesised that tuning the imidazole concentration during the affinity chromatography might help to separate ArnA from IdmO. Therefore, IMAC purification was performed using a step-gradient approach, whereby the bound proteins were eluted using 50 mM, 100 mM, 125 mM, 150 mM and 175 mM imidazole concentrations. As can be seen from Figure 4-11, most IdmO, together with ArnA and other contaminants, eluted at 100 mM imidazole concentration making this approach unsuited for the removal of ArnA.

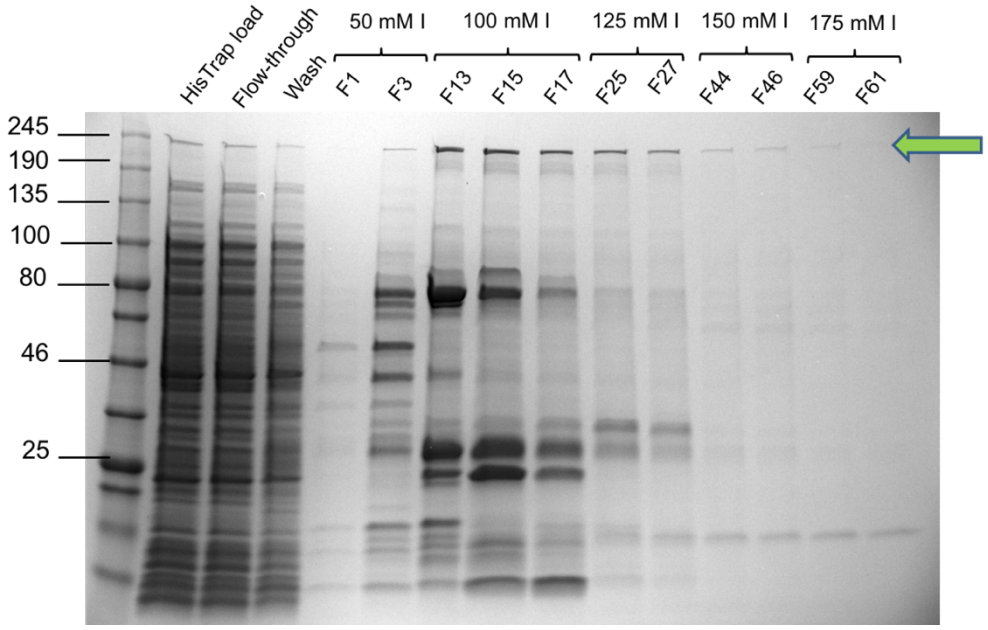


Figure 4-11 Tuning of the HisTrap elution step. *Elution was performed using increasing concentration of imidazole in a step-like fashion in an attempt to separate the elution of IdmO from the elution of ArnA. Most IdmO, together with ArnA and other contaminants, eluted at 100 mM imidazole concentration making this approach unsuited for the removal of ArnA.*

4.3.2 Exploring the differences in the isoelectric point of IdmO and ArnA

By using the ProtParam server (Gasteiger et al., 2005), the isoelectric points (pI) of IdmO (4.9) and ArnA (6.4) were calculated. The difference in pI values between the two proteins indicated that it might be possible to separate the two proteins by anion exchange chromatography. This is based on the assumption that at pH 7.0, IdmO will bind the anion exchanger more strongly than ArnA. To test this hypothesis, the elution pool from the 100 mM imidazole step was dialysed into a

low salt, no imidazole buffer (pH 7) and loaded onto an anion exchange column. Bound protein was eluted using an increasing salt concentration (Figure 4-12). Unfortunately, IdmO co-eluted with ArnA indicating that utilising anion chromatography is not likely to result in the complete separation of the two species. Therefore, it was decided that changing the purification tag or removing ArnA before the nickel affinity chromatography might be a more promising approach to produce pure IdmO.

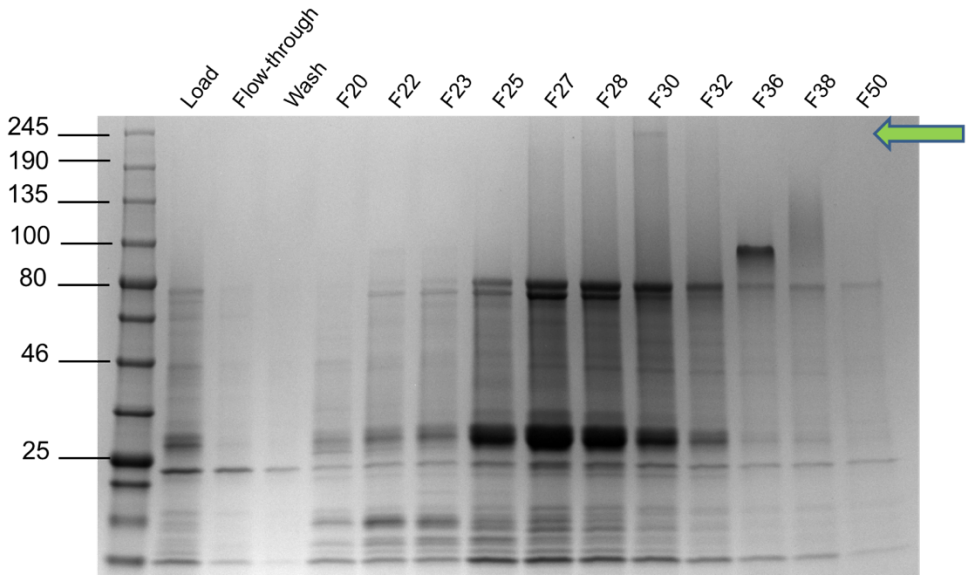


Figure 4-12 SDS-PAGE analysis of the anion exchange chromatography fractions. *IdmO (a band in F30 fraction noted by the green arrow) co-eluted with the contaminating protein ArnA (~70 kDa) indicating that the separation of IdmO from ArnA will not be possible using this chromatographic approach.*

4.3.3 Redesigning the construct to include a second affinity tag on the N-terminus

Affinity tags are indispensable tools for purifying high amounts of recombinant proteins of interest. However, no affinity tag is ideal with respect to yield, purity and cost (Lichty et al., 2005; Waugh, 2005). While a His tag purification procedure gives high yields and has low costs, metal binding proteins (as exemplified in section 4.2.3) present a need for further purification steps. Size-exclusion chromatography is one of the most widely used polishing steps in recombinant protein purification; however, it is not always optimum due to high costs (especially at large-scale) and, in this case, when contaminating proteins have very similar size. Combinatorial tagging, on the other hand, can deliver maximum benefit of affinity tags (Waugh, 2005; J. Nilsson et al., 1996). In such cases, a

high-throughput affinity tag (such as His tag) is used for the first chromatography step and then a second chromatography step with a different tag is performed to remove the unwanted proteins from the first purification step. What is more, placing the affinity tags on either terminus of the polypeptide allows to select only for the full-length proteins, and, in some cases, no buffer exchange is needed between the two chromatography steps allowing the process to be high-throughput.

We have chosen to keep the His tag on the C-terminus and add an additional affinity tag on the N-terminus of IdmO. Strep II tag was chosen because it is an excellent polishing affinity tag due to moderate cost, yield and extremely high specificity (Lichty et al., 2005; Schmidt and Skerra, 2007). Furthermore, both His and Strep II tags are small and hence do not significantly increase the metabolic burden to the expression host, and there is no need to exchange the buffer between the affinity columns, making the process high-throughput. As such, it was decided to make a new IdmO construct, termed Streptag/TEV (2). The Streptag/TEV construct would contain a Strep-tag on the N-terminus, separated from the gene of interest by a Tobacco Etch Virus (TEV) protease cleavage site, and a His-tag on the C-terminus (Figure 4-13).

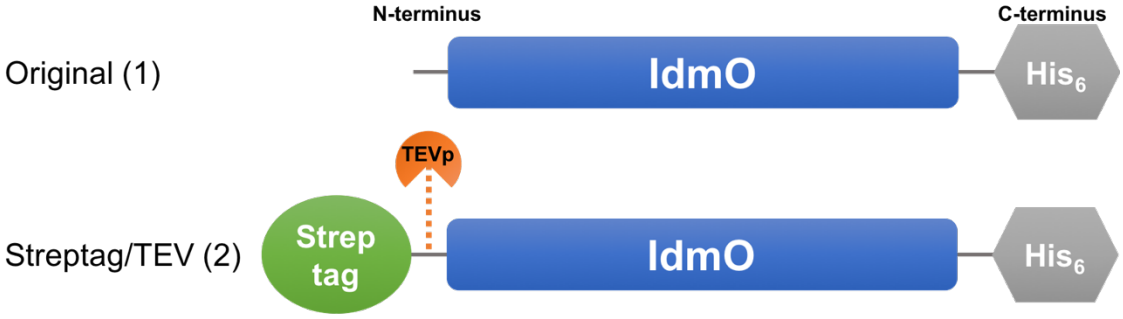


Figure 4-13 IdmO constructs used in this study. *The original construct, created as described in section 4.2.1 harbouring a C-terminal His-tag. The second construct, Streptag/TEV, harbours an N-terminal Strep-tag, separated from the gene by a TEV protease cleavage site, and a C-terminal His tag.*

4.3.3.1 Constructing the double-tagged IdmO vector

To construct the Streptag/TEV (2) double-tagged IdmO vector, two components were needed. First, an 80 bp fragment containing the Strep II tag and TEV protease cleavage size was amplified from the pET28c_SSstrep_TEV_GO plasmid (a kind gift from Prof. Michael McPherson, Faculty of Biological Sciences, University of Leeds, UK) (Figure 4-14a). This fragment was designed to have overhangs to the pET vector allowing a ligation-independent cloning strategy to be used (C. Li et al., 2011). Second, the vector harbouring the *idmO* gene with the C-terminal His tag had to be amplified. The attempts to linearise the whole 11.5 kb vector in one PCR reaction using identical cycling parameters in each cycle did not yield a single band of appropriate size (Figure 4-14b). Successful amplification was achieved using a touchdown PCR cycling protocol (Green and Sambrook, 2018), a method for increasing specificity of PCR reactions. In touchdown PCR, the initial annealing temperature is several degrees above the estimated T_m of the primers (72 °C was used in this case) and is gradually reduced between each cycle (-0.5 °C/cycle in this case) which favours very specific base pairing between the primer and the template. In two-step PCR, the annealing step is omitted altogether which in theory should also minimise non-specific base pairing; however, in this case, no products of 11.5 kb size were observed when using this method (Figure 4-14c).

Strep II tag and IdmO vector DNA fragments were mixed in 1:5, 1:1, 10:1 and 100:1 ratios, and incubated for an hour at 37 °C in the presence of *Dpn*I restriction enzyme as described by the FastCloning procedure (C. Li et al., 2011). This protocol yielded a total of ten colonies, which were all screened by performing a double digest with *Pst*I and *Xba*I restriction enzymes and by DNA sequencing using T7 and T7term universal primers (Table 2-4). One clone (termed pET28_stIdmO_c1) had the correct Strep II fragment insert and therefore was transformed into *E. coli* BL21(DE3) for protein expression.

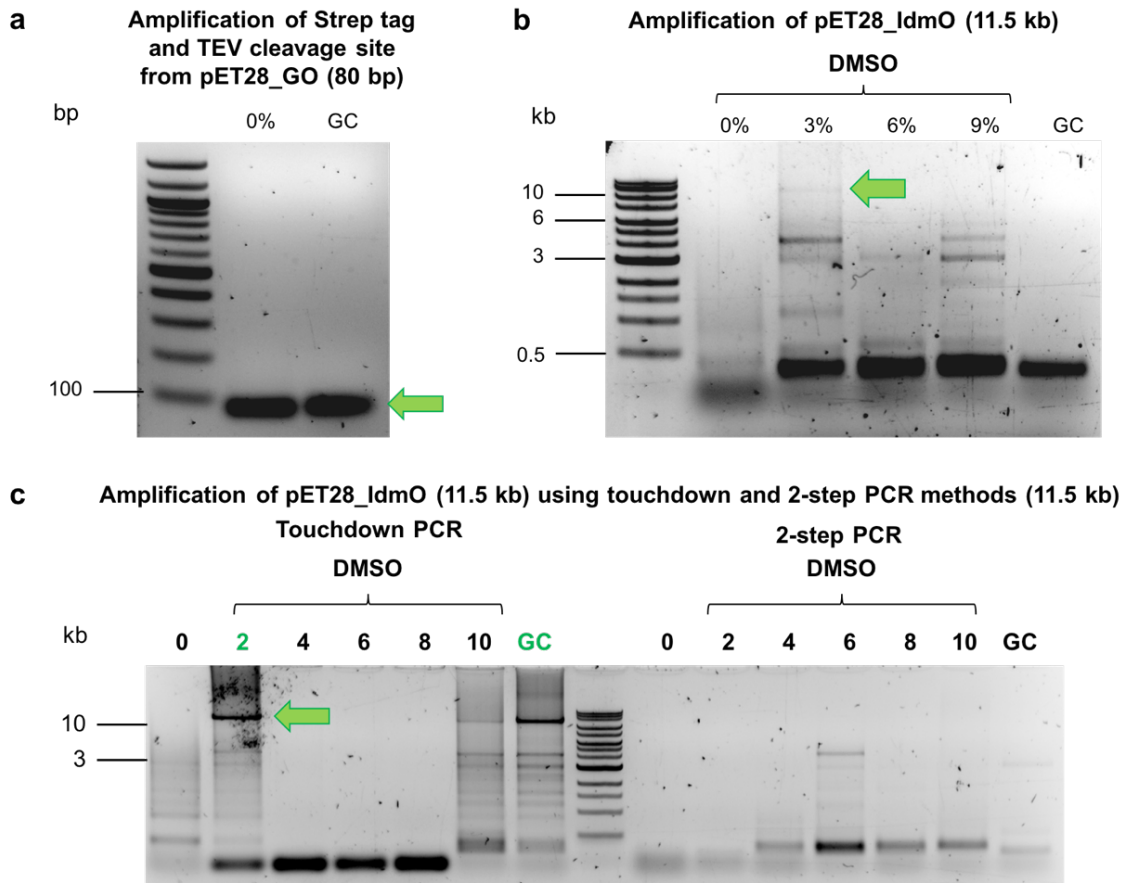


Figure 4-14 Constructing the double-tagged IdmO vector. *(a) Amplification of the Strep II tag and the TEV protease cleavage site was achieved using pET28c_SSstrep_TEV_GO vector as a template. (b) Attempts to linearise the pET28_IdmO plasmid using classic PCR method. A number of reactions with and without solvents (DMSO and GC) were set up to alleviate problems associated with high GC content of Streptomyces genes. (c) Linearisation of pET38_IdmO vector by touchdown and 2-step PCR. Only the touchdown PCR reactions supplemented with 2% DMSO and GC enhancer (NEB) yielded a PCR fragment of the correct length (11.5 kb). 2-step PCR did not yield any fragments of the correct length.*

4.3.4 Expression and purification of the double-tagged IdmO

Recombinant protein expression and IMAC purification were performed as described previously (section 4.2.2). The eluate from the IMAC was loaded onto a StrepTactin column (Iba-lifesciences), and the bound protein was eluted with 2.5 mM desthiobiotin. Samples from the purification procedure were analysed by reducing SDS-PAGE (Figure 4-15). As can be seen from the gel, the dual-affinity tag approach was successful at eliminating the ArnA contaminant from the purified IdmO sample. However, one major drawback of this approach was a very low yield. Most of IdmO failed to bind to StrepTactin beads during the application and eluted in the flow-through. As a result of this, the most concentrated

StrepTactin fraction had a concentration of 0.05 mg/ml IdmO (~200 nM). While this concentration of the sample was not sufficient for functional studies or crystallographic/cryo-electron microscopy experiments, it was, however, sufficient for negative-stain EM analysis.

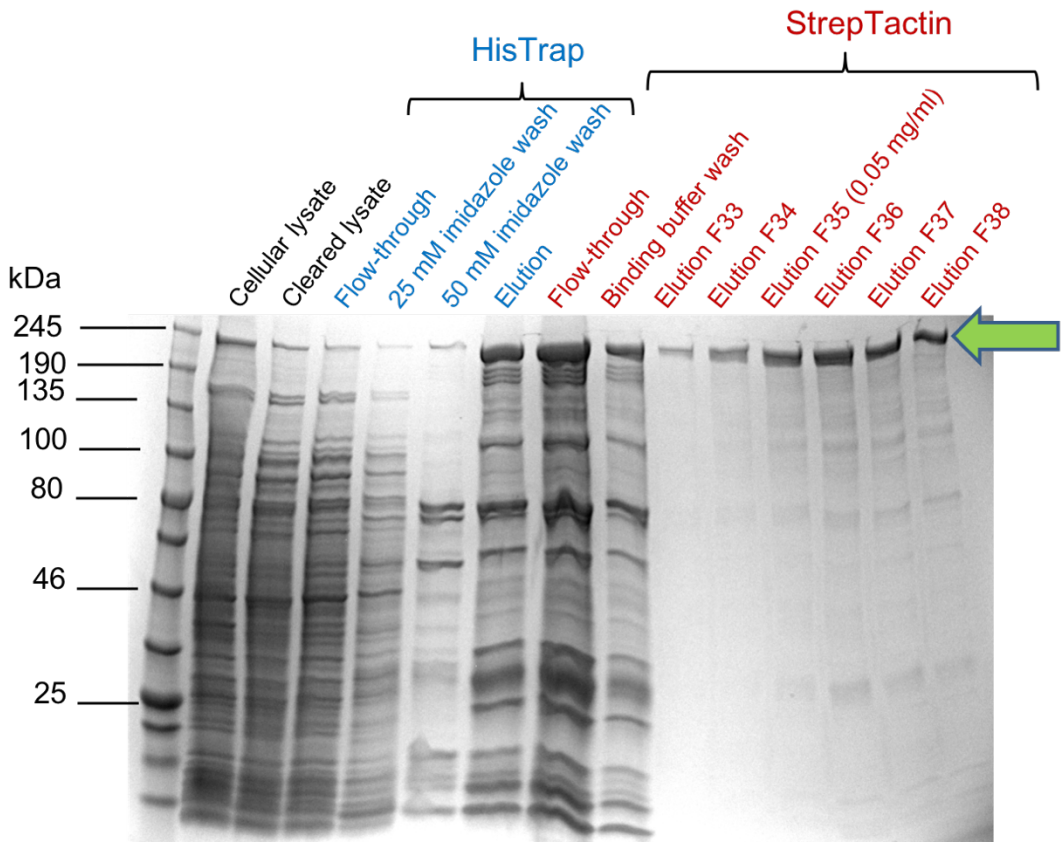


Figure 4-15 Purification of the double-tagged IdmO. *Samples from HisTrap are labelled blue, and StrepTactin chromatography samples are labelled red. The green arrow indicates IdmO. As can be seen from StrepTactin elution fractions, the dual-affinity tag approach was successful at eliminating the ArnA contaminant from the purified IdmO sample.*

4.4 Preliminary characterisation of IdmO

4.4.1 Peptide mass fingerprinting to confirm the completeness of IdmO

Accurate mass measurement by LC/MS is limited by the mass of the protein, and 228 kDa is too big for this approach to be feasible. To make sure the purified IdmO had no truncations and to determine whether the phosphopantetheinyl post-translational modification was present in the ACP domain (Nakano et al.,

1992; Quadri et al., 1998), a sample from StrepTactin chromatography elution, in the form of a 1D gel, was submitted for protein identification analysis by peptide mass fingerprinting (Webster and Oxley, 2012).

In-gel tryptic digestion, followed by LC-MS/MS and protein coverage analysis was performed which identified peptides covering both N- and C-termini (Figure 4-16). The overall coverage reached 73% (Table 4-3). Unfortunately, the conserved serine of the ACP (Ser2047) was not covered by any peptides, and it was impossible to determine whether the post-translational modification had occurred (although a non-specific phosphopantetheinylation of Ser1175 was detected).

Table 4-3 Peptide mass fingerprinting results of the double-tagged holo-IdmO band in the SDS-PAGE gel.

Protein Group	Accession	Coverage (%)	# Peptides	Avg. Mass	Description
1	Holo-IdmO	73	188	230278	Ieva 0701004B Holo-IdmO

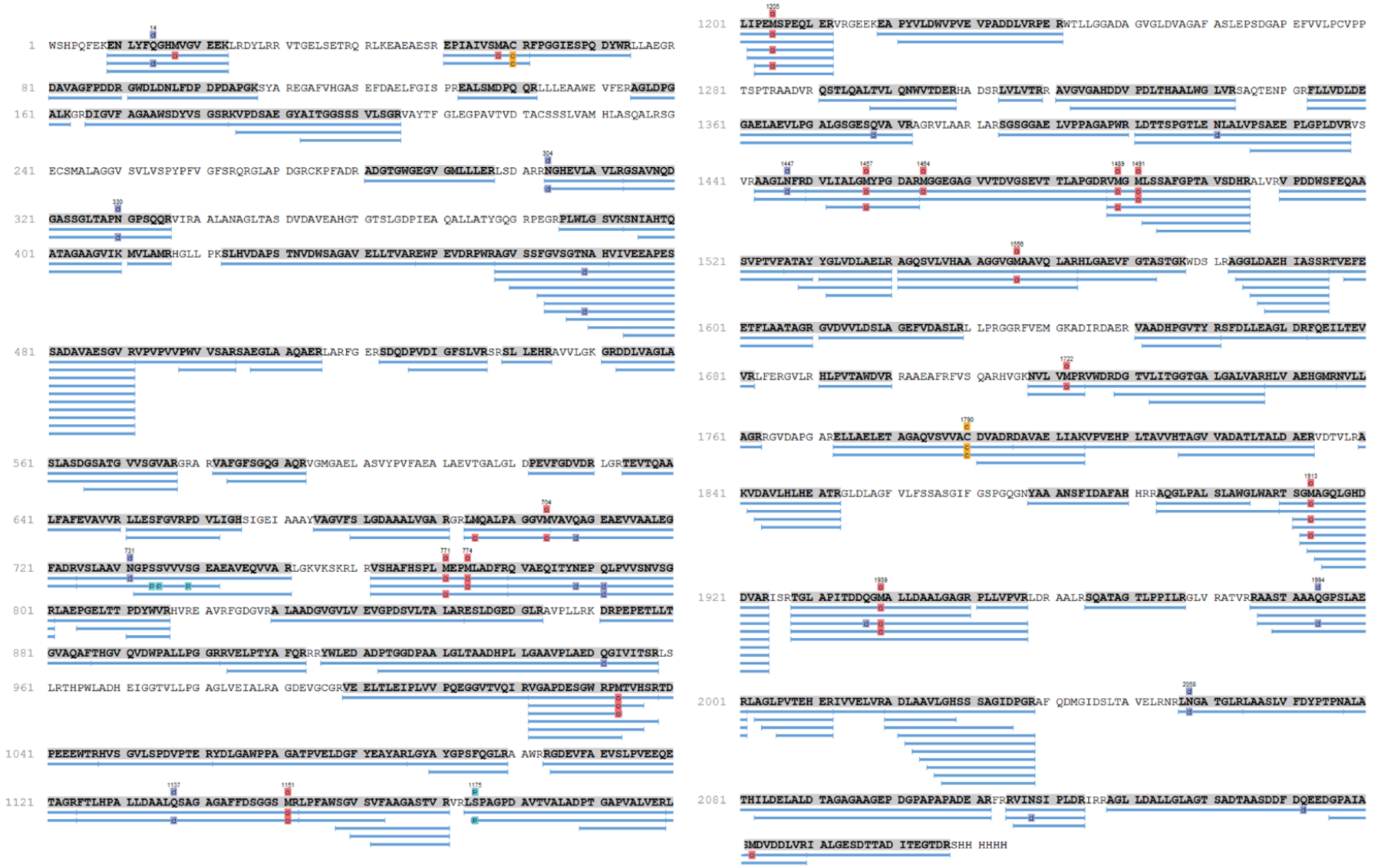


Figure 4-16 Peptide fingerprinting analysis of holo-IdmO. *Peptides identified are shown in blue. Post-translational modifications are shown as squares (orange – carbamidomethylation, blue – deamidation, red – oxidation, green – phosphopantetheinylation).*

4.4.2 Characterisation of IdmO by negative stain EM

Negative-stain EM was utilised to begin elucidating the structure of IdmO. StrepTactin elution fraction F35 (Figure 4-15) was desalted into buffer containing 10 mM Tris pH 7.4, 50 mM NaCl, and negative stain EM grids were made immediately using 1% uranyl acetate solution. A number of areas on the grid were imaged which revealed mono-dispersed particles with a length of approximately 20 nm (Figure 4-17a). Collection of 100 micrographs, picking of approximately 5000 individual particles and subsequent 2D class averaging (Figure 4-17b) and 3D refinement resulted in a low-resolution model of IdmO (Figure 4-17c), which revealed a dimeric U-shape architecture with a proposed reaction chamber located in the middle.

The low-resolution model of IdmO was compared to the 9.2 Å cryo-EM structure of *holo*-PikAIII (Figure 4-17d), the only other full-length PKS module characterised to date (Dutta et al., 2014). The comparison revealed that the two structures exhibited ample similarities in the overall architecture and the direction of the twist. What is more, PikAIII is four domains smaller which is reflected in the smaller size of this complex.

Since IdmO is a homo-dimer, C2 symmetry was applied during the reconstruction. To account for the effects of imposed symmetry, reconstruction with no symmetry (C1) was also calculated (Figure 4-17e). The overall architecture of the C1 model was observed to be highly similar to the C2 model, but of poorer quality, as expected without the constraints of symmetry and consistent for a very flexible molecule.

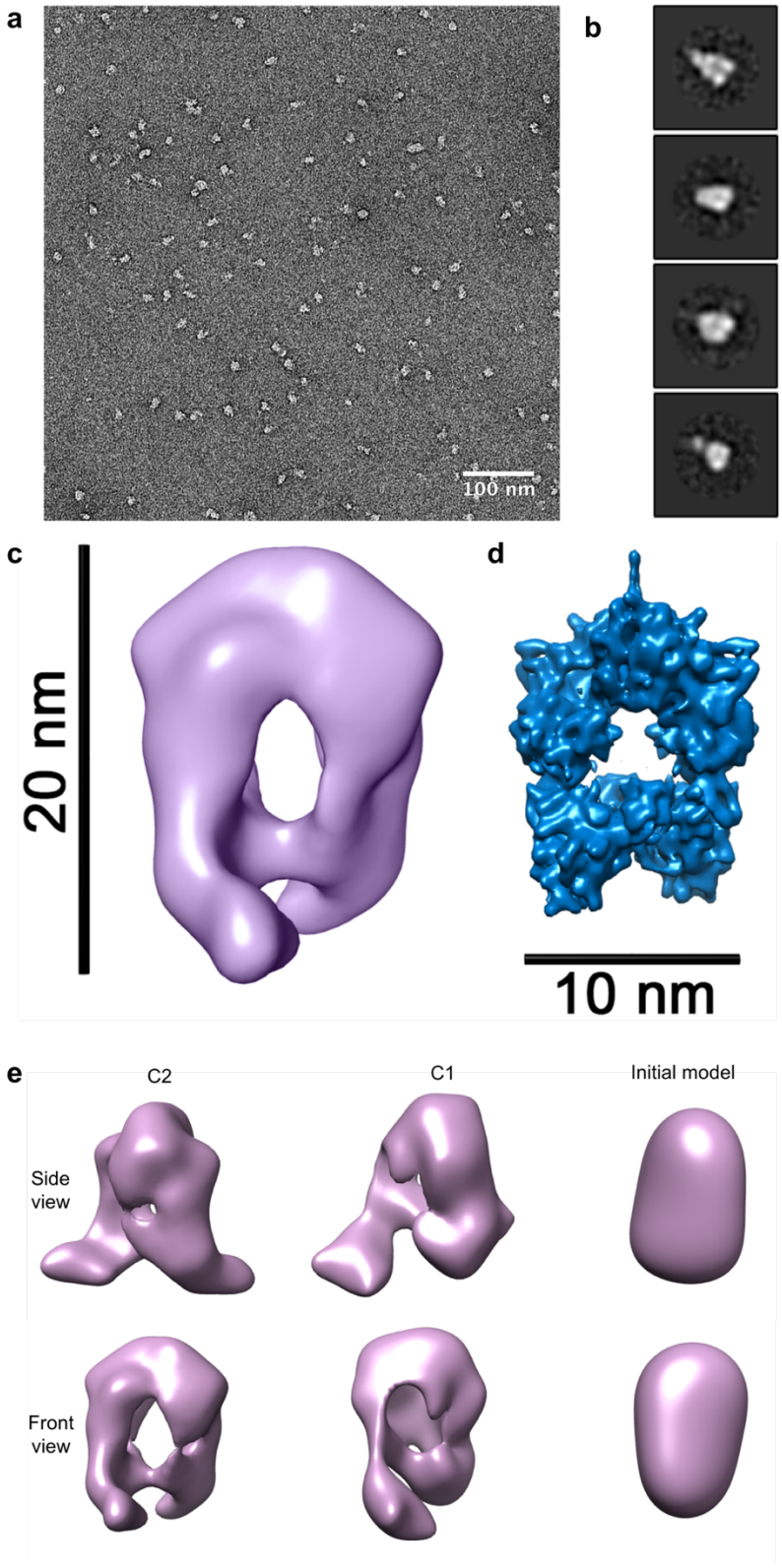


Figure 4-17 Negative stain EM studies of double-tagged IdmO. *(a) Raw micrograph with IdmO molecules embedded in 1% uranyl acetate stain. The image was recorded using a magnification of 30,000x and electron acceleration of 120 keV. Scale bar represents 100 nm. (b) 2D class averages of IdmO. (c) 3D reconstruction of IdmO (purple) compared to PikAIII cryo-EM map (EMDB entry 5647) (Dutta et al., 2014). (d) Comparison of IdmO reconstructions with (C2) and without (C1) symmetry as well as the starting model used in 3D refinement.*

4.5 Increasing the concentration of purified IdmO

The negative stain EM data were promising; therefore, the next logical step was to prepare an IdmO sample concentrated enough for analysis by cryo-EM. A rule of thumb is that negative stain EM samples need to be 10-100 times more concentrated for cryo-EM analysis to be feasible; therefore, the IdmO sample needs to have a concentration in the range of 0.5 - 5 mg/ml. The concentration of the peak StrepTactin fraction was 0.05 mg/ml, meaning, to increase the concentration of IdmO 10-100 times, the purification protocol had to be further optimised.

4.5.1 Attempts to optimise the StrepTactin chromatography step

Attempts to concentrate the StrepTactin elution fractions using centrifugal concentrator resulted in protein aggregation. Starting the StrepTactin purification with more material by performing protein expression in twelve litre culture volume and circulating the loading material over the column several times also did not result in increased amount of purified IdmO. It was hypothesised that the problem arose from IdmO molecules either being unable to bind to the StrepTactin column (due to the Strep II tag unavailability) or the bound IdmO molecules preventing the binding of more IdmO molecules to the neighbouring StrepTactin beads.

A third affinity chromatography step, in the form of small-scale IMAC spin column (His SpinTrap, GE), was briefly trialled. In this method, all StrepTactin elution fractions (an approximate total volume of 5 ml) were pooled together and bound to 100 µl of IMAC beads in a 1.5 ml Eppendorf spin tube. The bound material was eluted in 200 µl of elution buffer, concentrating the sample approximately 20 times. This resulted in the final IdmO concentration of 0.4 mg/ml (Figure 4-18a). Unfortunately, negative stain EM analysis revealed that this method of concentration significantly increased the amount of IdmO aggregates and therefore was not a suitable approach to derive material for structural studies by cryo-EM (Figure 4-18b).

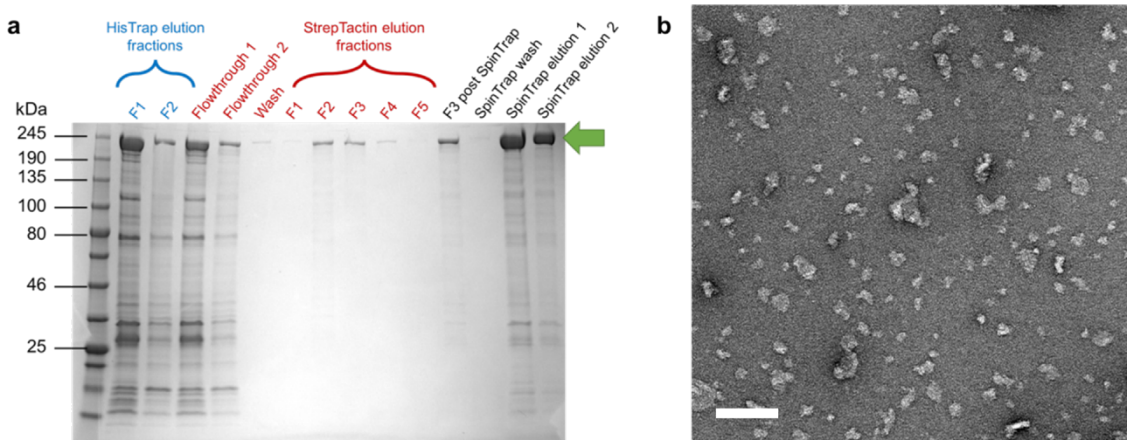


Figure 4-18 Concentrating IdmO by performing a third affinity chromatography step. *(a) Reducing SDS-PAGE analysis of samples from the three steps of affinity chromatography: HisTrap (IMAC), StrepTactin and His SpinTrap. The green arrow indicates IdmO. (b) A negative stain EM micrograph of desalted SpinTrap elution sample of IdmO. Large aggregates can be seen on the image. Scale bar represents 100 nm.*

4.5.2 Redesigning the construct to include a TEV protease cleavage site on the C-terminus

Another approach to get rid of contaminants without performing a second affinity or size-exclusion chromatography step is by cleaving the affinity tag and performing chromatography in reverse, i.e. binding the contaminants instead of the protein of interest to the column. In this approach, the first affinity chromatography step is performed as usual, and the affinity tag is cleaved off before the second chromatography step (identical to the first one) is performed to “catch” the unwanted contaminants and the cleaved affinity tag. The protein of interest does not have the affinity tag anymore and therefore elutes in the flow-through. Having decided on this approach, a TEV/Histag (3) construct had to be re-designed to contain a TEV protease cleavage site between the gene and the His tag on the C-terminus of IdmO (Figure 4-19).

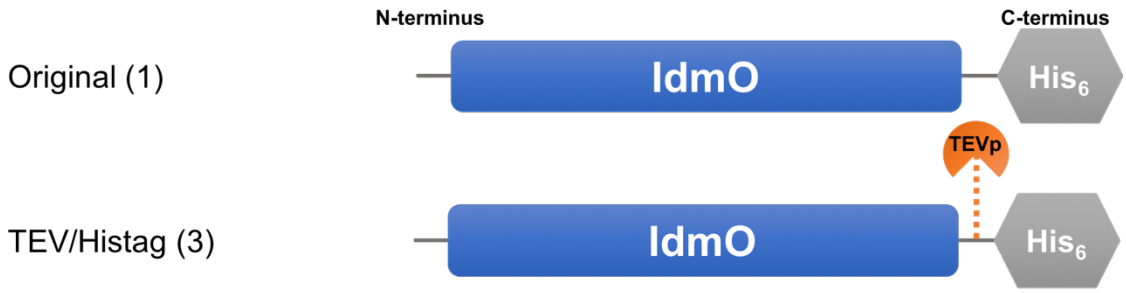


Figure 4-19 IdmO constructs used in this study 2. *(1) The original construct, created as described in section 4.2.1 harbouring a C-terminal His-tag. (3) TEV/Histag construct containing a C-terminal His tag, separated from the gene by a TEV protease cleavage site.*

4.5.2.1 Cloning of TEV/Histag (3) construct

To construct the TEV/Histag (3) IdmO vector, two DNA fragments were amplified using the original pET28a_IdmO vector as a template (Figure 4-19). The first fragment contained the linearised pET28 vector backbone (~5.2 kb) and the second fragment contained the *idmO* gene, which was PCR-amplified to contain a TEV cleavage site and overhangs to the vector backbone (~6.5 kb). Successful amplification of the pET28 backbone was achieved using a standard PCR cycling protocol, while the successful amplification of the gene insert was only achieved in the presence of the GC enhancer (Figure 4-20).

Both DNA fragments were mixed in a 1:1 ratio and incubated for an hour at 37 °C in the presence of *Dpn*I restriction enzyme before transformation into *E. coli* XL10 strain as described by the FastCloning procedure (C. Li et al., 2011). This cloning protocol yielded hundreds of colonies, twenty of which were screened by performing a double digest with *Pst*I and *Xba*I restriction enzymes and by DNA sequencing using T7 and T7term universal primers (Table 2-4). Three clones (termed pET28_IdmO_TEV16-TEV19) had the correct insert, and therefore pET28_IdmO_TEV16 was transformed (together with the PPTase gene harboring plasmid) into *E. coli* BL21(DE3) for protein expression.

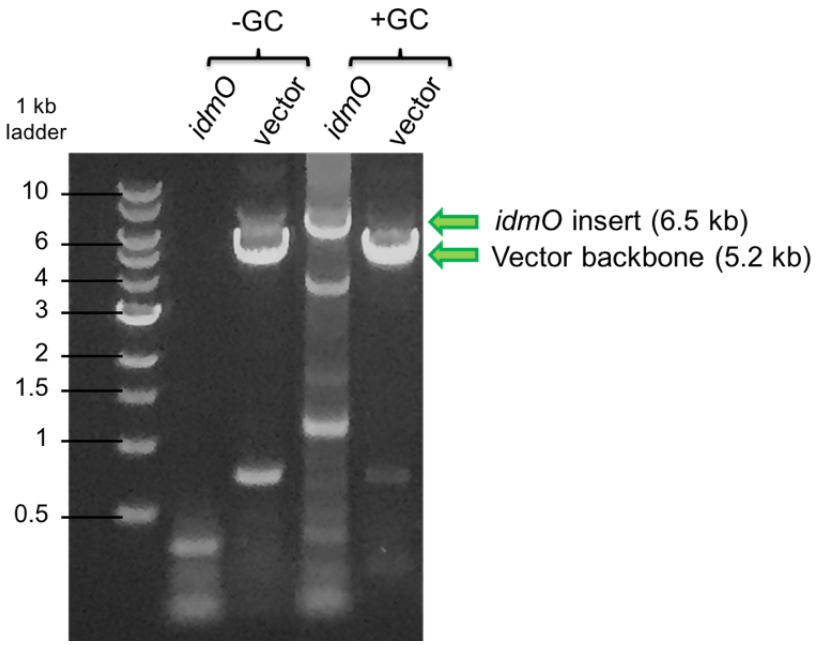


Figure 4-20 Amplification of (3) TEV/Histag IdmO vector fragments. *Fragments containing the vector backbone (5.2 kb) and gene insert together with the TEV protease cleavage site (6.5 kb) were amplified using an original pET28_IdmO vector as a template in the presence and absence of the GC enhancer. Both reactions yielded a vector backbone fragment of appropriate size whereas the insert was only amplified in the presence of the GC enhancer.*

4.5.2.2 Expression and purification of the TEV/Histag construct

Holo-IdmO was expressed as previously described (section 4.2.2). The purification protocol had to be adapted to include the desalting, proteolytic cleavage and size-exclusion steps as described in Figure 4-21a. Briefly, the cell harvest from six litres of *E. coli* culture was lysed, cleared and passed through an IMAC column (methods section 2.3.2.3). Eluted material was desalted into a buffer appropriate for TEV protease cleavage reaction (20 mM $NaPO_4$ pH 7.4, 0.1 M NaCl, 10% glycerol, 5 mM β-mercaptoethanol) and the His tag was cleaved at 4 °C overnight. The next day, the cleavage reaction was passed through a new IMAC column to “catch” the contaminants and the short His peptide before performing size-exclusion chromatography.

The IMAC “catch” step proved to be an efficient method for removing ArnA (Figure 4-21c). The size-exclusion chromatography step resulted in two peaks; the volume of the first peak was consistent with the void volume of the column while the volume of the second peak was consistent with a 456 kDa IdmO. It is worth noting that the shape of the IdmO peak was not completely symmetrical

and showed a “tail” indicating that it may comprise a heterogeneous population of protein particles. The most concentrated IdmO peak fraction (C7) had a concentration of 0.34 mg/ml, and all IdmO fractions combined contained a total of 2.5 mg of IdmO, making the overall yield of this IdmO purification procedure to be 0.42 mg recombinant IdmO per litre of *E. coli* BL21(DE3).

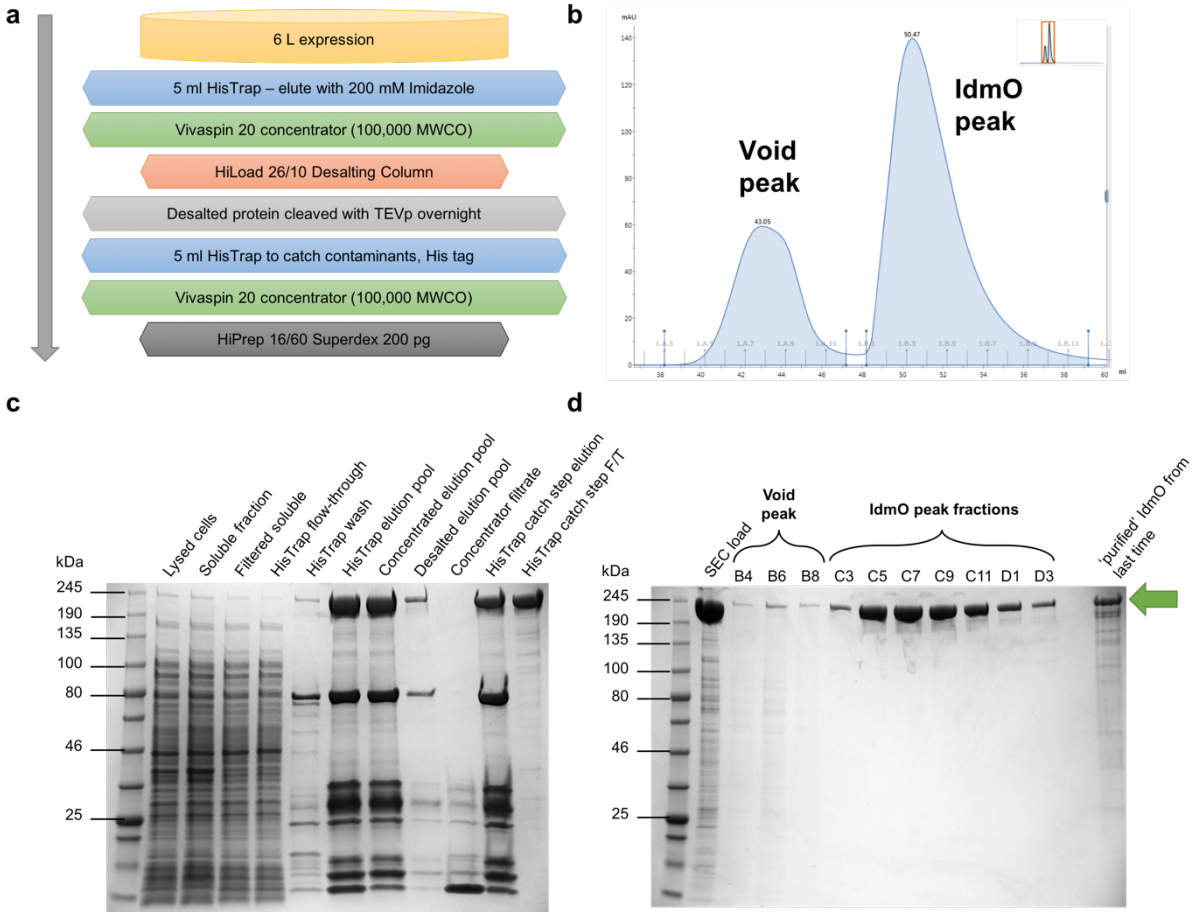


Figure 4-21 TEV/Histag IdmO construct purification. *(a) Purification method outline. Recombinant holo-IdmO was expressed in large-scale (12 x 0.5 litre culture volume), and first purification step was performed on a 5 ml IMAC column. The eluate was then desalted to get rid of most NaCl and imidazole, and the proteolytic cleavage of the His tag was performed at 4 °C overnight using TEV protease. The next morning most IMAC contaminants together with the His tag peptide were separated from the IdmO using a new IMAC column. The flow-through of this step was then further purified by a size-exclusion column. (b) Size-exclusion chromatography curve. The first peak contains aggregates and the second peak contains dimeric IdmO. (c) and (d) Reducing SDS-PAGE gels of samples taken throughout the purification. The green arrow indicates recombinant holo-IdmO.*

4.5.2.3 Characterisation of IdmO by size-exclusion chromatography multi-angle laser light scattering

Previous studies have reported PKSs to associate into homo-dimers, and therefore the expected mass of the IdmO dimer was calculated to be 456 kDa. Size-exclusion chromatography (Figure 4-21b) revealed an asymmetrical shape peak for IdmO, and therefore, it was decided to perform size-exclusion chromatography multi-angle laser light scattering (SEC-MALLS) analysis to gain insight into how homogenous the IdmO sample is.

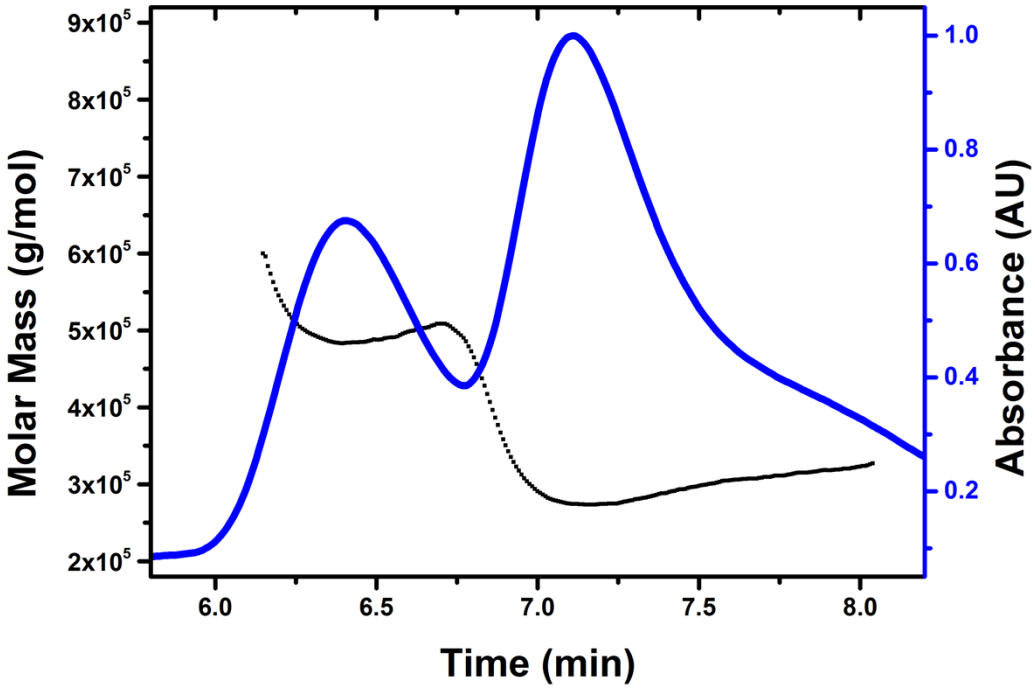


Figure 4-22 SEC-MALLS analysis of IdmO. *The dotted line represents the molar mass while the blue line represents the absorbance at 280 nm. The analysis was performed using a WYATT WTC silica column at a flow rate of 1 ml/min. The first peak was calculated to have an approximate molar mass of 480 kDa and the second peak – 275 kDa.*

SEC-MALLS analysis identified the presence of two clear peaks for the IdmO sample (Figure 4-22). The first peak molar mass was calculated to be approximately 480 kDa, consistent with a dimeric IdmO. The second peak molar mass was calculated to be 275 kDa, consistent with monomeric IdmO. When the same SEC-MALLS analysis was repeated with an analytical grade Superdex 200 column, only one peak was observed making it unsurprising that both dimeric and monomeric species of IdmO were eluting as one peak from the prep grade

Superdex 200 size-exclusion chromatography columns. This can be explained in terms of the shape of the molecule. It is likely that both dimeric and monomeric IdmO have a similar hydrodynamic radius and therefore travel down the size-exclusion chromatography column at a similar speed. The silica stationary phase used during the SEC-MALLS analysis offers a lot better resolution, which explains the presence of both dimeric and monomeric IdmO peaks observed during this analysis.

4.5.2.4 Negative stain EM of IdmO derived using a new purification protocol

To test whether IdmO, purified using the new protocol, still forms an arch-like architecture and to improve the resolution of the first reconstruction of IdmO (Figure 4-17), negative stain EM analysis was repeated. The method was identical to the one described in the section 4.4.2, except for the EM instrument used. This time a Tecnai TF20 instrument (instead of the Tecnai T12) was used and the images were recorded on a more advanced camera (FEI Ceta, 4k x 4x CMOS instead of Gatan US4000/SP 4k x 4k CCD camera).

A total of 100 micrographs were recorded at a magnification of 29,000, and approximately 30,000 particles were automatically picked. Four rounds of 2D class averaging were performed which yielded classes shown in Figure 4-23.

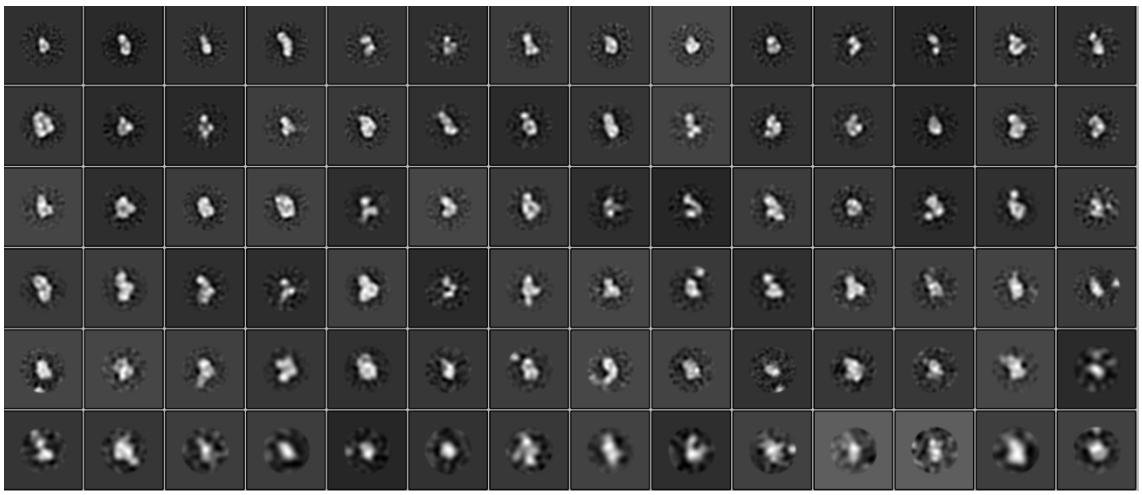


Figure 4-23 2D class averages of IdmO. *Parameters used in this data processing were as follows: 128-pixel box size (which equates to 450 Å at a pixel size of 3.51 Å/pix) and 250 Å diameter mask.*

Variability in the size and shape of IdmO particles was observed in the 2D classes with some classes (first row) resembling a monomer and some classes appearing a lot bigger and resembling what could be an IdmO dimer. When these particles were used to calculate a 3D reconstruction, the resulting map exhibited a lot more disorder (data not shown). The reasons for this could be a different IdmO purification procedure (a two-day method involving a cleavage step by TEV protease as opposed to a one-day method involving two affinity chromatography steps performed after one another) or a superior camera capable of resolving the heterogeneity of the sample at a better detail.

What is also worth mentioning, the results observed here may not be entirely representative of the sample if the uranyl acetate stain is too thick. If that is the case, a part of the molecule might get buried under the layer of stain, and therefore the resulting particles might appear smaller than they are leading to incorrect 2D class averages.

Nevertheless, these data have shown that IdmO can be purified to homogeneity and at sufficient quantities for structural studies. Next, we wanted to know whether this heterologously-produced IdmO is catalytically active.

4.6 Acyltransferase activity assay

4.6.1 Principle

To determine whether the purified recombinant *holo*-IdmO was active, a continuous, coupled acyltransferase activity assay (Figure 4-24) was adapted from previously published methods (Dunn et al., 2013; Molnos et al., 2003; McMinn and Ottaway, 1975; Kim et al., 2000).

During its catalytic cycle, the AT domain binds the extender unit (in this case malonyl-CoA) forming an AT-malonate intermediate and releasing free coenzyme A (CoASH) (Figure 1-8). Malonate is then transferred to a phosphopantetheinyl prosthetic group of ACP_9 via the nucleophilic attack and is ready for the chain elongation reaction catalysed by KS. The AT can then turnover another malonyl-CoA molecule; however, without the rest of the condensation reaction components, each IdmO dimer can only turnover two malonyl-CoA molecules. Therefore, to regenerate IdmO AT domain during a continuous assay, standalone *holo*-ACP_9 has to be added in excess.

In this assay, the release of CoASH is coupled to the formation of reduced nicotinamide adenine dinucleotide (NADH) in the α-ketoglutarate dehydrogenase (α-KGDH) assay. αKGDH, in the presence of thiamine pyrophosphate (TPP), converts CoASH, α-ketoglutarate and NAD+ into succinyl-CoA, CO_2 and NADH, the formation of which can be measured fluorometrically.

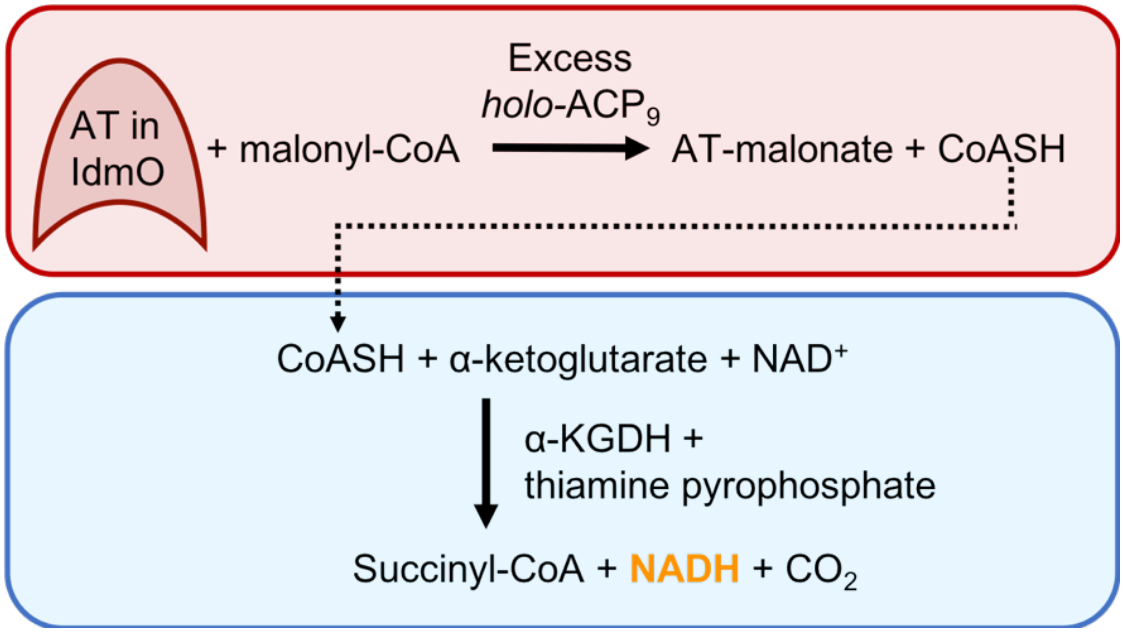


Figure 4-24 IdmO activity assay. *The box in red outlines the reaction catalysed by IdmO AT domain whereby malonyl-CoA binds the acyltransferase (AT) releasing a CoA molecule. The box in blue outlines coupling of AT catalysis to αKGDH activity. This coupled reaction results in the production of NADH which can be detected fluorometrically.*

4.6.2 Cloning, expression and purification of ACP_9

4.6.2.1 Summary of the initial attempts to produce recombinant ACP_9

As mentioned previously, an excess of *holo*-ACP_9 needs to be added to the reaction to regenerate the AT domain. First, the wild-type ACP_9 was cloned from the genomic DNA of *S. antibioticus*; however, the expression screens revealed no soluble expression. Codon usage analysis indicated that ~20% of ACP_9 codons were used by *E. coli* less than 10% of the time indicating a possibility for poor gene expression. To overcome this, a codon-optimised ACP_9 gene was synthesised. Unfortunately, the codon-optimised gene also resulted in very little soluble expression. We hypothesised that since the protein was very small (9 kDa), the protein translation was happening too fast for the protein folding machinery to operate correctly, leading to the formation of inclusion bodies.

4.6.2.2 Cloning of the pET_SUMO_ACP$_9$ fusion construct

Some affinity tags are known to improve the expression levels and solubility of their fusion partners (Marblestone et al., 2006). We decided to attach a SUMO tag to the ACP$_9$ in an attempt to improve the yield of recombinant *holo*-ACP$_9$. SUMO tag was PCR-amplified from Champion™ pET_SUMO vector (a kind gift from Dr Glyn Hemsworth, University of Leeds, UK).

The ACP$_9$ gene insert, together with the DNA overhangs to the vector, was amplified from the codon-optimised ACP$_9$ construct (Figure 4-25). Vector backbone amplification was successfully achieved as exemplified by a single band in the agarose gel. The amplification of the insert resulted in two bands on the agarose gel with the top band appearing to be of expected, ~300 bp, size and the bottom band of approximately 100 bp comprising non-specific amplification product.

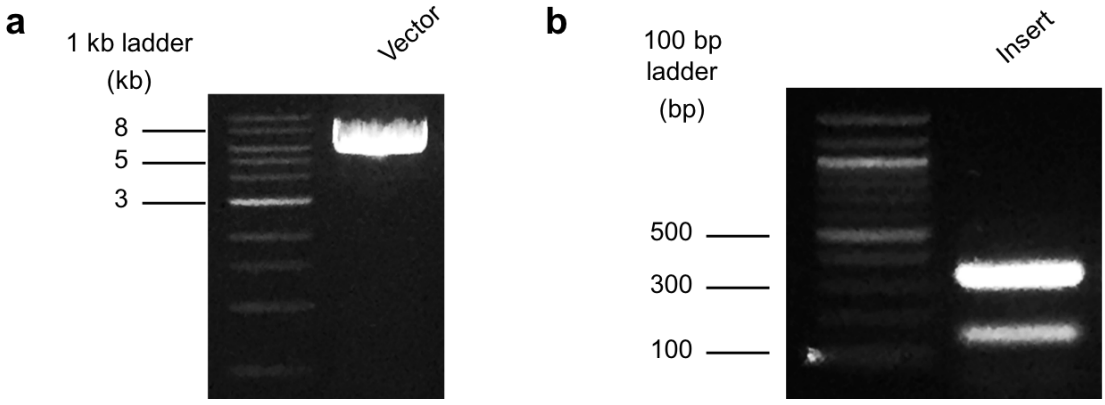


Figure 4-25 Amplification of pET_SUMO vector backbone and ACP$_9$ gene insert. *(a) pET_SUMO vector backbone with an expected size of 5.6 kb. (b) The ACP$_9$ gene insert with an expected size of 300 bp. Amplification of the gene insert resulted in an additional DNA fragment of approximately 100 bp.*

Despite the non-specific amplification, both the vector and the insert fragments were mixed in 1:1 ratio and incubated for an hour at 37 °C in the presence of *Dpn*I restriction enzyme before transformation into *E. coli* XL10 strain as described by the FastCloning procedure (C. Li et al., 2011). This cloning protocol yielded hundreds of colonies, ten of which were screened by performing a double digest with *Hind*III and *Xba*I restriction enzymes and by DNA sequencing using T7 and T7term universal primers (Table 2-4). Six clones (termed pET_SUMO_ACP2, 3, 6, 7, 8 and 9) had the correct size insert, and therefore pET_SUMO_ACP2 was

transformed (together with the PPTase gene harboring plasmid) into *E. coli* BL21(DE3) for protein expression.

4.6.2.3 Expression and purification of *holo*-ACP_9

Holo-ACP_9 was expressed as previously described (section 4.2.2) except for the growth medium (LB rather than 2TY). The purification protocol was very similar to the one used to purify TEV/Histag *holo*-IdmO (section 4.5.2.2). Briefly, the cell harvest from four litres of *E. coli* culture was lysed, cleared and passed through an IMAC column (methods section 2.3.2.3). Eluted material was desalted into a buffer appropriate for SUMO protease cleavage reaction (50 mM Tris pH 8.0, 0.2 M NaCl, 5 mM DTT, 5 mM $MgCl_2$) and the SUMO tag was cleaved at 4 °C for 90 hours. Finally, the cleavage reaction products were purified according to their size by size-exclusion chromatography.

The size-exclusion chromatography resulted in four peaks. The volume of the first peak was consistent with the void volume of the column and most likely contained protein aggregates. The second peak was thought to contain uncleaved SUMO-ACP_9. To investigate the composition of peaks three and four, LC/MS analysis was performed (Figure 4-27, Figure 4-28 and Figure 4-29).

LC/MS analysis revealed that the third peak contained varying lengths of cleaved SUMO tag and the fourth peak contained mostly *holo*-ACP_9. The fraction in between the two peaks contained very little *holo*-ACP_9 and therefore was not used for the assay.

All *holo*-ACP_9 peak fractions combined contained a total of 16.5 mg of recombinant protein, making the overall yield of this preparation to be 4.125 mg recombinant ACP_9 per litre of *E. coli* BL21(DE3) culture.

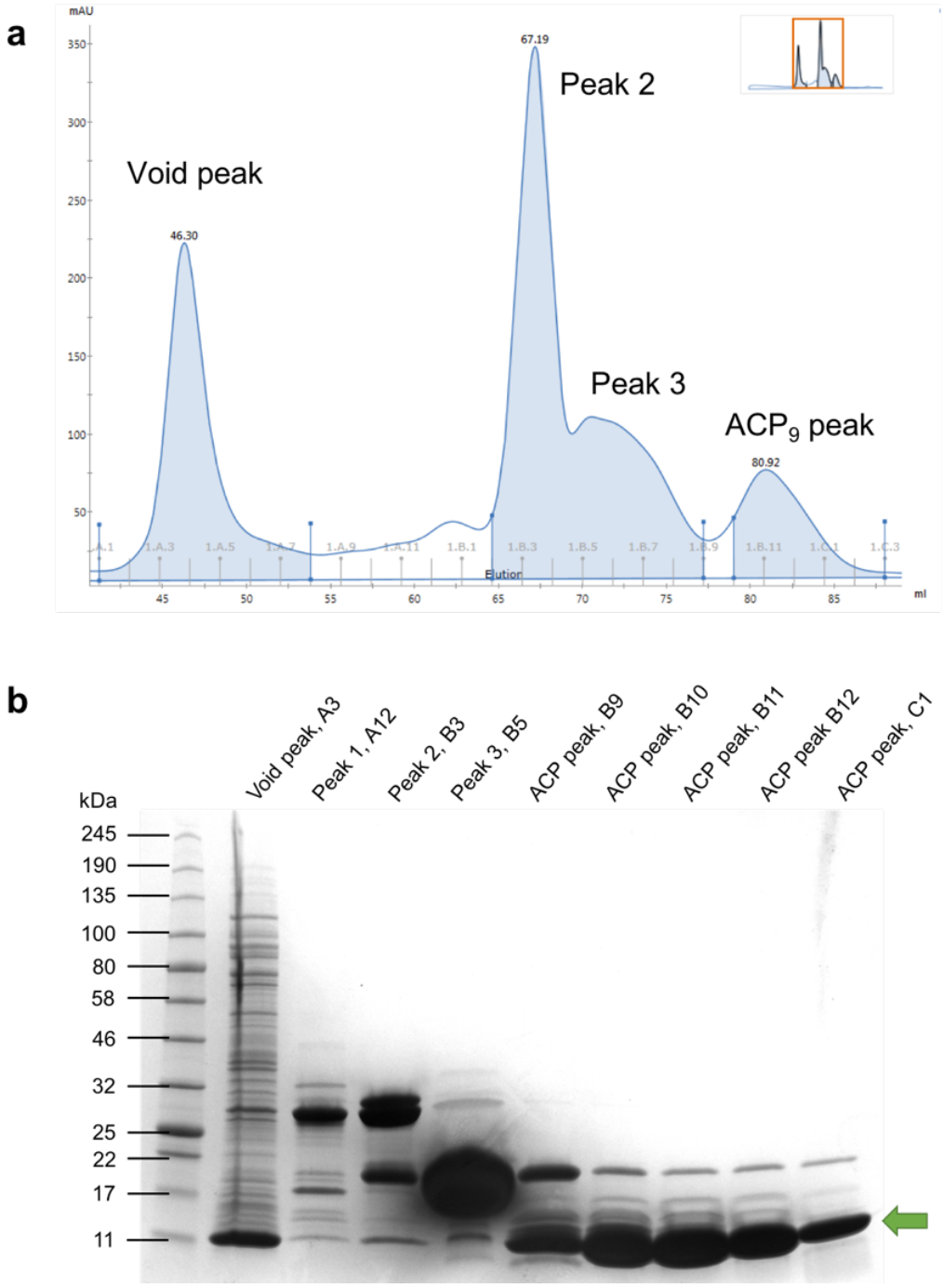


Figure 4-26 SUMO-ACP$_9$ purification by size-exclusion chromatography. *(a) Size-exclusion chromatography curve. The first peak contains aggregates, peaks two and three contain uncleaved SUMO-ACP$_9$ and SUMO tag while the last peak contains ACP$_9$. (b) Reducing SDS-PAGE gel of size-exclusion chromatography fractions. The green arrow indicates recombinant ACP$_9$.*

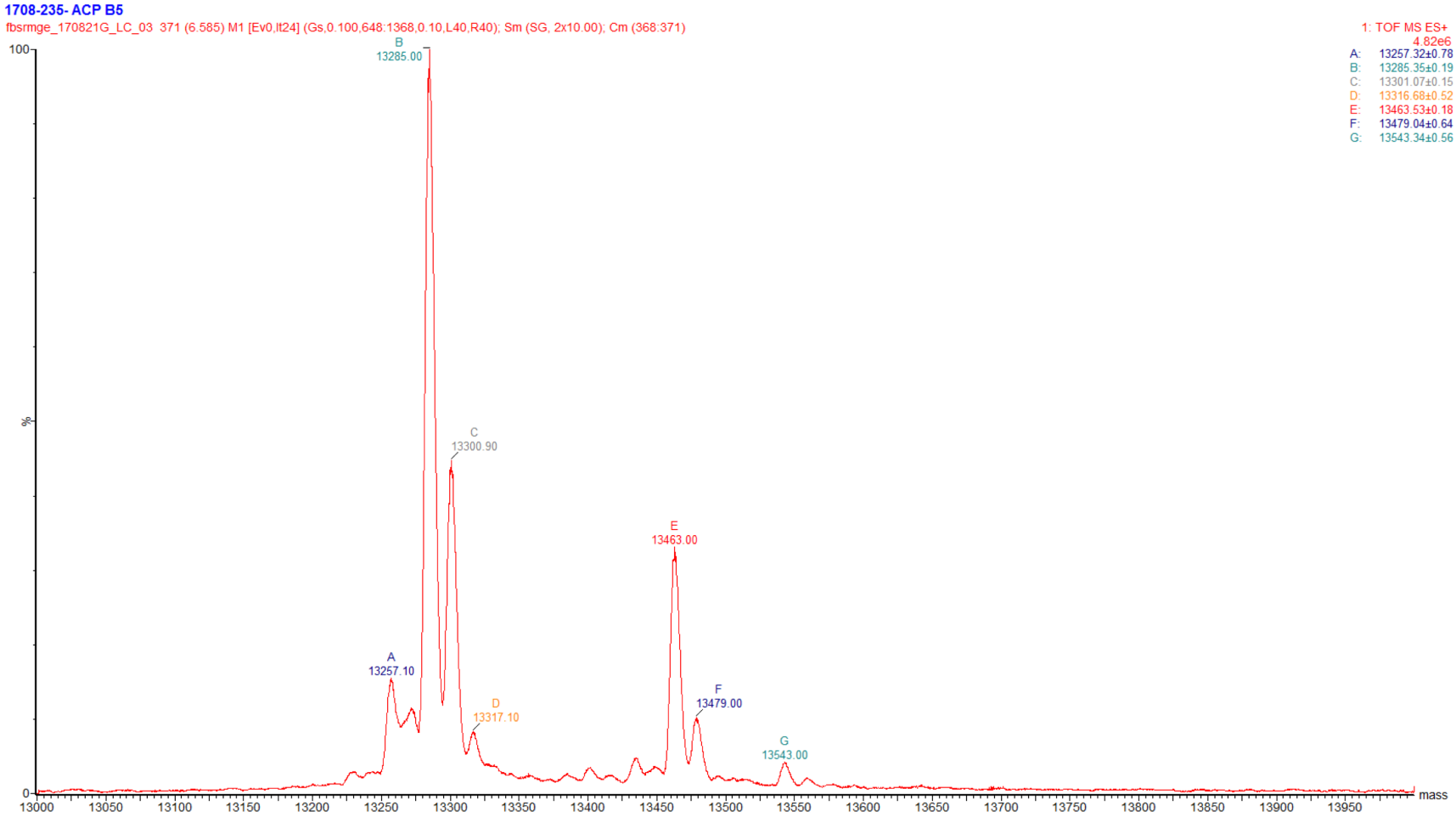


Figure 4-27 Analysis of the third holo-ACP$_9$ size-exclusion elution peak contents by LC/MS. *Here, a deconvoluted spectrum is shown for a number of proteins with measured masses between 13,257 and 13,543. Calculated mass for the SUMO tag is 13,285 indicating that this size-exclusion peak contained SUMO tag of various lengths.*

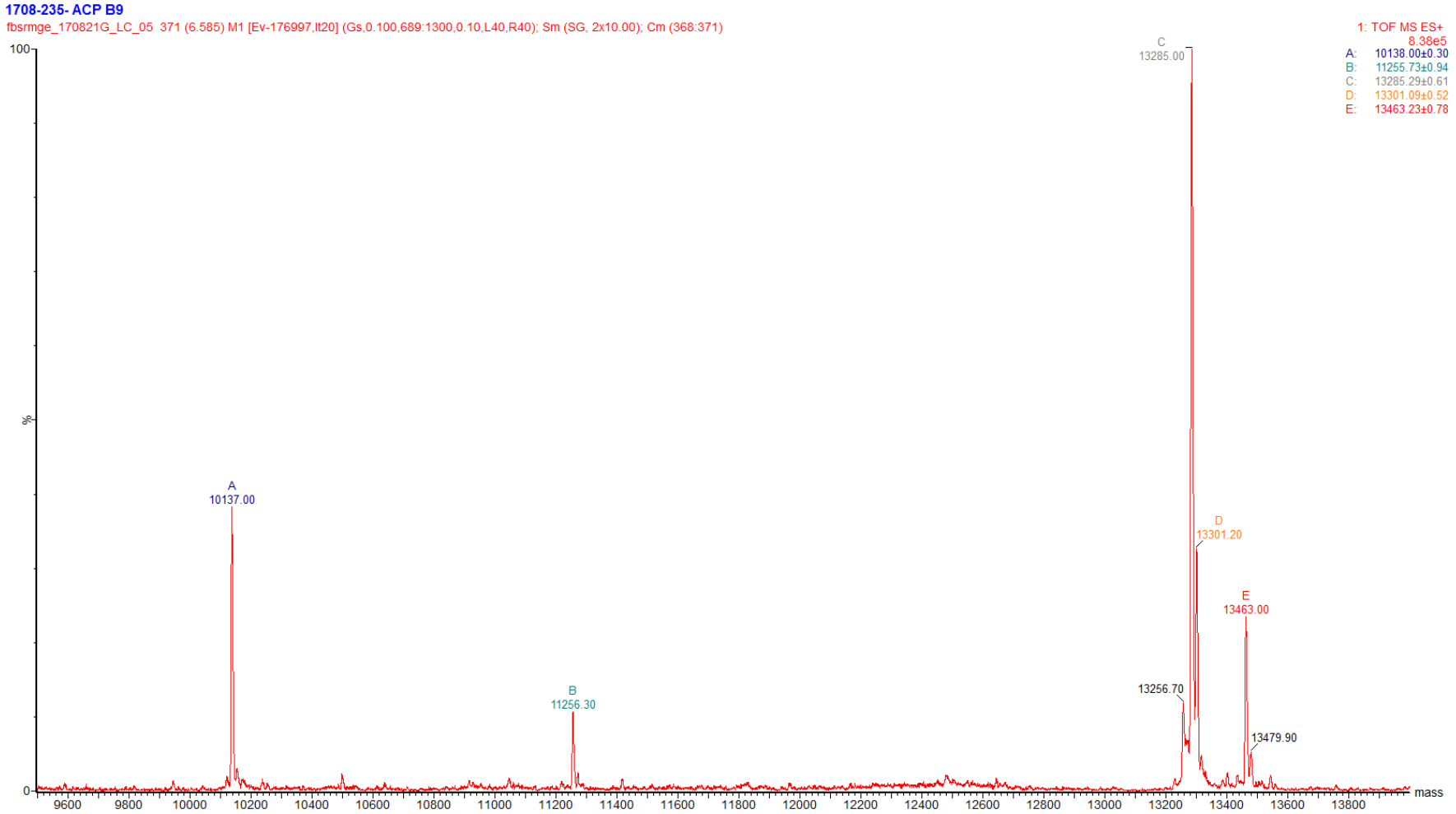


Figure 4-28 Analysis of the holo-ACP$_9$ fraction B9 (located in-between the third and fourth chromatography peaks) by LC/MS. *Here, a deconvoluted spectrum is shown with a number of proteins with measured masses between 10,138 and 13,463. The mass of the second peak (11,256.30) matches a predicted mass of holo-ACP$_9$ (11,256) indicating that holo-ACP$_9$ represents a very small part of this fraction with the rest being unwanted contaminants.*

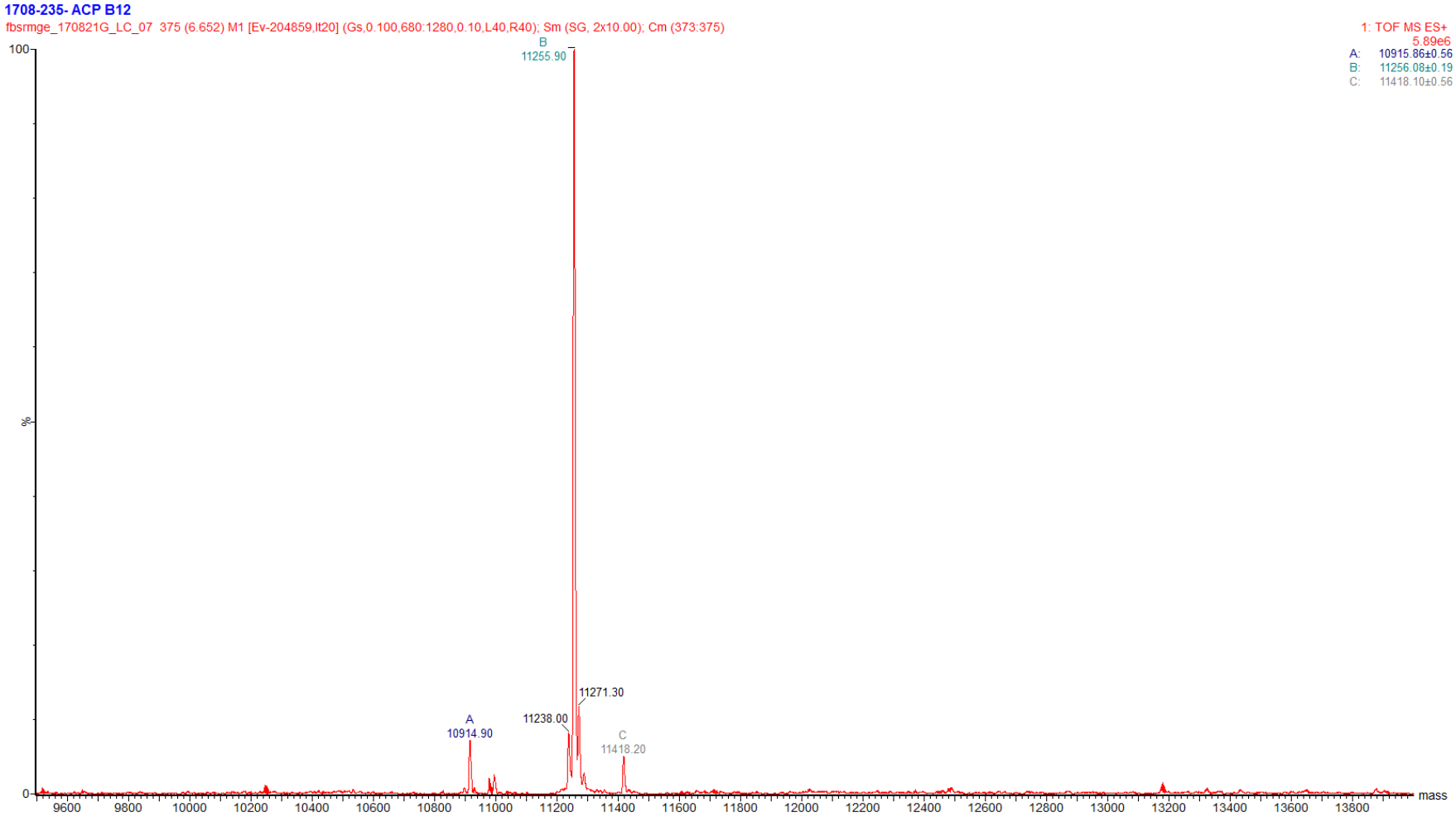


Figure 4-29 Analysis of the holo-ACP$_9$ peak fraction pool by LC/MS. *Here, a deconvoluted spectrum is shown with two peaks with measured masses of 10,914.90 and 11,255.90. The mass of the small peak matches a predicted mass of apo-ACP$_9$ (10,915) while the mass of the big peak matches a predicted mass of holo-ACP$_9$ (11,256) indicating that the ACP$_9$ was successfully phosphopantetheinylated and that the holo-ACP$_9$ represents the majority of this sample.*

4.6.3 Coupling of the AT catalysis to αKGDH activity

Even though the assay conditions were adapted from a method described previously (Dunn et al., 2013), initial attempts to couple the AT catalysis to αKGDH activity were fruitless. A lot of troubleshooting and optimisation took place to develop a method described below.

Specifically, assay components were prepared in seven different solutions and combined in three stages as outlined in Table 4-4. First, all ingredients highlighted in yellow (αKGDH, NAD+, TPP, α-ketoglutaric acid and ACP_9) were combined and incubated at 30 °C. Care needs to be taken at this step to make sure the pH of this solution is between 7.0 and 8.0. α-ketoglutaric acid is very acidic and therefore without an appropriate pH adjustment step will strongly influence the pH of the whole mixture of assay components rendering αKGDH inactive (working pH range of αKGDH is between pH 7.0 and 8.0). Second, the substrate (malonyl-CoA in this case) was added, and the reaction mixture was left to incubate at 30 °C for five minutes to allow any free CoA from the substrate to be turned over. Finally, the coupled reaction was initiated by the addition of IdmO and fluorescence emission was monitored for five minutes using 360 nm excitation and 460 nm emission filters.

Black polystyrene, flat bottom 96-well plates were used for all experiments and reactions were run in parallel with NADH standards and no enzyme controls. Sensitivity (also termed gain) parameters were tested and set on the day of the experiment. A buffer used for making up solutions was prepared daily and contained 20 mM Na_2PO_4 pH 7.4, 300 mM NaCl, 10% glycerol, 2 mM DTT and 1 mM EDTA.

Table 4-4 AT assay components. *Abbreviations: αKGDH, α-ketoglutarate dehydrogenase; NAD+, nicotinamide adenine dinucleotide; TPP, thiamine pyrophosphate. Ingredients were added together in three stages: yellow, blue and finally green.*

Solution	Ingredient	Stock conc.	Assay conc.	Volume for a 100 µl reaction
1	αKGDH	10x (400 mU)	40 µM	10 µl
2	NAD^+	40x (16 mM)	0.4 mM	2.5 µl
3	TPP	40x (16 mM)	0.4 mM	2.5 µl
4	α-ketoglutaric acid	20x (40 mM)	2 mM	5 µl
5	ACP_9	2.5x (250 µM)	100 µM	40 µl
6	Malonyl-CoA	10x (100 µM)	10 µM	10 µl
7	IdmO (AT)	5x (250 nM)	50 nM	20 µl

One final aspect to consider is the quality of αKGDH. Three different lots of αKGDH (Sigma) were used in this study, one of which turned out to be completely inactive while the other two displayed comparable activity. Therefore, after purchasing a new aliquot of αKGDH, it is advisable to test the αKGDH activity by setting up only the second part of the coupled reaction using free Coenzyme A substrate (blue part of Figure 4-24).

When all reaction components were assembled in the order described above, an increase in fluorescence was observed (Figure 4-30). This verified that IdmO is indeed active and capable of turning over malonyl-CoA molecules when expressed recombinantly. All relevant controls were also run in parallel. Reactions with no substrate, no nicotinamide adenine dinucleotide and no αKGDH resulted in no fluorescence produced while a reaction with no IdmO exhibited some fluorescence, presumably due to the leftover CoA in the malonyl-CoA solution.

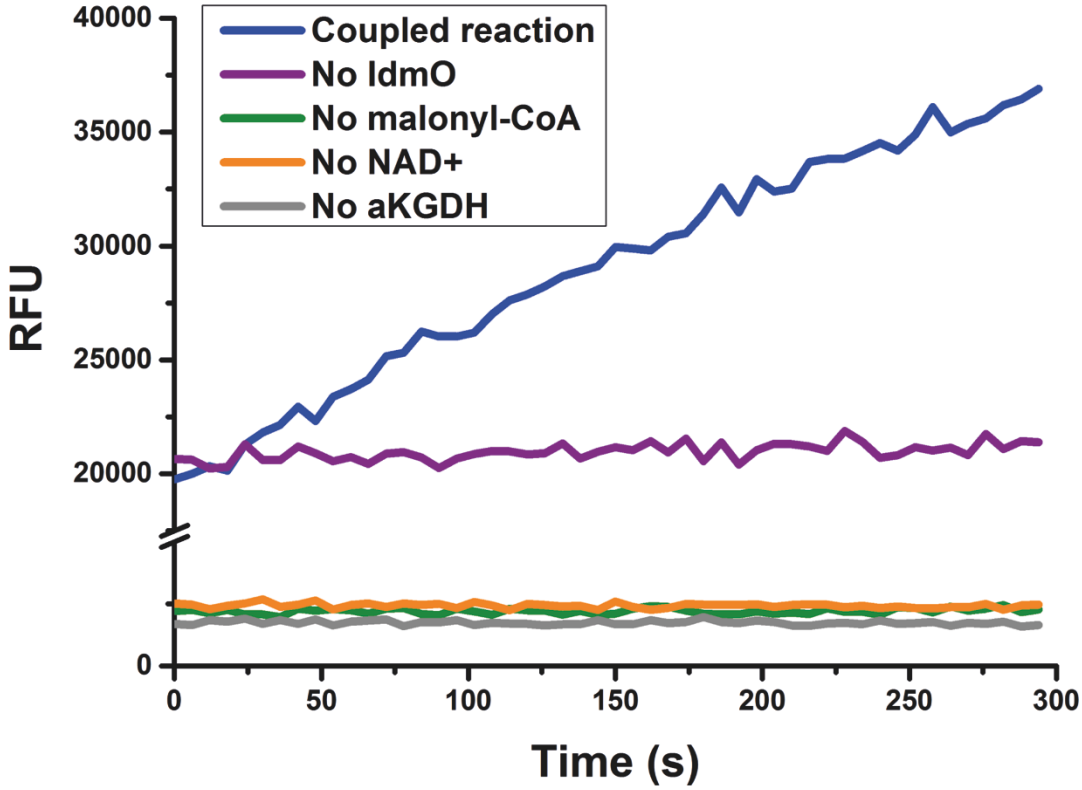


Figure 4-30 Coupling of AT catalysis to αKGDH activity. *Progress fluorescence curves can be seen for the coupled reaction and relevant controls. RFU: Relative Fluorescence Units.*

4.6.4 Preliminary kinetic characterisation

To assess the efficiency of the IdmO AT domain and to compare it against standalone AT domains studied before (Dunn et al., 2013), preliminary kinetic analysis of IdmO was performed.

Kinetic measurements need to be taken at a linear part of the progress curve where the enzyme of the study is limiting. Therefore, an optimum working concentration of IdmO was determined for 25 μM malonyl-CoA concentration (while other reaction components were kept as described in Table 4-4).

To do this, five coupled reactions were set up using 12.5 nM, 25 nM, 50 nM, 100 nM and 200 nM concentrations of IdmO and the total amount of fluorescence emitted in five minutes was calculated (Figure 4-31). As can be seen from Figure 4-31, the concentration of NADH doubles from 12.5 to 25 nM IdmO, but it does not increase proportionally for 50, 100 and 200 nM IdmO. Therefore, it was concluded that the optimum concentration of IdmO to use for the kinetic

measurements (when using up to 25 µM substrate concentration) is less than 25 nM.

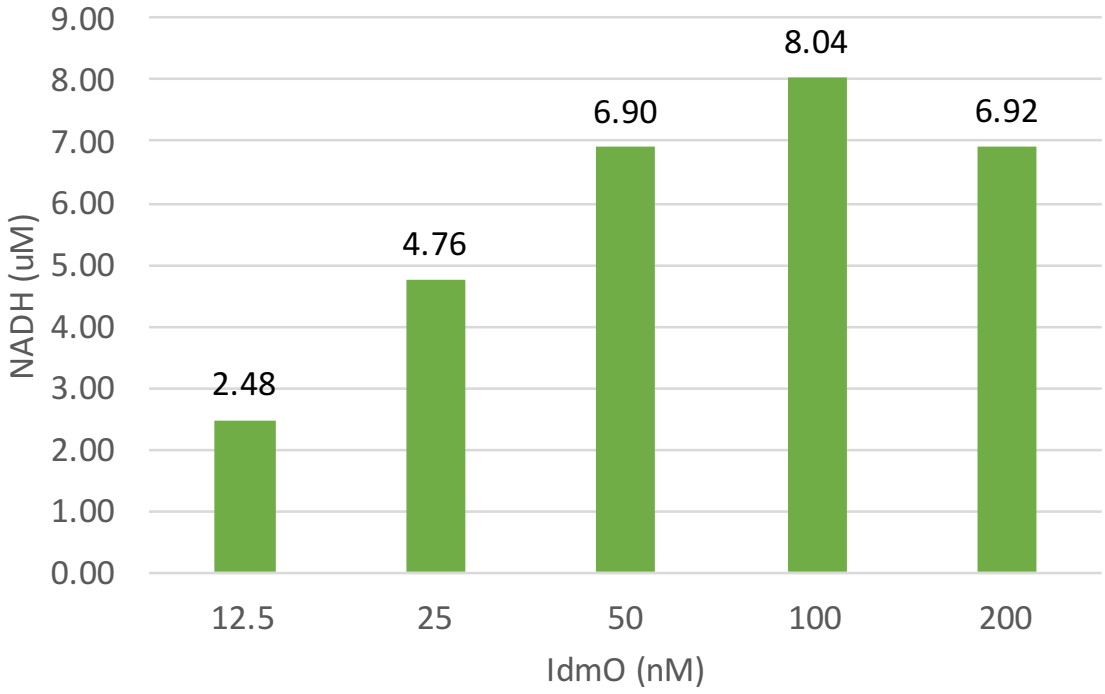


Figure 4-31 Determining the optimum concentration of IdmO to use during the kinetic measurements. *Coupled reactions were set up with varying concentrations of IdmO (12.5 to 200 µM) and the total increase in NADH concentration in five minutes was measured.*

To measure the saturation kinetics, the reactions were run with 50 nM IdmO in the presence of 100 µM *holo*-ACP$_9$, with malonyl-CoA concentrations of zero to 12.5 µM. Initial rates of the no-IdmO control reactions were subtracted from their enzyme-containing counterparts. To define kinetic parameters (Figure 4-32), initial velocity vs concentration curves were fit to the Michaelis-Menten equation using the curve fitting tool in OriginPro software (OriginLab©).

The observed K_M of IdmO was 3.7 ± 0.6 µM, which is comparable to the values reported for DEBS AT domains (2-7 µM) (Dunn et al., 2013). Interestingly, the k_{cat}/K_M rate of malonyl-CoA transacylation was approximately 500-fold higher than that of DEBS AT3-mediated transacylation of its native substrate, methylmalonyl-CoA (Dunn et al., 2013). Since DEBS AT3 domain was expressed and studied as a standalone protein, these results suggest that intramodule interactions within IdmO may play a crucial role in speeding up the reaction. However, more in-depth kinetic analysis and repeats in triplicate need to be

obtained before further interpretation of these kinetic parameters can be performed.

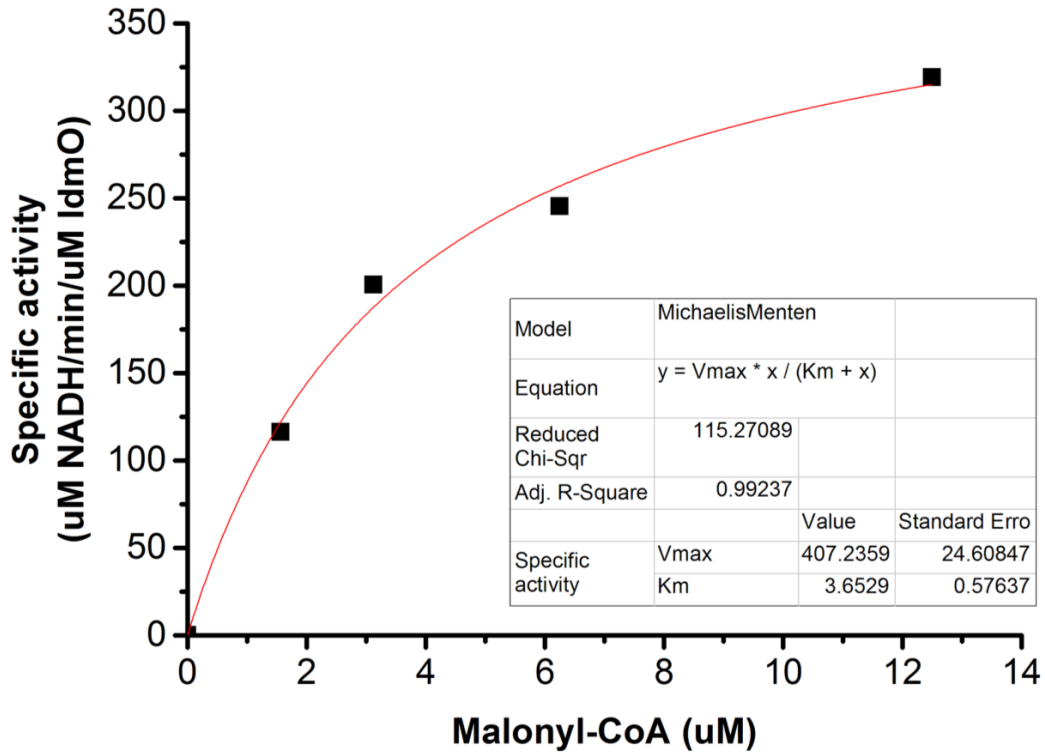


Figure 4-32 Hydrolysis of malonyl-CoA by IdmO AT domain. *Kinetic parameters were measured by varying malonyl-CoA concentrations in the presence of holo-ACP$_9$ co-substrate. The kinetic parameters were defined by fitting initial velocity vs concentration curves to the Michaelis-Menten equation using the curve fitting tool in OriginPro.*

4.6.5 Initial work towards the specificity engineering

The polyketide natural products have proven to be excellent therapeutics, including antibiotics, immunosuppressants and anticancer agents (Newman and Cragg, 2012). Polyketide backbone is mainly dependent on which extender unit the acyltransferase selects for the incorporation by the ketosynthase. Thus, engineering the specificity of the AT domains presents a promising approach to diversify the existing polyketide natural product libraries (Dunn and Khosla, 2013). Polyketide synthase AT domains have previously been shown to be highly specific to their native substrates, and the identification of several residues conveying specificity (Yadav et al., 2003; Minowa et al., 2007; Dunn and Khosla, 2013) led to a number of AT domain engineering studies (Dunn et al., 2013; Wang

et al., 2015; Dunn et al., 2014; Reeves et al., 2001; Del Vecchio et al., 2003; Sundermann et al., 2012).

We were interested in evaluating the possibility to engineer indanomycin derivates. One way to do this would be to engineer the malonyl-CoA-specific IdmO AT domain to accept non-native extender units, such as methylmalonyl-CoA, ethylmalonyl-CoA and propionyl-CoA. Historically, engineering studies were performed *in vivo* (Del Vecchio et al., 2003); however, such experiments are harder to control due to the abundance of many extender units in the environment. Since the AT assay conditions were established already (section 4.6.3), we decided to perform the mutagenesis and activity screening *in-vitro*. To do this, five IdmO AT mutants (outlined in Table 4-5) were made using site-directed mutagenesis.

Table 4-5 IdmO AT mutants generated. *Fingerprint residue sequences for specific substrates were acquired from the literature (Dunn and Khosla, 2013; Yadav et al., 2003; Starcevic et al., 2008).*

	Motif	**Theoretical substrate**
Wild-type	HAFH	Malonyl-CoA
AT1	YASH	Methylmalonyl-CoA
AT2	HASH	Hybrid between wild-type and AT1 #1
AT3	YAFH	Hybrid between wild-type and AT1 #2
AT4	VASH	Ethylmalonyl-CoA
AT5	AAAH	Promiscuous

All mutants were successfully cloned and expressed in their *holo-* form in *E. coli* Nico21(DE3). In this strain, four main metal binding contaminants from *E. coli* (SlyD, Can, ArnA, and AceE) have a chitin binding domain (CBD) fused to them (Robichon et al., 2011). This allows the removal of these proteins from the purified protein solution by chitin affinity chromatography. The use of this particular strain allowed the modification of the purification procedure (Figure 4-33a) so that all six AT mutant (including the wild-type) could be purified in parallel (Figure 4-33b).

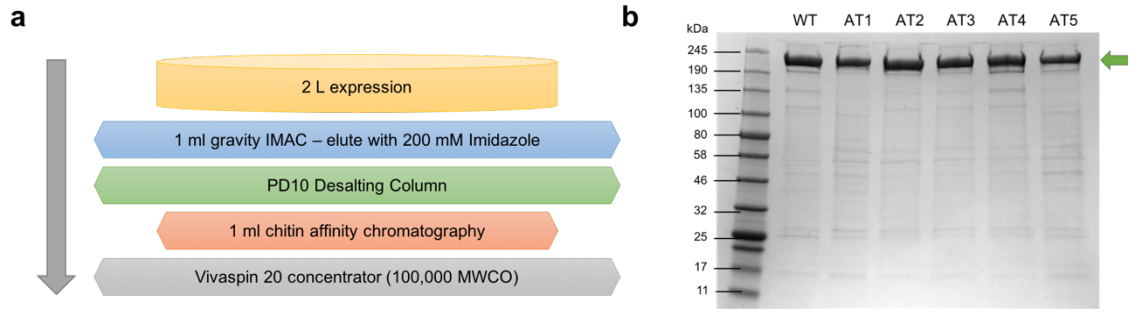


Figure 4-33 Purification of the AT mutants. *(a) Purification protocol scheme utilising two affinity chromatography steps (IMAC and chitin). (b) Reducing SDS-PAGE gel of purified AT mutants.*

Unfortunately, all attempts to perform the AT assay with the AT active site mutants failed to produce interpretable data. One assay attempt takes the full day, and it appeared that the mutants were losing their activity very soon after being thawed. Going forward, time needs to be taken to investigate the optimum purification and buffer conditions to retain the activity of the enzymes for at least several days.

It is also possible that an alternative purification strategy, involving *E. coli* Nico21(DE3) strain played a role in reducing the activity of recombinant IdmO. To control for this, IdmO would need to be purified using the old and the new protocols and tested in parallel.

4.7 Cryo-electron microscopy

4.7.1 Initial cryo-EM experiments

Single particle cryo-EM analysis was performed to study the architecture and domain interactions of the fully-reducing IdmO module. An IdmO sample from section 4.5.2.2 was applied to a cryo-EM grid and plunge frozen into liquid ethane for vitrification. The grid was screened on a Titan Krios microscope operated at 300 keV. Figure 4-34 shows a raw cryo-EM micrograph with several particles resembling a 2D class average from the preliminary negative stain EM analysis (Figure 4-10).

Unlike in negative stain EM analysis, a lot more heterogeneity was observed in this cryo-EM sample with proteinaceous species of different shapes and lengths observed throughout the micrographs. Nevertheless, some particles carried a significant resemblance to the previously observed 2D class averages, and,

therefore, a single particle data collection was performed. Nearly 5000 micrographs were collected, from which particles were picked using a semi-automatic approach, and class averages were produced using 2D classification algorithms (Figure 4-35).

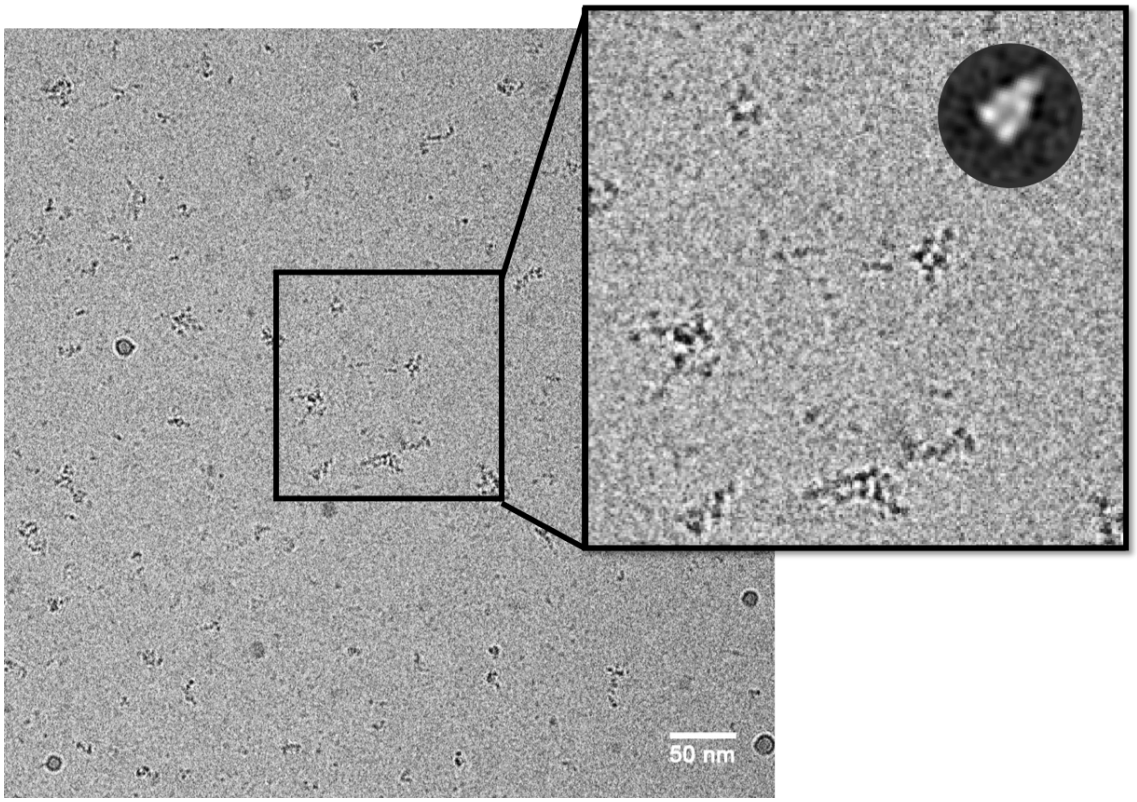


Figure 4-34 Raw cryo-EM micrograph of TEV/Histag IdmO. *Proteinaceous species of different shapes and sizes can be seen throughout. Inset: a magnified portion of the micrograph highlighting several particles resembling a 2D class average (shown in a dark circle) from negative-stain EM analysis (Figure 4-17). Scale bar represents 50 nm.*

Unfortunately, 2D class averages resembled spaghetti-like structures and did not reveal any high-resolution features indicating strong heterogeneity within the sample. Similar flexible architecture was reported previously for a multi-module non-ribosomal peptide synthetase (Tarry et al., 2017).

Some of the classes appeared to be a lot smaller than others signifying the equilibrium between monomeric and dimeric IdmO in the sample. A similar problem of equilibrium between oligomeric states with most particles being monomers was previously observed for cryo-EM samples of a fungal iPKS

(Herbst et al., 2018). In this particular study, the dimeric species were selected for through the optimisation of the buffer composition and purification strategy.

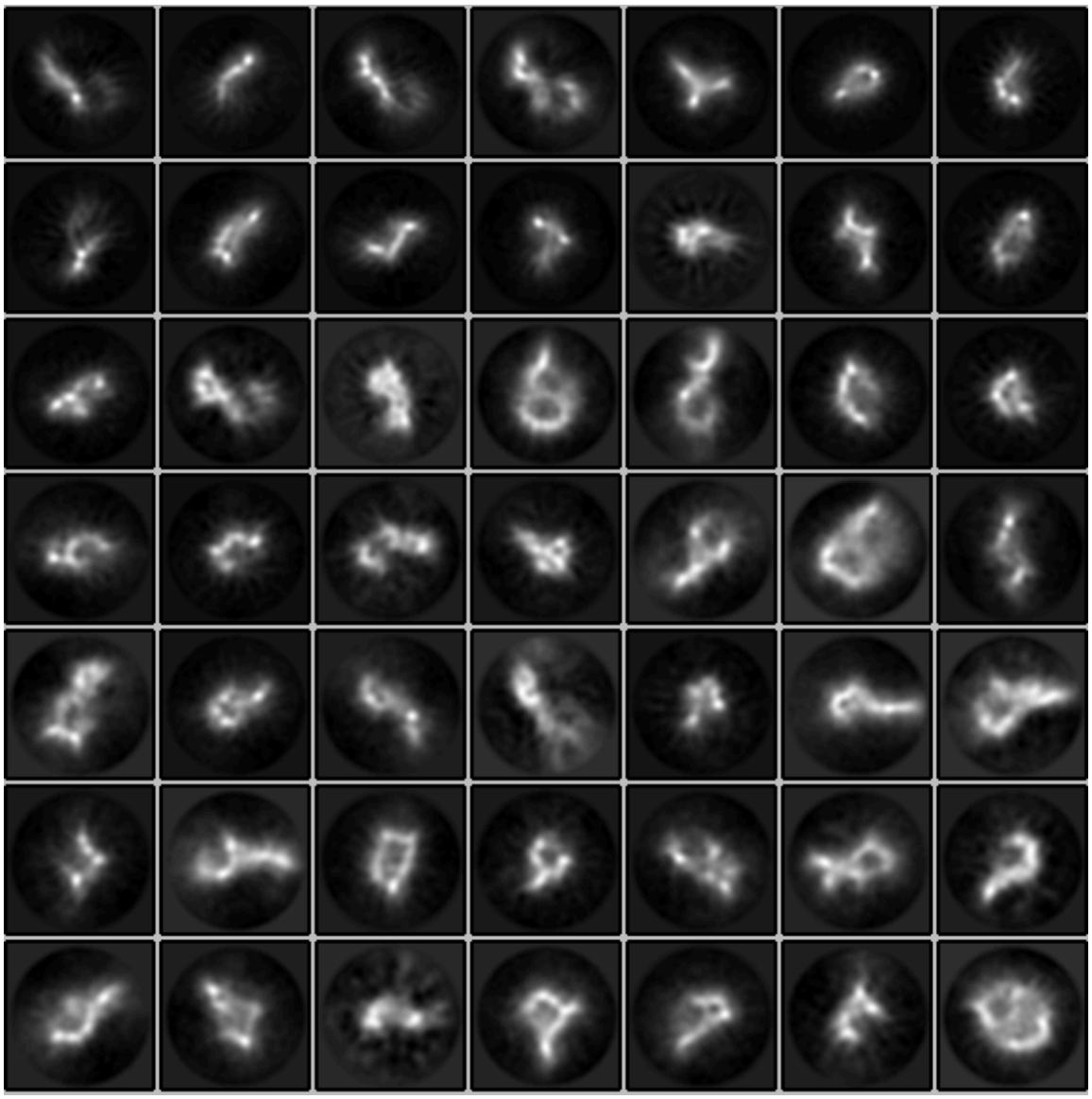


Figure 4-35 2D class averages produced from the first cryo-EM data collection of IdmO. *The absence of the high-resolution features in the 2D class averages signifies high heterogeneity within the sample. Classes in the first couple of rows seem to be significantly smaller than in the remaining rows indicating the possibility of these classes to consist of IdmO monomer species and the rest of the rows to consist of dimeric species.*

Another reason for the high level of heterogeneity within the sample could arise from cryo-EM sample preparation (Drulyte et al., 2018). When a sample is applied to a cryo-EM grid, the excess solution is blotted off using filter paper and only a very thin (<100 nm) layer of the sample is left behind. Due to Brownian motions

(Trurnit, 1960), the protein molecules will collide with the air-water interface at thousands of times per second in-between the blotting and plunging. While some proteins are not affected by this interaction, most of the protein molecules will adhere to it (either adopting random or preferred orientations) and some will either partially or fully denature (Noble et al., 2018).

We envisioned that these interactions at the air-water interface were the main reason for the dotty appearance of IdmO sample in the cryo-EM images. Since a single IdmO dimer comprises twelve catalytic domains linked together by a single polypeptide chain, it is possible that the air-water interface, coupled with the surface tension, leads to the unfolding of the quaternary structure and hence the appearance of dots (single domains) in the micrographs.

Problems associated with cryo-EM sample preparation have not been extensively studied or discussed and is a relatively new aspect of the field. In fact, we recently published the first article (Drulyte et al., 2018) reviewing different approaches one can take to optimise particle distribution in cryo-EM sample preparation. We decided to test several of these sample optimisation approaches to see if IdmO heterogeneity could be reduced allowing us to perform high-resolution cryo-EM single particle analysis.

4.7.2 Optimisation of cryo-EM sample preparation

This section summarises the approaches taken to optimise IdmO sample for high-resolution cryo-EM.

4.7.2.1 Buffer optimisation

Making and screening cryo-EM grids is both a time- and cost-ineffective way to assess the sample stability in different buffers. Therefore, a negative-stain EM protocol was used to test the following parameters:

1. Flash freezing and thawing.
2. Addition of β-mercaptoethanol reducing agent.
3. Addition of 1% glycerol stabilising agent.
4. Storing the sample at 4 °C for one week.
5. Sodium phosphate vs HEPES buffer.
6. Different pH (6.8, 7.5, 8.2).

The screening protocol involved purifying IdmO as described in section 4.5.2.2, dialysing the sample into the buffer of interest overnight, making a negative stain EM grid and screening it on a Tecnai T12 electron microscope.

The results indicated that there was no difference in particle distribution and homogeneity between fresh, one week old and flash frozen-thawed IdmO samples. Addition of β-mercaptoethanol or 1% glycerol did not affect the particle distribution, and there were no changes observed when using HEPES rather than the original phosphate buffer or changing the pH.

The only difference observed was in particle numbers. Dual-affinity tag (old) purification protocol elution sample at 0.04 mg/ml can be seen in Figure 4-36b while the new purification protocol (including an overnight TEV protease cleavage step and size-exclusion chromatography) elution sample at 0.05 mg/ml can be seen in Figure 4-36a. While the difference in measured concentration was only 0.01 mg/ml, a significant increase in IdmO particles was observed in the sample prepared using the new protocol. This was rationalised to the use of size-exclusion chromatography, since all particles seemed to be approximately the same size, while the old purification protocol resulted in many aggregates.

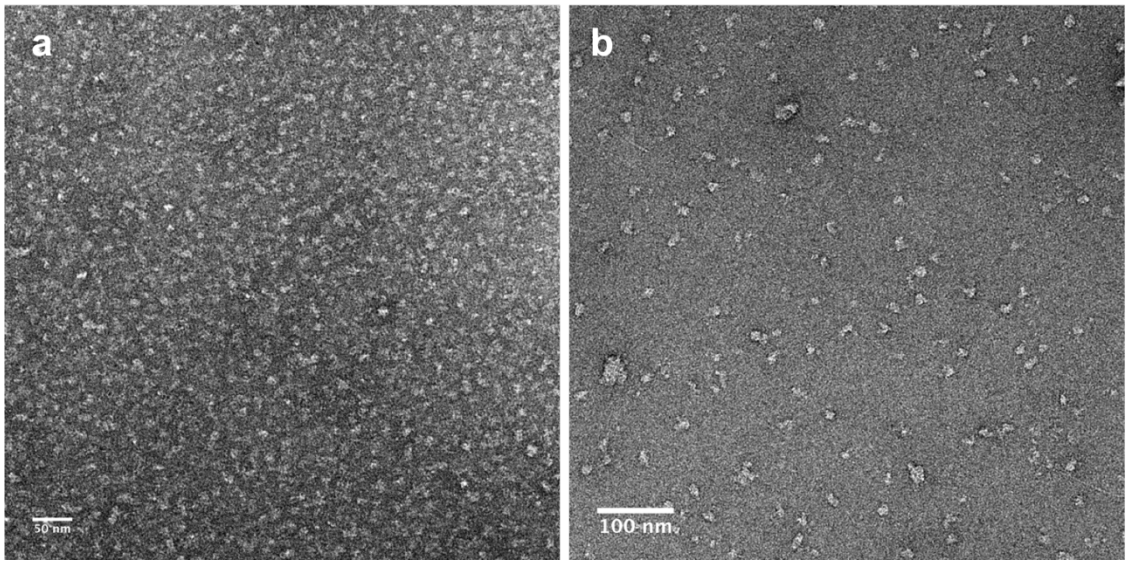


Figure 4-36 Comparison of IdmO sample purified using two different protocols. *(a) IdmO sample at 0.05 mg/ml concentration purified using a new protocol (section 4.5.2.2). (b) IdmO sample at 0.04 mg/ml concentration purified using an old dual-affinity chromatography protocol (section 4.3.4).*

4.7.2.2 Sample concentration

To assess the effects of diluting the sample, four cryo-EM grids were made:

1. Neat size-exclusion chromatography fraction at 0.5 mg/ml concentration.
2. Diluted sample at 0.25 mg/ml concentration.
3. Diluted sample at 0.1 mg/ml concentration.
4. Diluted sample at 0.05 mg/ml concentration.

Neat concentration grid was comparable to the sample from the first cryo-EM data collection (Figure 4-34) while all grids prepared with diluted sample displayed tiny dots seen throughout the images indicating denatured IdmO. Therefore, it was concluded that dilution of the sample should be avoided, or at least until the very last minute during cryo-EM grid preparation.

4.7.2.3 Continuous carbon EM supports and surfactants

As discussed above, the air-water interface is a harsh environment for many proteins; and sticking to this interface often leads to a degree of denaturation. In the case of IdmO, the denaturation seems almost complete as very few full-size particles are observed, and the micrographs mostly consist of individual domain (10-80 kDa) speckles (or “dots”). Using an EM support with a continuous layer of thin carbon presents an opportunity to protect the molecules from the air-water interface by positioning them at the support-water interface instead. It is worth mentioning, however, that this tactic does not come without its drawbacks. A layer of 2-3 nm carbon significantly increases the noise level in the micrographs, and some particles may not have much affinity to it, or if they do, they might also partially denature upon binding, or exhibit a preferential orientation (Drulyte et al., 2018).

To assess whether denaturing at the air-water interface could be resolved using this approach, grids containing a continuous layer of carbon (Quantifoil R2/2 overlaid with 2 nm carbon and Lacey carbon with 3 nm carbon) were trialled. Surprisingly, no particles at all (denatured or not) could be observed on the lacey grid (Figure 4-37a). Continuous carbon Quantifoil grid showed some particles of approximately the right size and shape as IdmO; however, many dots remained on the grid indicating that some IdmO molecules either did not bind to the continuous support or they bound and denatured nonetheless (Figure 4-37b).

Since no protein molecules were observed on a lacey grid overlaid with 3 nm carbon, we decided to try more-transparent graphene oxide instead. Unfortunately, none of the graphene oxide grids screened displayed globular particles and, what's more, all grids were significantly contaminated or wrinkled (which are the two main disadvantages of this method).

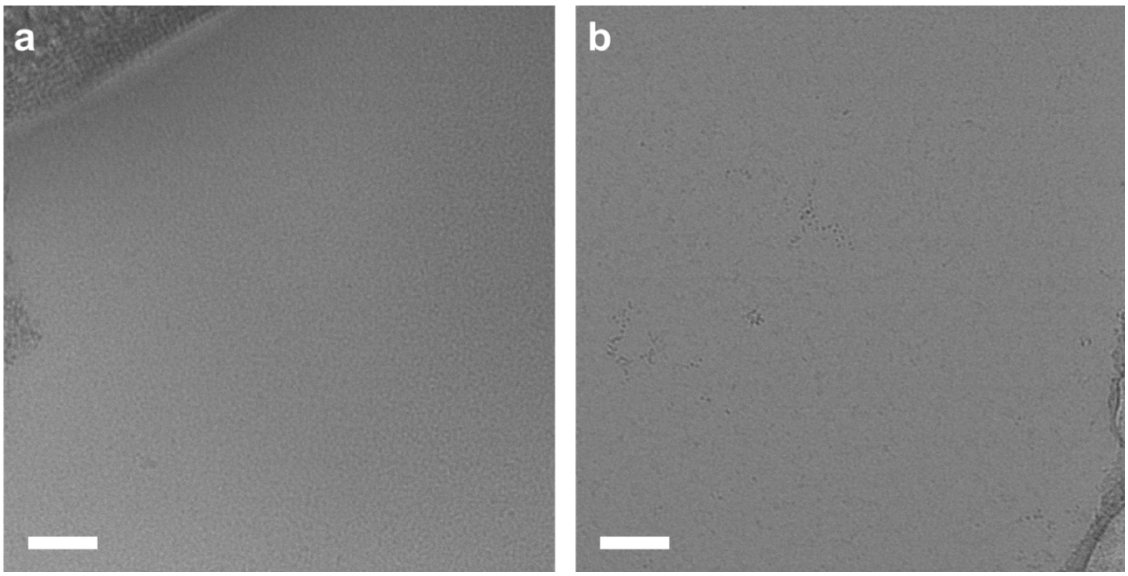


Figure 4-37 The effects of a continuous carbon film on the appearance of IdmO. *(a) No particles can be seen on the lacey EM support (Agar Scientific) with a continuous layer (3 nm) of amorphous carbon. (b) Some particles (both intact and completely denatured) can be seen on the Quantifoil R2/2 EM support (Quantifoil) with a continuous layer (2 nm) of amorphous carbon. Scale bars represent 50 nm.*

Another approach to protecting the sample from the forces present at the air-water interface includes adding a surfactant such as dodecylmaltoside (DDM) (Glaeser et al., 2016). The hydrophobic part of the detergent, when added at a correct concentration to the sample buffer (below the critical micelle concentration), forms a monolayer at the air-water interface, shielding the specimen from this interface (Glaeser and Han, 2017).

The effects of adding 0.0025% and 0.005% DDM to IdmO sample buffer were evaluated. The sample containing 0.0025% DDM displayed a previously observed distribution of "speckles" while the sample containing 0.005% DDM had fewer speckles but displayed some larger proteinaceous aggregates (Figure 4-38).

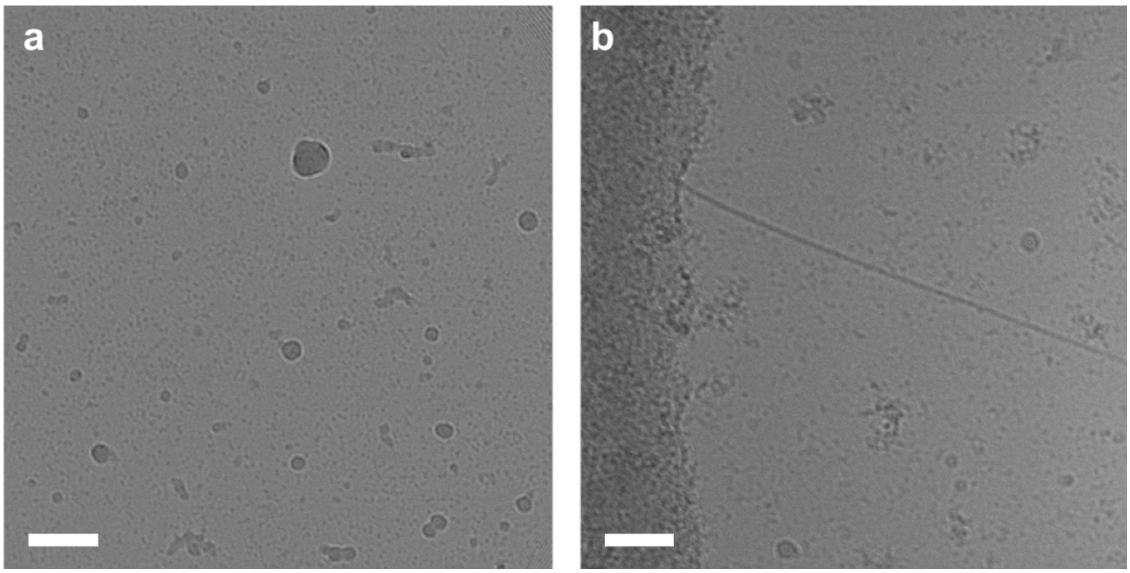


Figure 4-38 Investigating the effects of surfactant dodecylmaltoside on the IdmO cryo-EM sample quality. *(a) Sample with 0.0025% DDM shows many denatured IdmO speckles. (b) Sample with 0.005% DDM shows fewer speckles but also some larger aggregates. Scale bars represent 50 nm.*

4.7.2.4 Glutaraldehyde crosslinking

Another approach commonly used to reinforce fragile complexes during cryo-EM sample preparation is chemical crosslinking (Engel et al., 2016; Nguyen et al., 2018; Poepsel et al., 2018; Kasinath et al., 2018; Kang et al., 2017). First, the optimum concentration of glutaraldehyde (crosslinker) and the duration of the crosslinking reaction to obtain the most crosslinked dimeric IdmO was determined. A small-scale experiment varying these parameters was set up, and the level of crosslinking (as measured by the ratio of monomeric and dimeric, or crosslinked, IdmO) was evaluated by SDS-PAGE. The results of this experiment suggested that the optimum crosslinking was achieved when 0.05% glutaraldehyde was used, and when the crosslinking was performed at 4 °C for two hours.

A larger scale crosslinking was performed, and the crosslinking was evaluated by performing negative stain EM analysis. A small dataset was collected, and particles were extracted for 2D classification. Some 2D class averages are shown in Figure 4-39. Unfortunately, 2D class averages had even fewer features than in the non-crosslinked dataset (Figure 4-23), and 3D refinement did not result in a sensible solution. One of the possible explanations for these featureless, fuzzy

classes is over-crosslinking, which can happen when glutaraldehyde forms polymers. Going forward, only EM-grade glutaraldehyde should be used, and it should also be stored in small aliquots at -80 °C upon delivery.

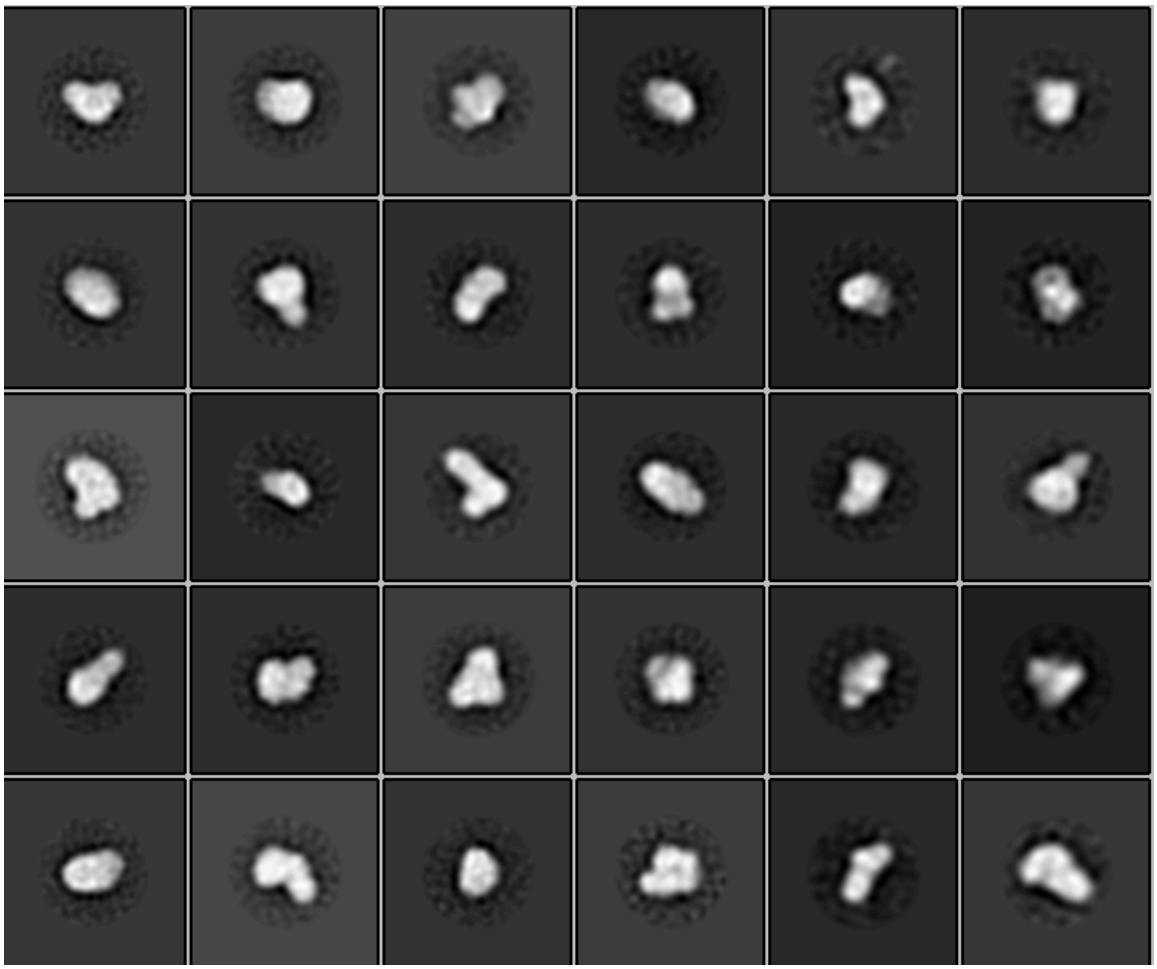


Figure 4-39 2D class averages of glutaraldehyde-crosslinked IdmO. *The class averages appeared fuzzy with very few features suggesting that the sample might have been over-crosslinked. Parameters used in this data processing are as follows: 128-pixel box size (which equates to 450 Å at a pixel size of 3.51 Å/pix) and 250 Å diameter mask.*

Besides crosslinking in solution, another popular chemical crosslinking approach is GraFix (Stark, 2010). When using this method, a sample is crosslinked as it travels down the density gradient during ultracentrifugation. The main downside of this approach is the need to remove the density sugar prior to cryo-EM grid preparation, which can be detrimental to the sample quality.

BS3 (bis(sulfosuccinimidyl)suberate) crosslinker has become more widely used in cryo-EM recently (Nguyen et al., 2018; Poepsel et al., 2018; Kasinath et al.,

2018). It is an amine-to-amine crosslinker and therefore is more specific compared to glutaraldehyde. What is more, it does not polymerise and therefore is thought to be a safer option for gentle crosslinking of cryo-EM specimens.

4.7.2.5 Hofmeister series and second cryo-EM dataset

We hypothesised that the unfolding at the air-water interface happens as a result of relatively week interdomain interactions. Certain ions, as described in the Hofmeister series, are capable of increasing the solvent surface tension and decreasing the solubility of a molecule, which then results in the strengthening of the hydrophobic interactions. This phenomenon is often referred to as salting-out (Hyde et al., 2017) and is commonly used in protein purification through the use of ammonium sulfate precipitation. To test whether the hydrophobic interactions within an IdmO dimer can be strengthened using this approach, purified IdmO was dialysed into either Na_2HPO_4 or Na_2SO_4 buffer. Cryo-EM grids were made of all samples and, to our surprise, a sample dialysed into Na_2SO_4 buffer displayed globular particles across all micrographs taken (Figure 4-40).

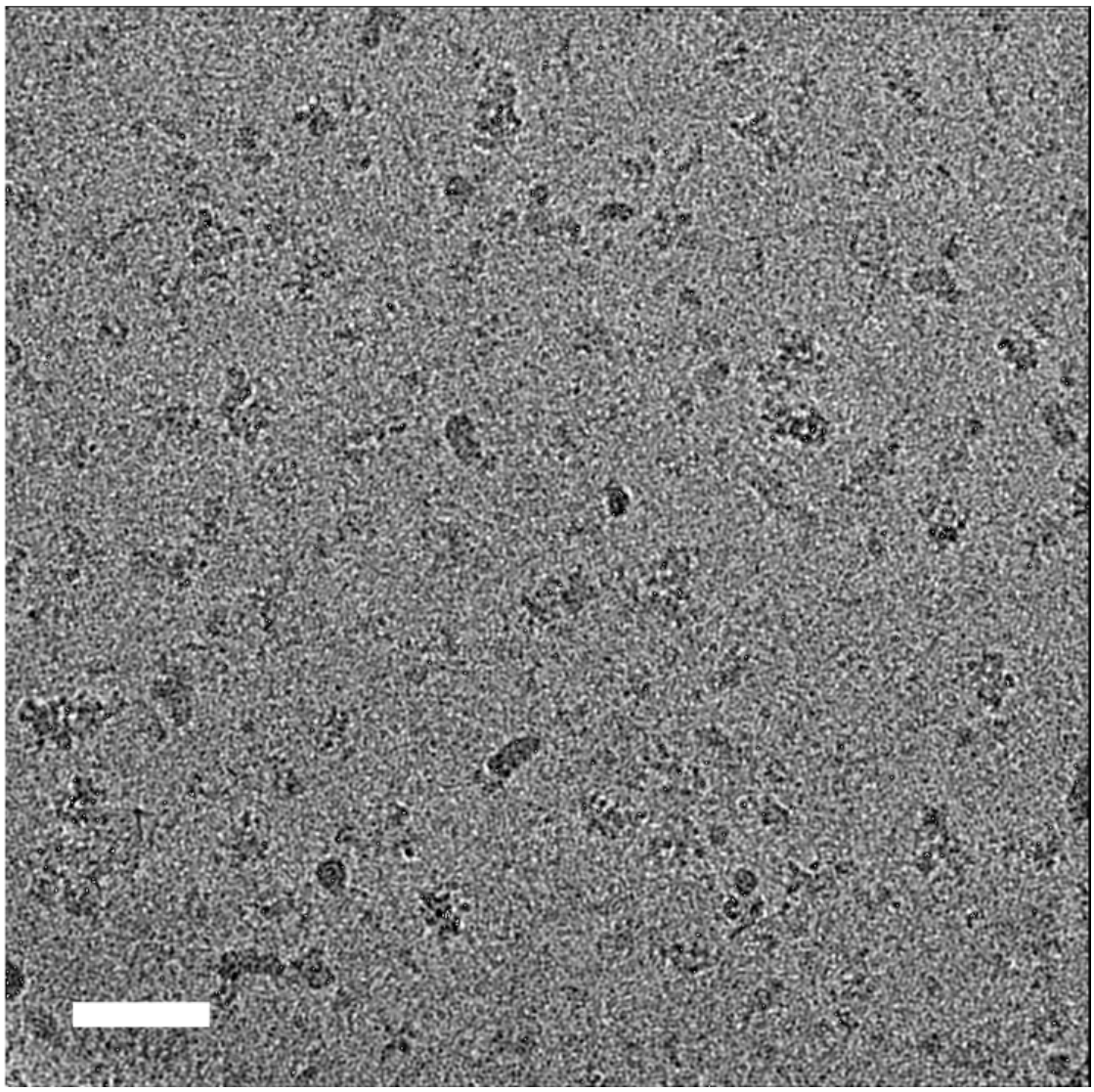


Figure 4-40 Raw cryo-EM micrograph of IdmO sample dialysed in 50 mM Na_2SO_4, 200 mM NaCl buffer. *Despite the high background noise, some proteinaceous, globular species could be seen on the micrograph. Scale bar represents 50 nm.*

The noise of the micrographs was a lot higher than observed previously which was thought to be a result of either high ionic strength buffer or a layer of denatured protein at both air-water interfaces. Nevertheless, particles from the micrographs taken during a screening session were manually picked, which resulted in a total of 400 particles. These particles were classified into four 2D classes, which are shown in Figure 4-41a. Surprisingly, these classes displayed a strong resemblance to the negative-stain class averages (Figure 4-41b) of double-tagged IdmO (Figure 4-17).

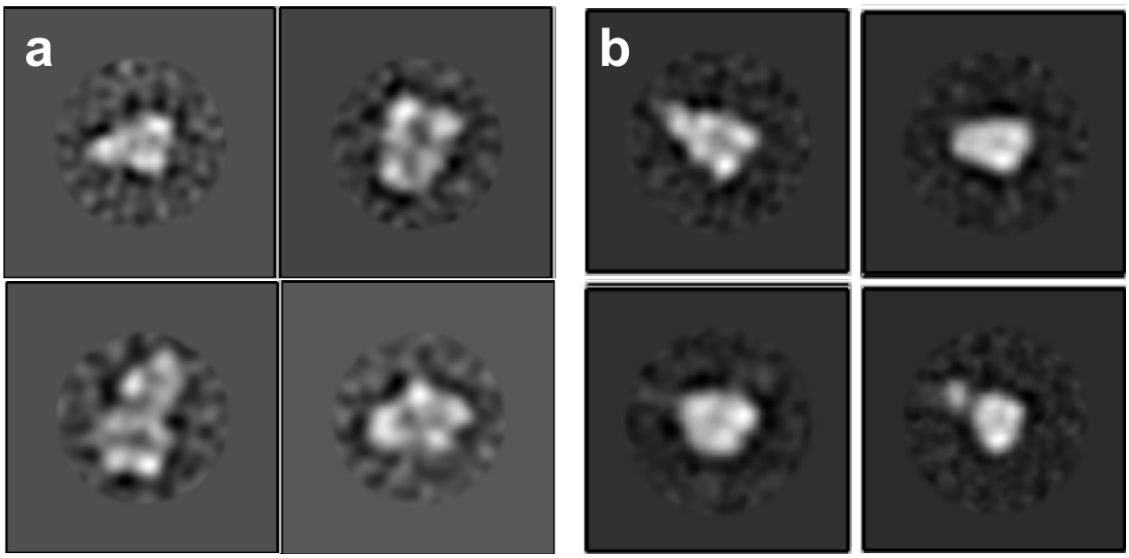


Figure 4-41 Comparison of cryo-EM 2D class averages to the negative-stain EM 2D class averages. *(a) Cryo-EM 2D class averages derived from 400 manually picked particles from the non-motion- or CTF-corrected micrographs. (b) Negative-stain EM 2D class averages of double-tagged IdmO (Figure 4-17).*

Since these screening micrographs were acquired without recording movie frames and hence were not motion- or CTF-corrected, a short (24 hour) dataset to evaluate the quality of the sample in more depth was collected. Approximately 2000 micrographs were acquired of IdmO sample in sodium sulphate buffer. From this dataset, nearly 100,000 particles were auto-picked, and four rounds of 2D classification were performed which reduced the set of particles in half. 2D class averages from the fourth classification job can be seen in Figure 4-42. Dimeric “twisted ribbon” shape could be seen in several of the 2D class averages, despite very poor signal to noise ratio. The most likely reason for such high background and poor-quality images is the presence of two layers of completely denatured IdmO, coating both air-water interfaces. The complete coverage of these interfaces would then allow fully assembled particles to be suspended in the vitreous layer intact, resulting in these classes resembling previously acquired negative stain class averages.

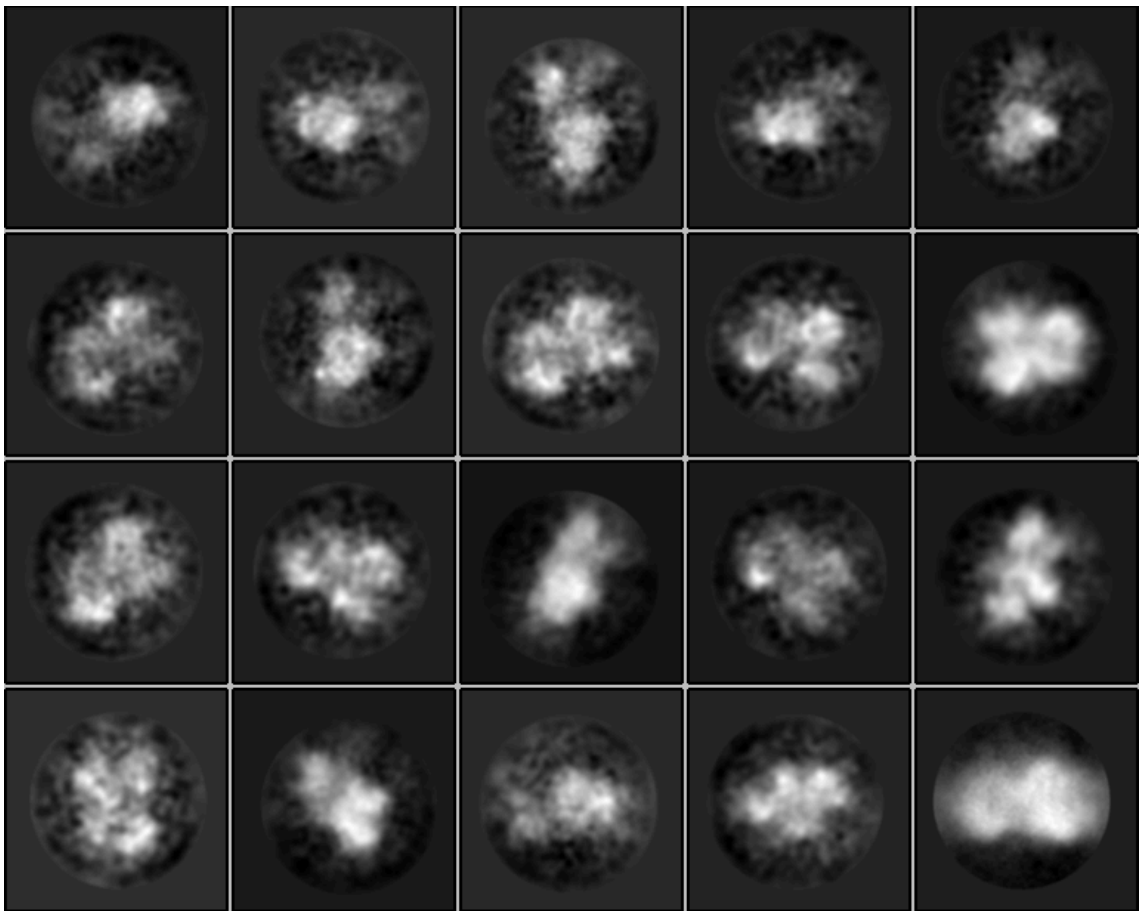


Figure 4-42 2D class averages of IdmO in sodium sulphate buffer. *Nearly 2000 micrographs were acquired which resulted in a dataset containing 100,000 particles which were reduced in half over four rounds of 2D classification. Dimeric symmetry could be seen in many of these averages; however, high background noise, coupled with the heterogeneity of the sample, resulted in very poor images.*

The work described in this chapter goes from constructing a plasmid vector suitable for producing recombinant *holo*-IdmO to the preliminary analysis of this module by an enzymatic assay and several EM techniques. During this project, many pitfalls were identified and successfully overcome. However, more challenges, such as the high heterogeneity and instability of IdmO in thin aqueous layers, need to be overcome to determine the high-resolution structure of this polyketide synthase module. Hopefully, this work will pave the way towards this goal in the future.

4.8 Summary

This chapter describes the steps taken to investigate the structure and function of IdmO, a fully-reducing polyketide module from indanomycin biosynthetic pathway (Figure 4-1).

The 6.5 kb gene encoding IdmO was amplified from the genomic DNA of *S. antibioticus* using nPCR approach (Figure 4-2 and Figure 4-3). This allowed the construction of pET28a_IdmO vector. Soluble expression of recombinant IdmO was achieved in *E. coli* BL21(DE3); however, due to low yield, metal binding proteins from *E. coli* prevented the obtaining of a homogenous sample (Figure 4-6). The incorporation of an N-terminal Strep II tag (Figure 4-14) and a subsequent modification of purification procedure to incorporate two affinity chromatography steps (Figure 4-15), allowed us to purify IdmO to homogeneity. Peptide fingerprinting analysis was performed (Figure 4-16) which confirmed the successful purification of the full-length IdmO.

Negative-stain EM analysis of IdmO was performed to obtain a low-resolution reconstruction of IdmO. The resulting map of IdmO revealed a dimeric architecture with a reaction chamber located in the middle of the arch (Figure 4-17). Similar domain arrangement has been reported previously and is likely to be a fundamental adaptation of mPKSs to facilitate a forward transfer of polyketide intermediates and increase the specificity of the reaction (Dutta et al., 2014). While IdmO model resembles that of PikAIII, it shows strong differences to previously reported X-ray structures of truncated PKS modules (Weissman, 2015b; Weissman, 2015a; Robbins et al., 2016) highlighting the suitability of electron microscopy for studying dynamic polyketide synthase modules.

To further increase our understanding of the interfaces between the functional domains and the rules governing the specificity of polyketide assembly lines, we set out to obtain a high-resolution structure of IdmO as well as to develop an activity assay which can be used to assess the indanomycin PKS engineering potential. First, the purification procedure was modified to increase the concentration of purified IdmO. This was achieved by cloning in a TEV protease cleavage site between the gene and the C-terminal hexahistidine tag (Figure 4-20). The purification procedure was modified to swap the second affinity chromatography step with an overnight tag cleavage by TEV protease, followed

by reverse IMAC metal affinity and size-exclusion chromatography (Figure 4-21). The purified IdmO sample was analysed by SEC-MALLS which revealed an equilibrium between monomeric and dimeric IdmO species (Figure 4-22).

To measure the transacylation activity of IdmO, a previously reported AT activity assay was adapted (Dunn et al., 2013). Michaelis-Menten kinetic calculations were performed from which the K_M of IdmO was calculated to be 3.6 μM (Figure 4-32), a value similar to the ones reported for the DEBS AT domains (2-7 μM) (Dunn et al., 2013). To the best of our knowledge, this was the first time such continuous assay was used to examine the activity and reaction kinetics of the AT domain in the full-length PKS module.

Despite our continuous efforts, we were unable to determine a cryo-EM structure of IdmO. Initially, cryo-EM micrographs revealed a lot more heterogeneity compared to the negative-stain EM data (Figure 4-34). The collection of the first dataset, auto-picking of particles and subsequent 2D classification revealed high flexibility of the complex, as well as the presence of both monomeric and dimeric species (Figure 4-35). We hypothesised that the heterogeneity and partial (or full) denaturation of IdmO was a result of the interactions at the air-water interface (Drulyte et al., 2018). Many parameters, such as grid preparation (Figure 4-37), crosslinking (Figure 4-39) and buffer components (Figure 4-38), were altered with the hope of stabilising the IdmO dimer; however, very little success was achieved.

The most progress in stabilising the interdomain hydrophobic interactions was made when using high ionic strength ions from the Hofmeister series (Hyde et al., 2017). A second cryo-EM dataset was collected from the sample dialysed into 75 mM Na_2SO_4 containing buffer. The picking of particles and the subsequent 2D classification revealed class averages with significant similarities to the 2D class averages from the negative stain EM data (Figure 4-41); however, the poor signal-to-noise ratio of the cryo-EM images prevented us from a successful calculation of a 3D volume. In our opinion, the high level of noise was arising from the monolayer of unfolded and denatured IdmO at both air-water interfaces. Going forward, the most promising approaches to deal with the sample preparation problems include the use of GraFix (Stark, 2010) and the substitution of glutaraldehyde with a more specific crosslinker BS3.

In summary, this chapter describes the cloning, recombinant expression and purification of the first fully-reducing PKS module IdmO. A number of biochemical

techniques (mass spectrometry, SEC-MALLS, enzyme kinetics) were used to characterise IdmO. Negative stain EM analysis was used to calculate a first three-dimensional map of IdmO which revealed a dimeric architecture. Cryo-EM analysis was attempted; however, further optimisation of the sample preparation needs to be carried out to overcome the problems associated with the denaturing of the IdmO dimer at the air-water interface.

Chapter 5 High-resolution cryo-EM structure of the full-length module from the rhizoxin polyketide synthase reveals novel insights into the ACP binding

5.1 Introduction

This chapter describes the structural work performed on rhizoxin PKS as part of a collaborative project between University of Leeds (Ranson and Berry groups), University of Tübingen (Dr Georg Zocher) and HKI-Jena (Hertweck group). All electron microscopy work was carried out by the candidate at the Astbury Biostructure Laboratory while all biochemical work was performed by Maria Dell (a PhD student from Prof. Christian Hertweck's group).

5.1.1 Challenges in polyketide synthase engineering

Type I modular PKSs (mPKSs) are giant assembly lines responsible for the production of structurally and functionally diverse polyketide natural products (Figure 1-2). Polyketides have proved to be excellent starting points for a wide variety of pharmaceutical drugs with immunosuppressant, antibacterial, antifungal or anticancer properties (Newman and Cragg, 2012; Fischbach and Walsh, 2006).

Since the discovery of PKSs in the early 1990s, considerable efforts have been made to understand the structure and organisation of individual domains and modules within PKSs, which would facilitate rational re-engineering of these systems (Wong and Khosla, 2012; Gokhale et al., 1999; Chen et al., 2006; Edwards et al., 2014; Weissman, 2015a; Marsden et al., 1998; Cortes et al., 1995; Weissman and Leadlay, 2005). While the prediction of the PKS module and domain arrangement and thus the resulting polyketide product from the genetic information alone seems fairly forward, most of the efforts to make chimeric PKS assembly lines result in truncated products, dramatically reduced yields or unexpected side products (Weissman, 2015a). This highlights a need to gain a better understanding of the structural arrangement of PKS modules and subunits (which contain more than one module).

X-ray crystallography has proven to be extremely useful in determining the structures of many individual domains and didomains from different PKS systems

(Tang et al., 2006; Scaglione et al., 2010; Tsai et al., 2001; Ames et al., 2012; Keatinge-Clay and Stroud, 2006; Liew et al., 2012); however, the heterogeneous nature of flexible PKS modules has led to all efforts to crystallise a full PKS module being unsuccessful.

5.1.2 Structural studies of full-length polyketide synthase modules

Structural information of intact mPKS modules is limited to several cryo-EM structures of the fifth module from pikromycin PKS (Dutta et al., 2014; Whicher et al., 2014) revealing a previously unobserved orientation of the catalytic domains forming an arch architecture (Figure 1-18), and low-resolution models of DEBS 3 and DEBS3-TE as determined by SAXS (Figure 1-19a) (Edwards et al., 2014).

Two fungal iterative polyketide synthases (iPKSs), which are single-module systems similar to mPKS, were investigated recently using hybrid approaches by the Maier group (Herbst et al., 2016; Herbst et al., 2018). Both iPKSs were shown to resemble better-characterised fatty acid synthases (Figure 1-17) from which PKSs are thought to have evolved. In the first study by Herbst et al. (Herbst et al., 2016), a hybrid crystal structure of the reducing fungal iterative mycocerosic acid synthase-like PKS (Figure 1-20) was reported which combined two independent crystal structures of its condensing and modifying regions.

In the second, more recent study (Herbst et al., 2018), a crystal structure of a condensing region of the non-reducing cercosporin PKS was reported together with a cryo-EM structure of the same region cross-linked to a single modifying region ACP. The cryo-EM results revealed considerably greater asymmetry than the crystal structure without the ACP (Herbst et al., 2018). What is more, the stoichiometry of the ACP binding was revealed to be 1:2 with only one ACP being able to bind to the iPKS at a given time. The binding event was proposed to induce a conformational change in the complex and thus prevent the binding of the second APC.

Understanding the structure and dynamics of fully-functional PKSs remains a central challenge in natural product research. Pikromycin mPKS and cercosporin and MAS-like iPKS structural data present first significant breakthroughs in visualising full-length PKS modules in action, and provide insights for successful re-engineering of polyketide assembly lines.

5.1.3 Non-canonical branching module from the rhizoxin PKS

Modular PKS assembly lines fuse the polyketide extender units in a head-to-tail linear fashion (Weissman and Leadlay, 2005). This leads to a linear polyketide intermediate which is often cyclised and/or modified by post-PKS tailoring enzymes (section 1.3.4.6) to derive the final product (Olano et al., 2010).

One of the possible additional post-PKS modifications is branching of the linear backbone. At least twelve *trans*-AT PKS products with β-branching have been identified, and all twelve biosynthetic gene clusters encoding the enzymatic machinery responsible for the generation of these polyketides have been shown to possess the same conserved β-branching cassette (Calderone, 2008). This cassette encodes five freestanding proteins which act *in trans* to incorporate a β-branch in the growing polyketide. These core proteins include ACP and KS domains (the latter of which lacks a conserved catalytic cysteine residue), a hydroxymethylglutaryl-CoA synthase homolog and two enoyl-CoA hydratase homologs, the first one serving as a dehydratase and the second one as a decarboxylase (Calderone, 2008).

Rhizoxin, a polyketide natural product (Figure 5-1) produced by the endofungal bacterium *Burkholderia rhizoxinica* (Partida-Martinez and Hertweck, 2005; Scherlach et al., 2012), is a strong antimitotic agent, the activity of which is dependent on a short (C2) carbon chain (shown in red in Figure 5-1) branching away from the main macrolide ring (Scherlach et al., 2006; Kusebauch et al., 2011; Schmitt et al., 2008; Hong and White, 2004).

Sequencing of the *B. rhizoxinica* genome failed to detect any enzymes, such as those identified in the AT-less PKS β-branching cassette (Calderone, 2008), which could act *in trans* to introduce this branching (Lackner, Moebius, Partida-Martinez and Hertweck, 2011; Lackner, Moebius, Partida-Martinez, Boland, et al., 2011). Further bioinformatic analysis and investigation of the prematurely released products of the rhizoxin assembly line led to the discovery of a novel type of PKS module (Figure 1-24) which itself catalyses a branching reaction in the polyketide backbone (Partida-Martinez and Hertweck, 2007; Kusebauch et al., 2009).

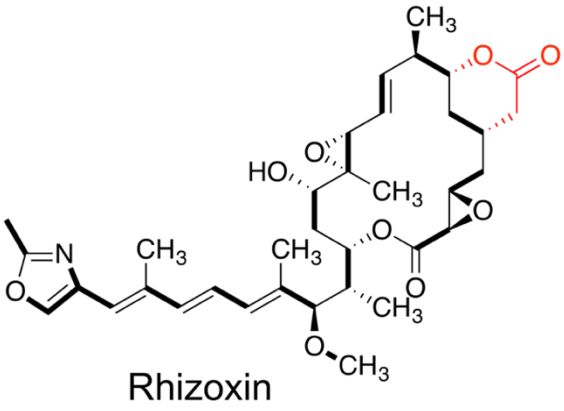


Figure 5-1 The rhizoxin polyketide natural product produced by the endofungal bacterium Burkholderia rhizoxinica (Partida-Martinez and Hertweck, 2005). *This natural product possesses strong antimitotic properties, which were shown to be a result of a short branch (shown in red) of the main macrolide ring (Scherlach et al., 2006; Kusebauch et al., 2011; Schmitt et al., 2008; Hong and White, 2004).*

This new type of module, termed “branching”, is part of RhiE subunit and is located in between modules 10 and 11 of the rhizoxin PKS (Bretschneider et al., 2013). It consists of three domains, a ketosynthase, an acyl-carrier protein and a branching domain (B), and has the mass of approximately 115 kDa (Figure 1-24).

In vitro reconstitution and mutation experiments were performed which confirmed that this module was indeed responsible for introducing the branch in the linear backbone of rhizoxin (Bretschneider et al., 2013). The proposed branching mechanism, involving the Michael addition of the malonyl unit and a subsequent lactonization of the thioester, is shown in Figure 5-2.

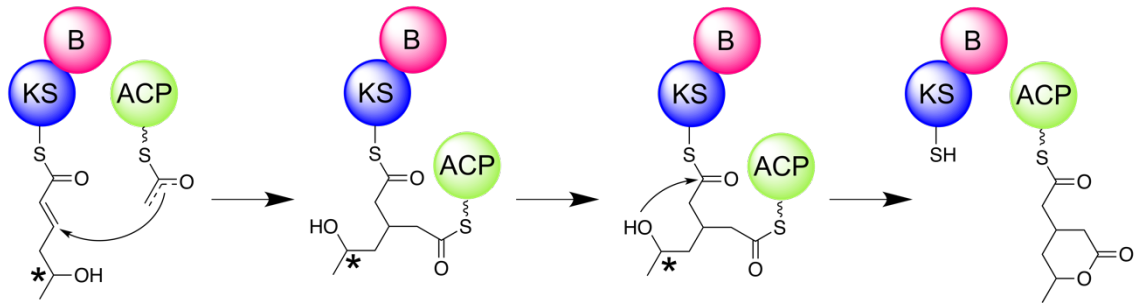


Figure 5-2 The proposed model of rhizoxin branching (Bretschneider et al., 2013). *First, the nucleophilic attack of the malonyl unit bound to the ACP results in the polyketide intermediate being covalently tethered to both the KS and the ACP. Then, the C5 hydroxyl moiety of the thioester intermediate attacks the C1 carbonyl bound to the KS resulting the formation of the lactone ring. In experiments including the deoxy substrate analogue, the C5 hydroxyl moiety (denoted with an asterisk) was missing. It was hypothesised that this deoxy analogue would still undergo the Michael addition of the ACP-bound nucleophile, but would be unable to attack the C1 carbonyl during the subsequent lactonization of the thioester.*

To further characterise this module and to visualise the cryptic B domain – which lacks sequence homology to any characterised PKS domains – a crystal structure of KS-B didomain (lacking the C-terminal linker and the ACP domain) was solved to 2.14 Å resolution (Figure 5-3) (Bretschneider et al., 2013).

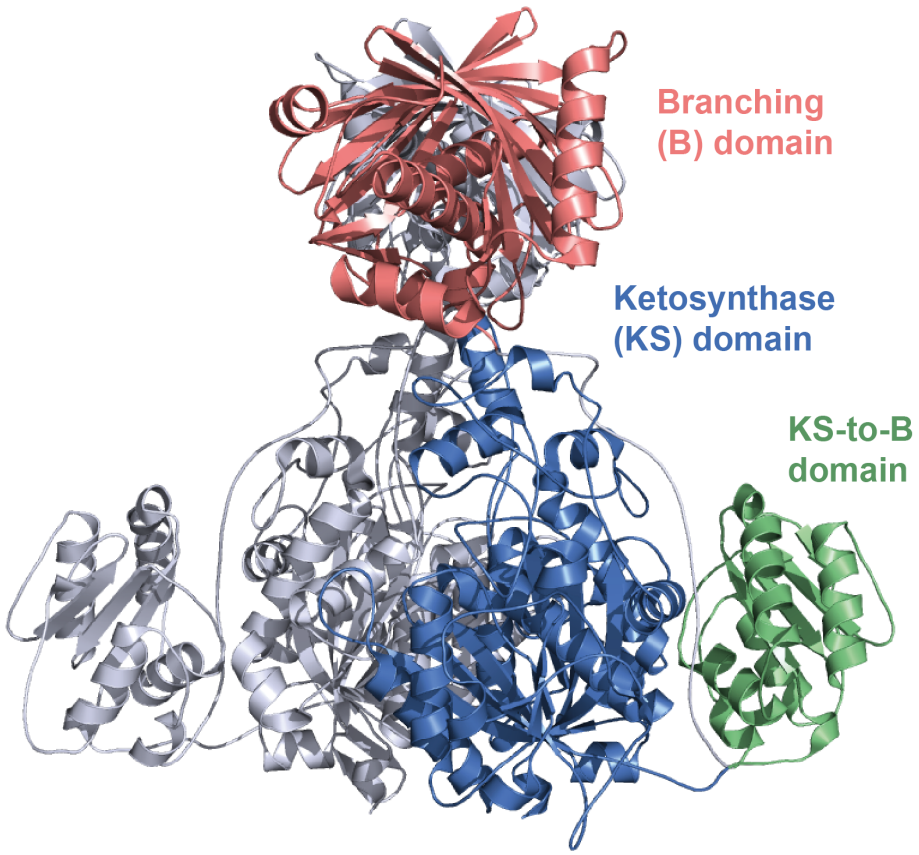


Figure 5-3 Crystal structure of the KS-B didomain from rhizoxin PKS (Bretschneider et al., 2013). *KS-B didomain dimer is shown, where chain A is shown in grey, and chain B is coloured according to the three folded domains it comprises. The ketosynthase domain is shown in blue, the KS-to-B linker domain is shown in green, and the branching domain is shown in red. The figure was created using PDB 4KC5.*

The crystal structure of the KS-B didomain revealed a homo-dimeric assembly comprising three distinct domains, ketosynthase, branching domain and KS-to-B linker domain (Figure 5-3). The KS domain appeared to have a typical thiolase fold (Sheryl Tsai and Ames, 2009) while the B domain, despite having less than 15% sequence homology to any known protein, exhibited a double hot dog motif, which has previously been observed in dehydratase domains (Maier et al., 2008).

To investigate which module, KS or B, was performing the branching reaction, site-directed mutagenesis of the active site residues of both domains was

performed. The results revealed that the novel polyketide branching reaction was performed by the KS domain; however this domain was only catalytically active when fused to the B domain (Bretschneider et al., 2013). This suggested that the B domain must play a crucial modulating role in catalysis.

As mentioned previously, in order to obtain crystals of the KS-B didomain, a large C-terminal part of the polypeptide, comprising the C-terminal acyl carrier protein and a 44 amino acid B-to-ACP linker, was removed. In addition, cross-linking experiments with a deoxy substrate analogue, between the KS-B didomain and a standalone ACP of the same module, revealed a band shift in the SDS-PAGE gel, confirming that the ACP is also needed for the branching catalysis to occur (Bretschneider et al., 2013). Therefore, the crystal structure of the KS-B didomain does not represent a fully functional branching module from the rhizoxin PKS.

5.1.4 Aim of this study

Significant efforts were spent by Dr Georg Zocher (University of Tübingen) attempting to determine the crystal structure of the fully functional branching module involving the flexible ACPs. A few crystals diffracting to ~ 20 Å resolution were obtained; however, no further progress was achieved (personal communication with Georg Zocher).

In recent years, the field of cryo-EM has seen the resolution revolution, with 831 cryo-EM maps of better than 4 Å resolution having been deposited in the EMDB since 2013, compared with just eighteen in the preceding ten years (Drulyte et al., 2018). This astonishing achievement is a result of the advancements in the direct electron detection technology, improved microscope hardware and innovations in image-processing (McMullan et al., 2016; Dutta, 2018).

Since the full-length KSB-ACP branching module appears to be too flexible for the crystallographic approach to be feasible, we envisioned that it would be a suitable target to be studied by cryo-EM. Our collaborators from the University of Tübingen and HKI-Jena prepared a sample of the full-length functional KSB-ACP module, in which the ACP is linked to the ketosynthase domain via a deoxy substrate analogue (Figure 5-2). This sample was termed KSB-ACP*S (*S denoting the linking via a substrate analogue). Our goal, therefore, was to determine the structure of the KSB-ACP*S.

This chapter describes the steps taken to investigate the structure of the full-length KSB-ACP*S module by electron microscopy techniques.

5.2 Negative stain EM of the KSB-ACP*S module

To evaluate the purity and homogeneity of the KSB-ACP*S sample, sent to us by our collaborators from HKI-Jena, we performed preliminary analysis by negative stain EM. The sample, diluted to 1 µg/ml, was applied to a glow-discharged amorphous carbon grid and stained with 1% uranyl acetate solution. A number of areas on the grid were imaged which revealed mono-dispersed particles with a length of approximately 10 nm (Figure 5-4a).

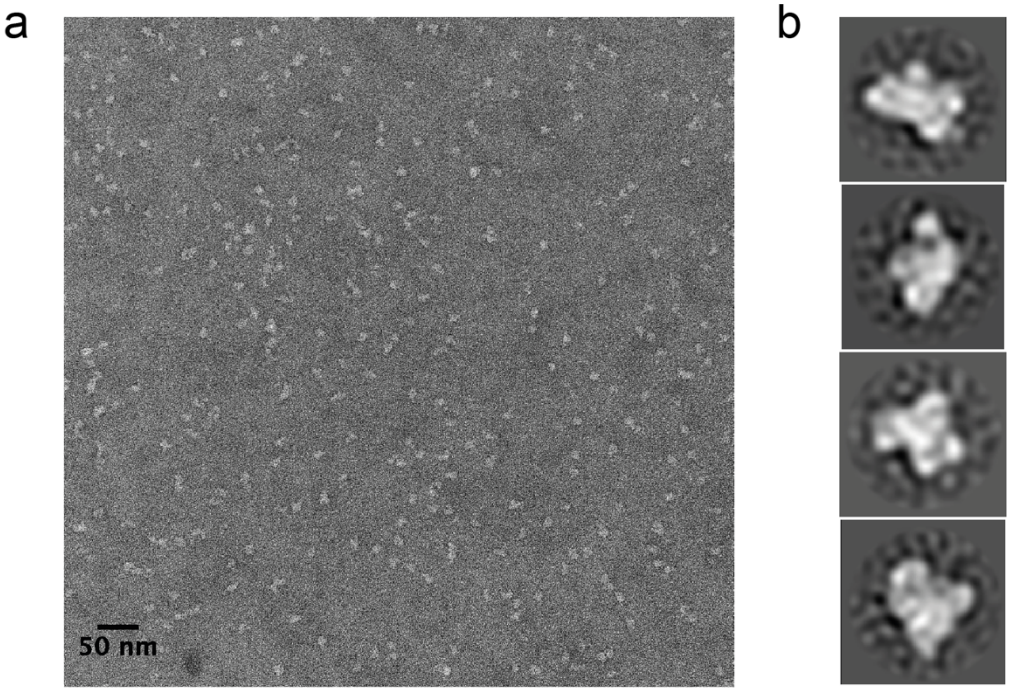


Figure 5-4 Negative stain EM of KSB-ACP*S. *(a) Raw micrograph with KSB-ACP*S molecules embedded in 1% uranyl acetate stain. The image was recorded using a magnification of 50,000x and electron acceleration of 200 keV. The scale bar is 50 nm. (b) 2D class averages of KSB-ACP*S. Parameters used in this data processing were as follows: 128-pixel box size (which equates to 256 Å at a pixel size of 2 Å/pix) and 220 Å diameter mask.*

Collection of 150 micrographs, picking of approximately 25,000 individual particles and subsequent 2D classification revealed star- and cross-shaped class averages shown in Figure 5-4b. Upon comparison to the KS-B didomain crystal structure projections generated *in silico* (not shown), it was decided that the

sample exhibited enough homogeneity and purity to commence the cryo-EM analysis.

5.3 Cryo-EM structure of the full-length rhizoxin PKS branching module

Cryo-EM of the branching module, KSB-ACP*S (where the *S denotes cross-linking between KS active site and the ACP via a deoxy substrate analogue), displayed good ice thickness, with particles being visible in holes with thinner ice (Figure 5-5a) and having a distinct cross shape (Figure 5-6a).

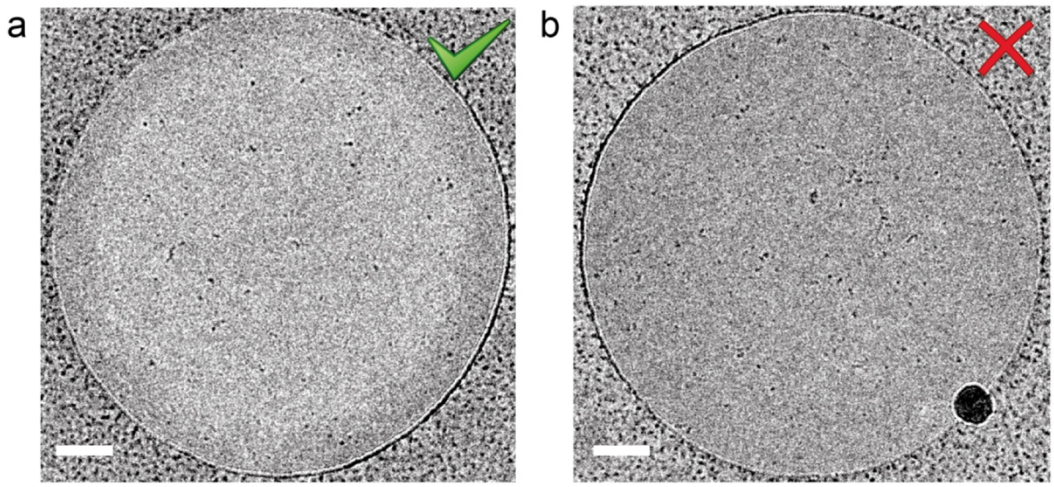


Figure 5-5 Comparison of ice thickness on cryo-EM grids. *Both images were taken at 11,000 magnification and show one full Quantifoil R1.2/1.3 hole. (a) An example of thin ice. Ice in the middle of the hole is thinner and hence appears lighter than near the edge of the whole. (b) An example of thick ice. Ice is the same shade throughout the hole. During this project, KSB-ACP*S particles could only be observed in thin ice (a). Scale bars represent 250 nm.*

4267 micrographs were acquired in a semi-automated fashion using the EPU software, and particles were auto-picked in Relion 2.0 (Scheres, 2012) which resulted in a dataset of approximately 1.6 million particles. The 2D classification was performed on the whole dataset, and a selection of the resulting 2D class averages is shown in Figure 5-6b. Details of the secondary structure, indicative of high-resolution data, could be seen in the images.

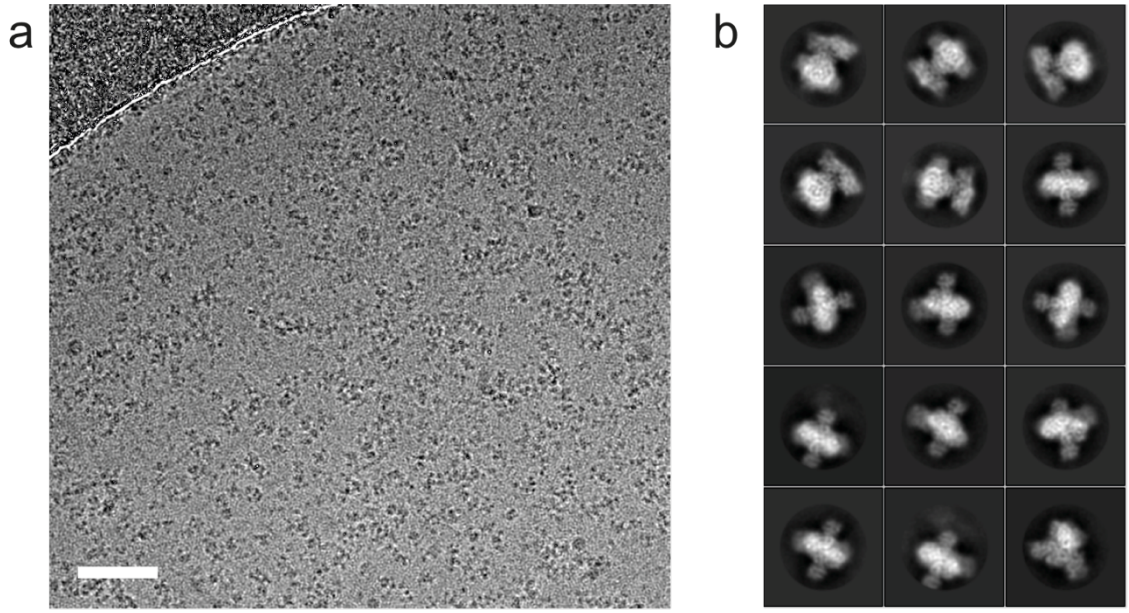


Figure 5-6 Initial cryo-EM analysis of the KSB-ACP*S. *(a) Raw cryo-EM micrograph of the KSB-ACP*S taken at 75,000 magnification at 300 keV. Proteinaceous species with a distinctive cross shape can be seen throughout. A total of 4,267 micrographs of the KSB-ACP*S were collected for this study. Scale bar represents 50 nm. (b) A selection of 2D class averages of the KSB-ACP*S generated from the ~ 1.6 million particle dataset. Parameters used in this data processing were as follows: 200-pixel box size (which equates to 213 Å at a pixel size of 1.065 Å/pix) and 150 Å diameter mask.*

Despite the promising 2D class averages, due to strong preferred orientation of the particles, the initial reconstruction of 2D classification-polished data, performed as described in the methods section 2.7.5, resulted in a poor-quality reconstruction exhibiting density “smearing” (Figure 5-7). To overcome this problem, one round of 3D classification was performed, which allowed us to obtain a subset of particles with full coverage of all Fourier components (Figure 5-8b). Particles belonging to the class with the highest resolution were used to generate the final cryo-EM reconstruction of KSB-ACP*S (Figure 5-8d) with a global resolution of 3.7 Å (Figure 5-8a).

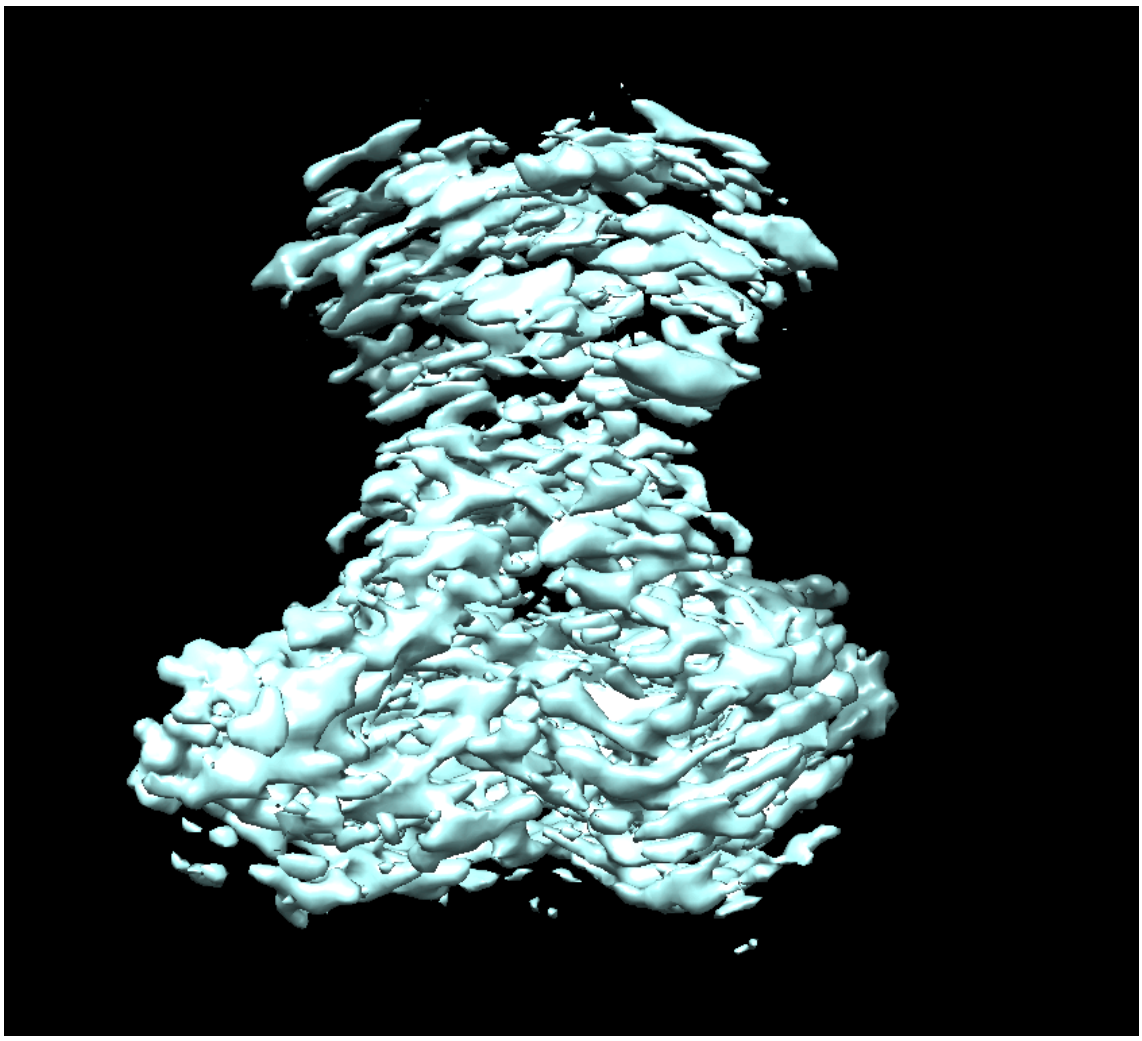


Figure 5-7 Cryo-EM map of the KSB-ACP*S with density "smearing" as a result of the sample exhibiting preferred orientation. *When the sample displays a preferential orientation, some Fourier components become oversampled and some under sampled which in turn results in density artefacts in the final reconstruction.*

The resulting map showed clear and continuous density for the KS and B domain protein backbone and was in good agreement with the previously reported KS-B didomain crystal structure (Bretschneider et al., 2013). The resolution of the KS core extended to 3.5 Å with side chains being visible (Figure 5-8c), while the B and KS-B linker domains appeared more flexible (Figure 5-8d). Surprisingly, no density indicating the bound ACP was visible adjacent to the KS active site.

Since this module was shown to perform the branching reaction *in vivo*, it is difficult to explain why an intramolecularly cross-linked ACP was not imaged in action, i.e. bound at the KS active site. We hypothesised that the deoxy substrate analogue molecule gets hydrolysed over time and therefore, the percentage of cross-linked particles in the sample on the grid might be low.

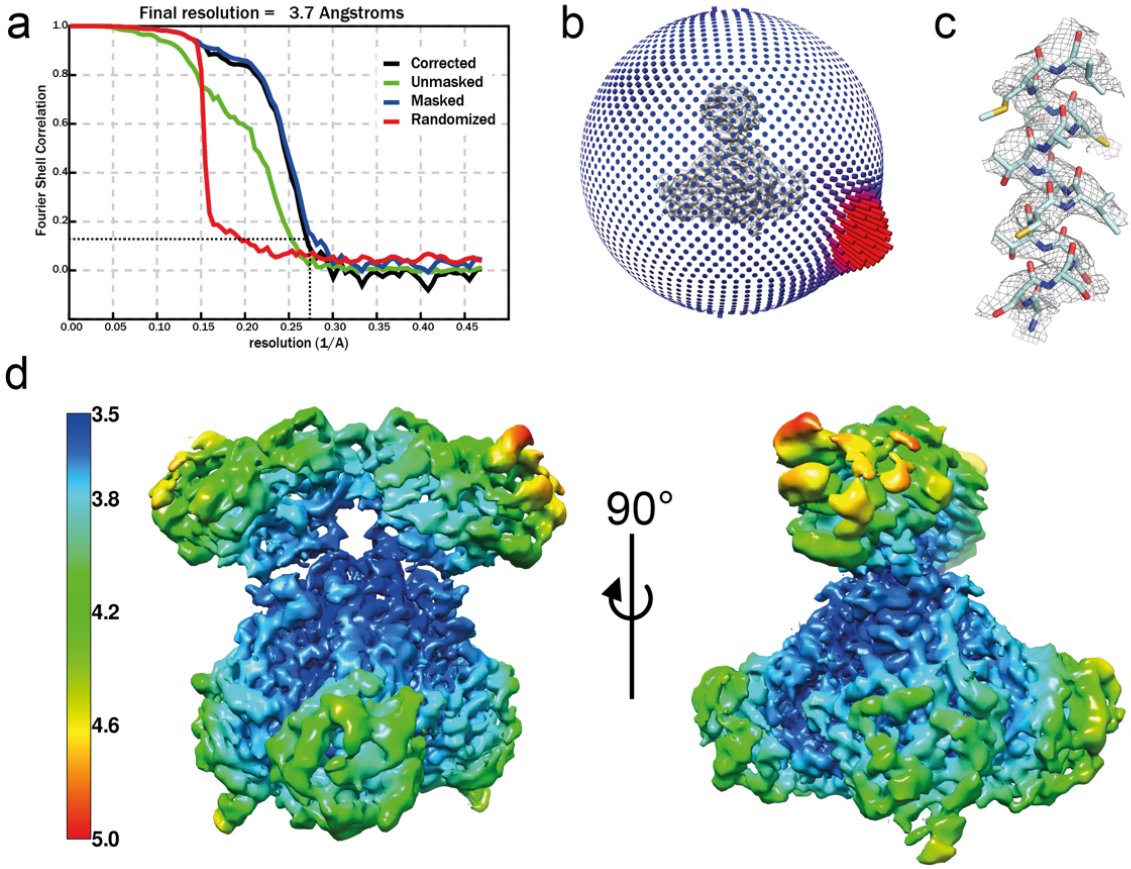


Figure 5-8 Cryo-EM reconstruction of the KSB-ACP*S complex. *(a) The FSC curve for the overall map of KSB-ACP*S resulting from gold-standard refinement. The global resolution of the map was calculated to be 3.7 Å. (b) A 3D angular distribution plot based on the alignment used for the final reconstruction. A clear preferred orientation can be seen; however, the full coverage of all other possible angles prevented the density artefacts. (c) A representative α-helix from the core of the KS domain rigid-body fitted with the KS-B atomic model (PDB 4KC5) (Bretschneider et al., 2013). (d) Cryo-EM reconstruction of the KSB-ACP*S module. The local resolution was estimated using RELION 2.0. Most of the ketosynthase core is resolved at 3.5 - 3.6 Å while the rest of the complex is resolved at 3.8 - 5.0 Å. Maps shown at 4.7 rmsd.*

5.4 Potential ACP binding sites as revealed by further 3D classification and cross-linking mass spectrometry

Traditionally, 3D classification is used to sort the data into subsets of particles exhibiting alternative conformations. However, 3D classification can also be used to improve the quality of the final reconstruction. This is done by performing the 3D refinement first and then running 3D classification without the alignment of particles (i.e. using particle coordinates from the refinement job rather than performing angular sampling again). In this study, 3D classification both with and

without the alignment was performed, and all data processing details are outlined in Figure 5-9 and methods section 2.7.6.

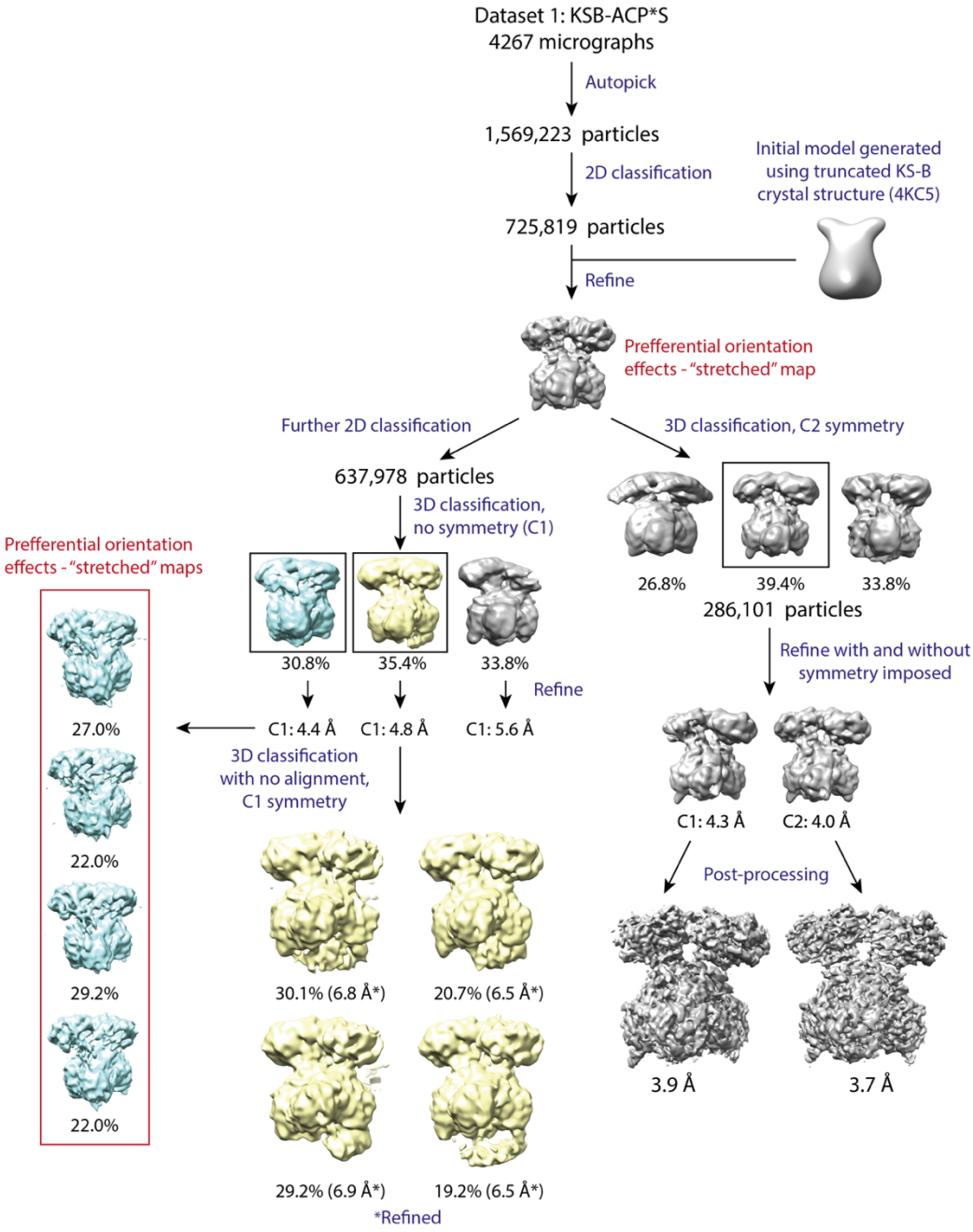


Figure 5-9 KSB-ACP*S data processing strategy used in this study. *All processing was performed with RELION 2.0. Two-fold symmetry was either imposed (C2) or not (C1). For 3D classification with no alignment, particle positions from 3D refinement were used and angular sampling was turned off.*

To investigate the possibility of ACP binding only in a small subset of particles, a second round of 2D classification followed by 3D classification with no symmetry operators were carried out. Interestingly, two out of the three resulting classes

(shown in blue and yellow in Figure 5-9) appeared to have additional density either above or below the KS domain suggesting the presence of two possible KS-ACP interaction sites.

Particles from each of the two classes (coloured blue and yellow in Figure 5-9) were further classified with no alignment. Classification of the first conformation (shown in blue) led to four identical classes all exhibiting density smearing artefacts (with all four classes highlighted with a red box in Figure 5-9). The classification of the second conformation (shown in yellow) resulted in an additional four distinguishable classes, each displaying slightly differently sized and positioned extra density underneath the KS dimerisation interface (Figure 5-9). These results suggest that, if the ACP is indeed binding to the KS at this location, the intrinsic flexibility of the PKS systems remains.

To see whether the additional density was big enough to accommodate a single copy of the ACP, a homology model of the branching module ACP was generated using the Phyre2 server (Kelley et al., 2015). The ACP was then fitted into both 3D classes exhibiting the extra density, which revealed that the density was of approximately the right size and shape to accommodate one copy of the branching module ACP (Figure 5-10).

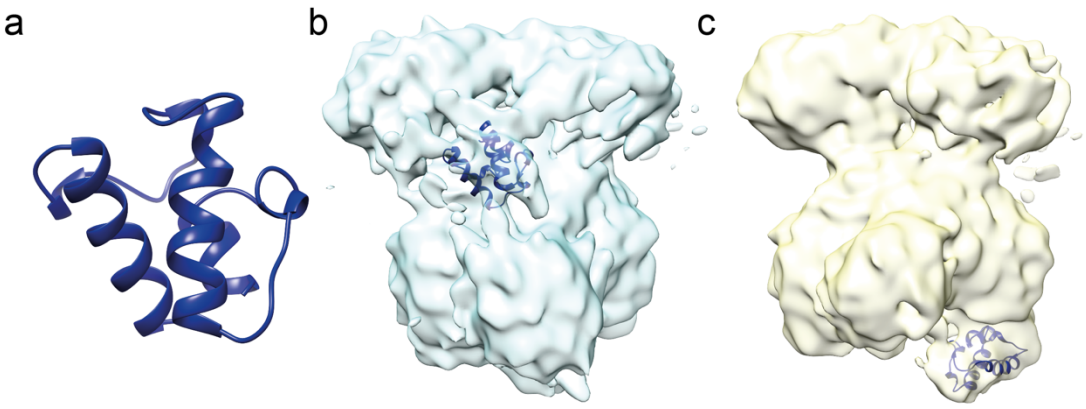


Figure 5-10 Fitting of the branching module ACP into the KSB-ACP*S 3D class averages exhibiting the extra density. *(a) Structural model of the branching module ACP, generated using Phyre2 server (Kelley et al., 2015). (b) ACP fitted into the extra density observed between the KS and B domains. (c) ACP fitted in the extra density observed underneath the KS domain.*

To validate these observations, two approaches were taken. First, a cryo-EM reconstruction of the KS-B didomain was determined as a control (described in more detail in the next section). Second, to probe the interactions between the ACP and the KS, our collaborators at HKI-Jena performed chemical cross-linking combined with LC/MS (Schmidt and Robinson, 2014). In this approach, KSB-ACP*S was cross-linked with a photoreactive crosslinker, fragmented using the protease trypsin, and the resulting peptides were analysed using LC/MS. A number of cross-links, including some between the ACP and the top and bottom sides of the KS domain, were identified, which were in agreement with the cryo-EM data (data not shown).

5.5 Cryo-EM structure of the KS-B didomain

To eliminate the possibility of the extra density being an artefact of image processing, a protein construct lacking the covalently-attached ACP altogether (and identical to the construct used to obtain the crystal structure) was investigated using cryo-EM. Data collection and image processing were identical to the KSB-ACP*S with the only difference being the size of the dataset (more details can be found in methods section 2.7.7). A total of approximately 0.8 million particles were auto-picked and polished by 2D and 3D classification algorithms. The resulting EM density map was comprised of approximately 100 thousand particles and displayed a global resolution of 4.4 Å (with C2 symmetry applied) (Figure 5-11).

In-depth 3D classification, with and without alignment and symmetry was performed identically to the KSB-ACP*S (Figure 5-9). None of the resulting 3D classes displayed any additional features like those observed for the KSB-ACP*S. What is more, the final KS-B map was in excellent agreement with the KS-B crystal structure (Bretschneider et al., 2013), providing us with further evidence that our method of generating KSB-ACP*S and KS-B maps was not introducing imaging artefacts, and that the additional density observed in the KSB-ACP*S reconstruction was accurate.

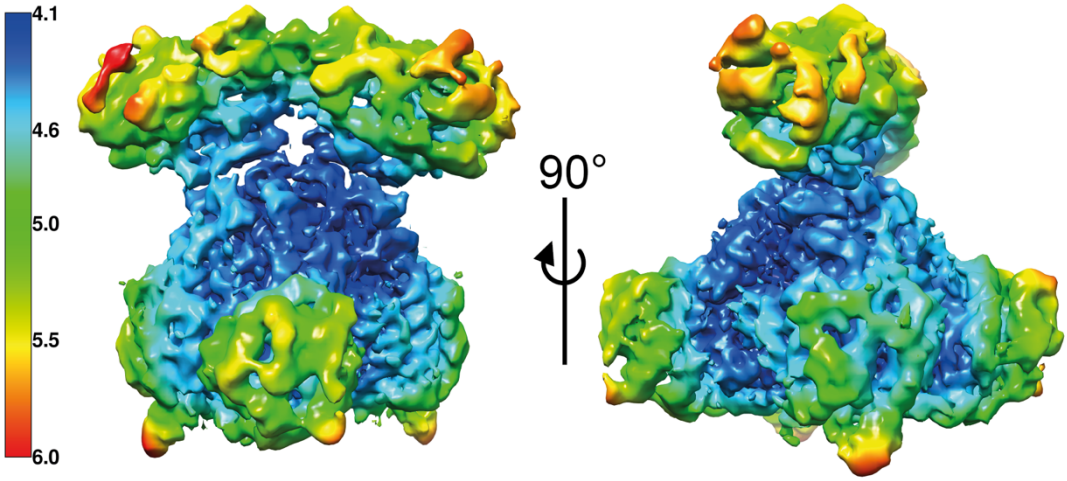


Figure 5-11 Cryo-EM structure of the KS-B didomain with a global resolution of 4.4 Å. *The local resolution was estimated using RELION 2.0. Most of the KS core is resolved at 4.1 - 4.5 Å while the rest of the complex is resolved at 4.6 - 6.0 Å. Maps shown at 4.1 rmsd.*

5.6 Summary

This chapter describes the steps taken to investigate the structure and function of the functional, full-length branching module from rhizoxin biosynthetic pathway (Partida-Martinez and Hertweck, 2007).

During previous work by our collaborators, a novel type of PKS module, responsible for introducing a short branch in rhizoxin macrolide ring, was identified (Figure 5-2) (Kusebauch et al., 2009). The module was studied in great detail, the branching reaction was reconstituted *in vitro*, and a crystal structure of the truncated module containing ketosynthase and branching domains (KS-B) was solved (Bretschneider et al., 2013).

We wanted to investigate the structure of the full-length module containing the ACP. To reduce the flexibility of the sample and to image the complex in action, the ACP was linked to the KS active site via a deoxy substrate analogue reported previously (Bretschneider et al., 2013) to produce a sample termed KSB-ACP*S.

The KSB-ACP*S module, as well as the KS-B didomain, as a control, were analysed by cryo-EM. The analysis led to two EM reconstructions with the global resolution of 3.7 Å and 4.4 Å, respectively. To our surprise, the initial 3D reconstruction of the KSB-ACP*S (Figure 5-8) did not reveal any differences from

either the previously reported crystal structure (Figure 5-3) or the cryo-EM reconstruction of the KS-B didomain (Figure 5-11).

We hypothesised that the ACP was only bound in a small subset of all particles. To test this, further *in silico* analysis, in the form of 2D classification and 3D classification with no angular sampling, was performed. This led to the identification of two potential binding sites for the ACP (Figure 5-9 and Figure 5-10). These observations were further supported by the cross-linking mass spectrometry and KS-B didomain cryo-EM analysis (sections 5.4 and 5.5).

The results of this cryo-EM project are impressive due to a relatively small size of the specimen (~230 kDa), low symmetry (C2) and relatively high resolution (3.7 Å and 4.4 Å) achieved. However, the two new ACP binding sites identified are not located near the conserved active site of the KS domain, raising further questions. Mainly, what is the reason for the ACP to bind away from the active site – perhaps there is a secondary active site present in this branching KS? In-depth biochemical studies by chemical cross-linking and mutagenesis could provide a better explanation for these findings. There is also a possibility that the results presented here are a product of promiscuous linking via the substrate analogue. To test for the latter, cryo-EM analysis of the KSB-ACP (where the ACP is not linked via a substrate analogue) should be performed.

The KSB-ACP*S map described here is the first, to our knowledge, example of a full-length PKS module resolved to better than 5 Å resolution. What is more, similarly to the MAS-like polyketide synthase (Herbst et al., 2018), our reconstruction revealed 1:2 stoichiometry with ACP bound to only one of the two monomers within a KS-B didomain dimer. It is, therefore, possible that the binding of one ACP to the KS prevents the binding of the second ACP.

Together with the biochemical data, the results presented in this chapter show the potential of cryo-EM to be used to study the structure and function of polyketide synthases. Studies like this are essential before rational engineering of these fascinating molecular machines can become routine.

Chapter 6 Summary, future perspectives and concluding remarks

6.1 Summary

This thesis has described the structural and functional studies of biosynthetic enzymes involved in natural product indanomycin and rhizoxin biosynthesis.

Chapter 1 introduced polyketide natural products and the use of their derivatives in modern medicine (Moloney, 2016; Newman and Cragg, 2012; Beutler, 2009), as well as a vast foundation of research into the discovery and engineering of the complex biosynthetic machinery required to produce them (Staunton and Weissman, 2001; Weissman, 2015a; Poust et al., 2014). It is now known that type I PKS assembly lines are divided into modules (Caffrey et al., 1992), with individual components of such modules having been studied in great detail (Robbins et al., 2016). Significant efforts have been placed over the last thirty years to generate the diversity in the existing polyketide libraries, both by synthetic and biological approaches (Weissman and Leadlay, 2005). Rational engineering of PKSs has resulted in some successes (Del Vecchio et al., 2003; Cortes et al., 1995; Rowe et al., 2001); however, the lack of high-resolution structural information of the intact PKS complexes remains the major obstacle before successful generation of polyketide analogues can become routine (Weissman, 2015b). Besides producing functional modular assembly lines, synthetic biologists also need a large “toolkit” of well-characterised tailoring enzymes which are capable of further altering the functionality of polyketide analogues (Olano et al., 2010).

The first results chapter focused on one such tailoring enzyme, IdmH, which has been proposed to catalyse a [4+2] cycloaddition reaction during the indane ring formation of indanomycin (C. Li et al., 2009; Rommel et al., 2011). The structural and functional characterisation of this putative polyketide cyclase began by cloning of the *idmH* gene, heterologous expression in *E. coli* and purification by standard chromatography techniques. Difficulties in crystallisation were overcome by producing an IdmH variant lacking residues 96-104 which were predicted to form a flexible loop. This has led to the determination of two crystal structures, one of wild-type IdmH at 2.7 Å and one of IdmH-Δ96-104 at 2.0 Å resolution. Both crystal structures of IdmH revealed homo-dimeric assembly

comprising two α + β barrels, with a large hydrophobic cavity located in each of the two barrels. We proposed this cavity to constitute the active site of the enzyme. To test whether proposed active site residues bind indanomycin product, NMR spectroscopy methods were applied. Specifically, the backbone resonance assignment of approximately 88% of IdmH residues was carried out, which was then followed by titration experiments and mapping of chemical shift perturbations. The results of the NMR experiments, as well as the *in-silico* docking and QM/MM simulations, performed by Dr Marc van Der Kamp from the University of Bristol, were shown to be consistent with IdmH acting as a Diels-Alderase to catalyse the formation of the indane ring in indanomycin.

The second results chapter described the successes and difficulties of performing structural studies on a fully-reducing module IdmO from indanomycin NRPS/PKS (C. Li et al., 2009). IdmO was successfully cloned from *S. antibioticus* genomic DNA; however, complications were encountered when soluble expression and purification of this large (~230 kDa) recombinant protein in sufficient quantities for structural studies was attempted. These issues were overcome by re-designing the construct and utilising an engineered *E. coli* strain Nico21(DE3) (Robichon et al., 2011). Subsequently, the catalytic activity of the recombinant IdmO was confirmed by a coupled αKGDH assay; however, specificity engineering attempts stalled due to the IdmO complex losing its activity too quickly. Finally, attempting structural studies by electron microscopy revealed further issues with the heterogeneity and poor stability of the sample when embedded in the thin layer of vitreous ice on a cryo-EM grid. Unfortunately, none of the measures taken to increase the stability of IdmO rectified the problem.

The third, final, results chapter focused on type I modular *trans*-AT PKS responsible for the production of rhizoxin natural product (Bretschneider et al., 2013). Rhizoxin's antimitotic activity depends on a short (C2) branch (Kusebauch et al., 2009), which is inserted into the growing polyketide by a novel type of branching module (Bretschneider et al., 2013). The purpose of this work was to investigate the structure of this novel module by cryo-EM and determine the structural basis for the branching reaction. Cryo-EM analysis yielded two high-resolution maps, one of the full branching module and one of the ketosynthase-branching didomain, at 3.7 Å and 4.5 Å resolution, respectively. The cryo-EM density map of the full complex revealed the ACP bound in two previously

unobserved locations. These results provided the basis for further analysis by biochemical techniques, such as mass spectrometry, and highlighted the potential of cryo-EM to be used to gain a better understanding of polyketide biosynthesis.

6.2 Continuation of the project and future perspectives

6.2.1 IdmH

While the data from the first results chapter strongly suggests that the IdmH acts as a Diels-Alderase to catalyse the formation of the indane ring in indanomycin, further analysis needs to be carried out to definitively characterise the function of this protein. The most obvious strategy to achieve this would be to synthesise a synthetic indanomycin substrate **4** and follow the conversion of **4** to indanomycin **1** in the presence and the absence of IdmH.

If IdmH acts similarly to other Diels-Alderases characterised to date and provides a template for the molecule to adopt a favorable orientation, it may also be able to catalyse the [4+2] cycloaddition in other natural products. Therefore, since the organic synthesis of **4** would be very challenging, it might be sensible to utilise a simpler strategy first, such as testing other known Diels-Alderase substrates, such as those used to study AbyU (Byrne et al., 2016). Finally, calculating the NMR structures for the free- and indanomycin bound-IdmH would provide further details into the catalytic mechanism of IdmH.

Following the successful functional characterisation of IdmH (provided the IdmH is indeed a Diels-Alderase), future work could involve the implementation of this enzyme in other polyketide biosynthetic pathways to generate novel polyketides. The Diels-Alder reaction, despite being widely used in synthetic organic chemistry, is an extremely rare reaction in nature, which is why it would also be advantageous to investigate whether IdmH could be utilised for synthetic cycloaddition reactions.

6.2.2 IdmO

Negative stain EM analysis revealed a high level of heterogeneity within the IdmO sample. While IdmO is expected to be flexible in solution, eliminating disassociated monomers from the sample would aid the downstream processing efforts to produce a more detailed negative-stain EM reconstruction. This could

be attempted by performing preparative SEC-MALLS in different buffers, collecting the resulting fractions and making grids from different parts of the elution peak. A similar approach has been utilised to analyse a MAS-like PKS module successfully (Herbst et al., 2018).

Cryo-EM analysis of IdmO was not possible due to strong denaturation of IdmO at the air-water interface. A number of measures (Drulyte et al., 2018), such as changing the buffer composition or trialling different types of grids were attempted, but with little success. Going forward, chemical cross-linking using ultracentrifugation appears to be the most promising approach to tackle low stability of IdmO (Stark, 2010), as exemplified by the successful use of this technique with other challenging samples (Nguyen et al., 2018; Poepsel et al., 2018; Kasinath et al., 2018).

Structural studies of intact PKS systems is of paramount importance and therefore following the successful cryo-EM characterisation of IdmO, the analysis containing multiple modules should be carried out. To start, investigating the interactions between the IdmO and the ACPs from the adjacent modules would provide an initial understanding of how the polyketide chain is transferred along this assembly line. Going from there, multiple modules could be expressed and purified individually and fused together using either native or synthetic docking domains (Broadhurst et al., 2003; Zheng et al., 2013) before performing a similar analysis by electron microscopy. Successful execution of such a project would provide invaluable insights into the intermodule communications, which is crucial for the successful engineering of functional polyketide assembly lines.

6.2.3 RhiE

Cryo-EM analysis of KSB-ACP*S revealed intriguing new details of two possible ACP binding sites; however, they were not located near the KS active site residues. To depict the mechanism of this module and to explain the unusual ACP binding, further characterisation of this module by cross-linking mass spectrometry could be carried out. Furthermore, it would be interesting to see whether the ACP would act similarly if the branching KS was replaced with an elongating KS from another module from the same biosynthetic assembly line. Finally, to investigate the interactions between the branching module and the

adjacent modules, multi-modular constructs could be made and analysed using the same single-particle cryo-EM pipeline.

The majority of polyketides are generated in linear head-to-tail fashion. Structural and functional characterisation of the branching module from the rhizoxin PKS represents the means to expand the structural space of current polyketide natural products. Identification of residues which convey the branching function upon a ketosynthase would allow switching an elongating KS into a branching KS using simple site-directed mutagenesis, instead of carrying out complex module swaps. Introducing one or several branching modules into existing polyketide synthases could generate further diversity in this class of natural products and create many natural product analogues which could be screened for novel or improved pharmacological properties.

6.3 Concluding remarks

This research has provided invaluable insights into different aspects of PKS biology ranging from polyketide tailoring and branching to the assembly of the intact modules. A number of difficulties were experienced when attempting the structural analysis of PKS polypeptides from the indanomycin NRPS/PKS. Nevertheless, structural and functional characterisation of a putative polyketide cyclase was successful. While cryo-EM analysis of IdmO was not successful, determination of optimum expression, purification and assay conditions for this PKS module provides a great starting point for further structural characterisation and specificity engineering. Finally, cryo-EM structures of the branching module from the rhizoxin PKS provide a strong basis for further characterisation of this novel type of branching module.

Taken together, all three PKS proteins studied during this work are important additions to the enzymatic toolbox of polyketide synthases and have great potential to be used to generate diversity in synthetically engineered biosynthetic assembly lines.

List of References

Aggarwal, R., Caffrey, P., Leadlay, P.F., Smith, C.J. and Staunton, J. 1995. The thioesterase of the erythromycin-producing polyketide synthase: mechanistic studies *in vitro* to investigate its mode of action and substrate specificity. *Journal of the Chemical Society, Chemical Communications*. **0**(15),pp.1519–1520.

Akey, D.L., Kittendorf, J.D., Giraldes, J.W., Fecik, R.A., Sherman, D.H. and Smith, J.L. 2006. Structural basis for macrolactonization by the pikromycin thioesterase. *Nat Chem Biol*. **2**(10),pp.537–542.

Akey, D.L., Razelun, J.R., Tehranisa, J., Sherman, D.H., Gerwick, W.H. and Smith, J.L. 2010. Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure*. **18**(1),pp.94–105.

Aldridge, S., Parascandola, J., Sturchio, J.L., Society, A.C. and Great Britain, R.S.O.C. 1999. *The Discovery and Development of Penicillin 1928-1945*.

Alekseyev, V.Y., Liu, C.W., Cane, D.E., Puglisi, J.D. and Khosla, C. 2007. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase*. Protein Science: A Publication of the Protein Society.* **16**(10),pp.2093–2107.

Ames, B.D., Nguyen, C., Bruegger, J., Smith, P., Xu, W., Ma, S., Wong, E., Wong, S., Xie, X. and Li, J.W.-H. 2012. Crystal Structure and Biochemical Studies of the *Trans*-Acting Polyketide Enoyl Reductase LovC from Lovastatin Biosynthesis. *Proceedings of the National Academy of Sciences*. **109**(28),pp.11144–11149.

Auclair, K., Sutherland, A., Kennedy, J., Witter, D.J., Van den Heever, J.P., Hutchinson, C.R. and Vederas, J.C. 2000. Lovastatin Nonaketide Synthase Catalyzes an Intramolecular Diels-Alder Reaction of a Substrate Analogue. *Journal of the American Chemical Society*. **122**(46),pp.11519–11520.

Bai, X.-C., McMullan, G. and Scheres, S.H.W. 2015. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences*. **40**(1),pp.49–57.

Beinker, P., Lohkamp, B., Peltonen, T., Niemi, J., Mäntsälä, P. and Schneider, G. 2006a. Crystal structures of SnoaL2 and AclR: two putative hydroxylases in the biosynthesis of aromatic polyketide antibiotics. *Journal of Molecular Biology*. **359**(3),pp.728–740.

Berman, H.M. 2000. The Protein Data Bank. *Nucleic Acids Research*. **28**(1),pp.235–242.

Beutler, J.A. 2009. Natural products as a foundation for drug discovery. *Current protocols in pharmacology*. **Chapter 9**(1),pp.Unit 9.11–9.11.21.

Bhattacharyya, M., Stratton, M.M., Going, C.C., McSpadden, E.D., Huang, Y., Susa, A.C., Elleman, A., Cao, Y.M., Pappireddi, N., Burkhardt, P., Gee, C.L., Barros, T., Schulman, H., Williams, E.R. and Kuriyan, J. 2016. Molecular mechanism of activation-triggered subunit exchange in Ca(2+)/calmodulin-dependent protein kinase II. *eLife*. **5**,p.213.

Birch, A.J., Massy-Westropp, R.A. and Moye, C.J. 1955. Studies in relation to biosynthesis. VII. 2-Hydroxy-6-methylbenzoic acid in *Penicillium griseofulvum* Dierckx. *Australian Journal of Chemistry*. **8**(4),p.539.

Boddy, C.N., Hotta, K., Tse, M.L., Watts, R.E. and Khosla, C. 2004. Precursor-Directed Biosynthesis of Epothilone in *Escherichia coli. Journal of the American Chemical Society*. **126**(24),pp.7436–7437.

Bonnett, S.A., Rath, C.M., Shareef, A.-R., Joels, J.R., Chemler, J.A., Hakansson, K., Reynolds, K. and Sherman, D.H. 2011. Acyl-CoA Subunit Selectivity in the Pikromycin Polyketide Synthase PikAIV: Steady-State Kinetics and Active-Site Occupancy Analysis by FTICR-MS. *Chemistry & Biology*. **18**(9),pp.1075–1081.

Bowler, M.W., Nurizzo, D., Barrett, R., Beteva, A., Bodin, M., Caserotto, H., Delagenière, S., Dobias, F., Flot, D., Giraud, T., Guichard, N., Guijarro, M., Lentini, M., Leonard, G.A., McSweeney, S., Oskarsson, M., Schmidt, W., Snigirev, A., Stetten, Von, D., Surr, J., Svensson, O., Theveneau, P. and Mueller-Dieckmann, C. 2015. MASSIF-1: A beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *Journal of Synchrotron Radiation*. **22**(6),pp.1540–1547.

Bowler, M.W., Svensson, O. and Nurizzo, D. 2016. Fully automatic macromolecular crystallography: the impact of MASSIF-1 on the optimum acquisition and quality of data. *Crystallography Reviews*. **22**(4),pp.233–249.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*. **72**,pp.248–254.

Breazeale, S.D., Ribeiro, A.A. and Raetz, C.R. 2002. Oxidative Decarboxylation of UDP-Glucuronic Acid in Extracts of Polymyxin-Resistant *Escherichia coli* Origin of Lipid a Species Modified with 4-Amino-4-Deoxy-L-Arabinose. *J Biol Chem*. **277**(4),pp.2886–2896.

Bretschneider, T., Heim, J.B., Heine, D., Winkler, R., Busch, B., Kusebauch, B., Stehle, T., Zocher, G. and Hertweck, C. 2013. Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature*. **502**(7469),pp.124–128.

Brignole, E.J., Smith, S. and Asturias, F.J. 2009. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nature Structural & Molecular Biology*. **16**(2),pp.190–197.

Brimble, M.A. 2000. Polyether Antibiotics *In*: *Kirk-Othmer Encyclopedia of Chemical Technology*. John Wiley & Sons, Inc.

Broadhurst, R.W., Nietlispach, D., Wheatcroft, M.P., Leadlay, P.F. and Weissman, K.J. 2003. The Structure of Docking Domains in Modular Polyketide Synthases. *Chemistry & Biology*. **10**(8),pp.723–731.

Bukhari, H.S.T., Jakob, R.P. and Maier, T. 2014. Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase. *Structure*. **22**(12),pp.1775–1785.

Bulkley, D., Innis, C.A., Blaha, G. and Steitz, T.A. 2010. Revisiting the structures of several antibiotics bound to the bacterial ribosome. *Proc Natl Acad Sci U S A*. **107**(40),pp.17158–17163.

Byrne, M.J., Lees, N.R., Han, L.C., van der Kamp, M.W., Mulholland, A.J., Stach, J.E.M., Willis, C.L. and Race, P.R. 2016. The Catalytic Mechanism of a Natural Diels-Alderase Revealed in Molecular Detail. *Journal of the American Chemical Society*. **138**(19),pp.6095–6098.

Caffrey, P., Bevitt, D.J., Staunton, J. and Leadlay, P.F. 1992. Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from *Saccharopolyspora erythraea*. *FEBS letters*. **304**(2-3),pp.225–228.

Calderone, C.T. 2008. Isoprenoid-like alkylations in polyketide biosynthesis. *Nat Prod Rep*. **25**(5),pp.845–853.

Carey, F.A. and Sundberg, R.J. 2013. *Part B: Reactions and Synthesis*. Springer.

Castonguay, R., He, W., Chen, A.Y., Khosla, C. and Cane, D.E. 2007. Stereospecificity of Ketoreductase Domains of the 6-Deoxyerythronolide B Synthase. *Journal of the American Chemical Society*. **129**(44),pp.13758–13769.

Chaillan-Huntington, C., Butler, P.J.G., Huntington, J.A., Akin, D., Feldherr, C. and Stewart, M. 2001. NTF2 monomer-dimer equilibrium. Edited by B. Holland. *Journal of Molecular Biology*. **314**(3),pp.465–477.

Chakravarty, B., Gu, Z., Chirala, S.S., Wakil, S.J. and Quiocho, F.A. 2004. Human fatty acid synthase: Structure and substrate selectivity of the thioesterase domain. *Proceedings of the National Academy of Sciences*. **101**(44),pp.15567–15572.

Chan, Y.A. and Thomas, M.G. 2010. Recognition of (2 S)-Aminomalonyl-Acyl Carrier Protein (ACP) and (2 R)-Hydroxymalonyl-ACP by Acyltransferases in Zwittermicin A Biosynthesis. *Biochemistry*. **49**(17),pp.3667–3677.

Chandran, S.S., Menzella, H.G., Carney, J.R. and Santi, D.V. 2006. Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains. *Chemistry & Biology*. **13**(5),pp.469–474.

Charlop-Powers, Z., Owen, J.G., Reddy, B.V.B., Ternei, M.A. and Brady, S.F. 2014. Chemical-biogeographic survey of secondary metabolism in soil. *Proc Natl Acad Sci U S A*. **111**(10),pp.3757–3762.

Chen, A.Y., Schnarr, N.A., Kim, C.-Y., Cane, D.E. and Khosla, C. 2006. Extender unit and acyl carrier protein specificity of ketosynthase domains of the 6-deoxyerythronolide B synthase. *Journal of the American Chemical Society*. **128**(9),pp.3067–3074.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*. **66**(1),pp.12–21.

Cheng, Y. 2015. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell*. **161**(3),pp.450–457.

Cheng, Y.-Q., Tang, G.-L. and Ben Shen 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proceedings of the National Academy of Sciences*. **100**(6),pp.3149–3154.

Clarke, A.C. 1973. *Profiles of the Future: An Inquiry into the Limits of the Possible*.

Colegate, S.M. and Molyneux, R.J. 2007. *Bioactive Natural Products*. CRC Press.

Collie, N. and Myers, W.S. 1893. VII.—The formation of orcinol and other condensation products from dehydracetic acid. *J. Chem. Soc., Trans.* **63**(0),pp.122–128.

Compton, S.J. and Jones, C.G. 1985. Mechanism of dye response and interference in the Bradford protein assay. *Analytical Biochemistry*. **151**(2),pp.369–374.

Cortes, J., Wiesmann, K.E., Roberts, G.A., Brown, M.J., Staunton, J. and Leadlay, P.F. 1995. Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science*. **268**(5216),pp.1487–1489.

Cortés, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. and Leadlay, P.F. 1990. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature*. **348**(6297),pp.176–178.

Cowtan, K. 2012. Completion of autobuilt protein models using a database of protein fragments. *Acta Crystallographica Section D: Biological Crystallography*. **68**(Pt 4),pp.328–335.

Cowtan, K. 2010. Recent developments in classical density modification. *Acta Crystallographica Section D: Biological Crystallography*. **66**(Pt 4),pp.470–478.

Cowtan, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography*. **62**(Pt 9),pp.1002–1011.

Crump, M.P., Crosby, J., Dempsey, C.E., Parkinson, J.A., Murray, M., Hopwood, D.A. and Simpson, T.J. 1997. Solution structure of the actinorhodin polyketide synthase acyl carrier protein from Streptomyces coelicolor A3(2). *Biochemistry*. **36**(20),pp.6000–6008.

Cummings, M., Breitling, R. and Takano, E. 2014. Steps towards the synthetic biology of polyketide biosynthesis. *FEMS Microbiol Lett*. **351**(2),pp.116–125.

David, B., Wolfender, J.-L. and Dias, D.A. 2014. The pharmaceutical industry and natural products: historical status and new trends. *Phytochemistry Reviews*. **14**(2),pp.299–315.

Davison, J., Dorival, J., Rabeharindranto, H., Mazon, H., Chagot, B., Gruez, A. and Weissman, K.J. 2014. Insights into the function of *trans*-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chemical Science*. **5**(8),pp.3081–3095.

Del Vecchio, F., Petkovic, H., Kendrew, S.G., Low, L., Wilkinson, B., Lill, R., Cortés, J., Rudd, B.A.M., Staunton, J. and Leadlay, P.F. 2003. Active-site residue, domain and module swaps in modular polyketide synthases. *Journal of Industrial Microbiology and Biotechnology*. **30**(8),pp.489–494.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR*. **6**(3),pp.277–293.

Dias, D.A., Urban, S. and Roessner, U. 2012. A Historical Overview of Natural Products in Drug Discovery. *Metabolites*. **2**(2),pp.303–336.

Dickens, M.L., Ye, J. and Strohl, W.R. 1995. Analysis of clustered genes encoding both early and late steps in daunomycin biosynthesis by *Streptomyces* sp. strain C5. *Journal of Bacteriology*. **177**(3),pp.536–543.

Dickson, M. and Gagnon, J.P. 2004. Key factors in the rising cost of new drug discovery and development. *Nature Reviews Drug Discovery*. **3**(5),pp.417–429.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. and Katz, L. 1991. Modular organization of genes required for complex polyketide biosynthesis. *Science*. **252**(5006),pp.675–679.

Drozdetskiy, A., Cole, C., Procter, J. and Barton, G.J. 2015. JPred4: a Protein Secondary Structure Prediction Server. *Nucleic Acids Research*.

Drulyte, I., Johnson, R.M., Hesketh, E.L., Hurdiss, D.L., Scarff, C.A., Porav, S.A., Ranson, N.A., Muench, S.P. and Thompson, R.F. 2018. Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallographica Section D: Structural Biology*. **74**(Pt 6),pp.560–571.

Dunn, B.J. and Khosla, C. 2013. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *Journal of the Royal Society, Interface*. **10**(85),pp.20130297–20130297.

Dunn, B.J., Cane, D.E. and Khosla, C. 2013. Mechanism and Specificity of an Acyltransferase Domain from a Modular Polyketide Synthase. *Biochemistry*. **52**(11),pp.1839–1841.

Dunn, B.J., Watts, K.R., Robbins, T., Cane, D.E. and Khosla, C. 2014. Comparative analysis of the substrate specificity of *trans*- versus *cis*-acyltransferases of assembly line polyketide synthases. *Biochemistry*. **53**(23),pp.3796–3806.

Dutta, M. 2018. Recent Advances in Single Particle Cryo-electron Microscopy and Cryo-electron Tomography to Determine the Structures of Biological Macromolecules. *Journal of the Indian Institute of Science*. **98**(3),pp.231–245.

Dutta, S., Whicher, J.R., Hansen, D.A., Hale, W.A., Chemler, J.A., Congdon, G.R., Narayan, A.R.H., Hakansson, K., Sherman, D.H., Smith, J.L. and Skiniotis, G. 2014. Structure of a modular polyketide synthase. *Nature*. **510**(7506),pp.512–517.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server *In*: J. M. Walker, ed. *The Proteomics Protocols Handbook*. Humana Press, pp. 571–607.

Edwards, A.L., Matsui, T., Weiss, T.M. and Khosla, C. 2014. Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. *Journal of Molecular Biology*. **426**(11),pp.2229–2245.

Edwards, M.P., Ley, S.V., Lister, S.G., Palmer, B.D. and Williams, D.J. 1984. Total Synthesis of the Ionophore Antibiotic X-14547A (Indanomycin). *The Journal of Organic Chemistry*. **49**(19),pp.3503–3516.

Eichner, S., Floss, H.G., Sasse, F. and Kirschning, A. 2009. New, Highly Active Nonbenzoquinone Geldanamycin Derivatives by Using Mutasynthesis. *ChemBioChem*. **10**(11),pp.1801–1805.

Emsley, P. and Cowtan, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*. **60**(Pt 12 Pt 1),pp.2126–2132.

Engel, C., Plitzko, J., Cramer P. 2016. RNA polymerase I–Rrn3 complex at 4.8 Å resolution. *Nature Communications.* **7**,p.12129.

Etienne, G., Malaga, W., Laval, F., Lemassu, A., Guilhot, C. and Daffé, M. 2009. Identification of the Polyketide Synthase Involved in the Biosynthesis of the Surface-

Exposed Lipooligosaccharides in Mycobacteria. *Journal of Bacteriology*. **191**(8),pp.2613–2621.

Evans, S.E., Williams, C., Arthur, C.J., Burston, S.G., Simpson, T.J., Crosby, J. and Crump, M.P. 2008. An ACP structural switch: conformational differences between the apo and holo forms of the actinorhodin polyketide synthase acyl carrier protein. *ChemBioChem*. **9**(15),pp.2424–2432.

Evans, S.E., Williams, C., Arthur, C.J., Płoskoń, E., Wattana-Amorn, P., Cox, R.J., Crosby, J., Willis, C.L., Simpson, T.J. and Crump, M.P. 2009. Probing the Interactions of early polyketide intermediates with the Actinorhodin ACP from S. coelicolor A3(2). *Journal of Molecular Biology*. **389**(3),pp.511–528.

Fafarman, A.T., Sigala, P.A., Schwans, J.P., Fenn, T.D., Herschlag, D. and Boxer, S.G. 2012. Quantitative, directional measurement of electric field heterogeneity in the active site of ketosteroid isomerase. *Proceedings of the National Academy of Sciences*. **109**(6),pp.E299–E308.

Fage, C.D., Isiorho, E.A., Liu, Y., Wagner, D.T., Liu, H.-W. and Keatinge-Clay, A.T. 2015. The structure of SpnF, a standalone enzyme that catalyzes [4 + 2] cycloaddition. *Nat Chem Biol*. **11**(4),pp.256–258.

Farmer, B.T., Constantine, K.L., Goldfarb, V., Friedrichs, M.S., Wittekind, M., Yanchunas, J., Robertson, J.G. and Mueller, L. 1996. Localizing the NADP+ binding site on the MurB enzyme by NMR. *Nature structural biology*. **3**(12),pp.995–997.

Ferrandi, E.E., Sayer, C., Isupov, M.N., Annovazzi, C., Marchesi, C., Iacobone, G., Peng, X., Bonch-Osmolovskaya, E., Wohlgemuth, R., Littlechild, J.A. and Monti, D. 2015. Discovery and characterization of thermophilic limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries. *FEBS Journal*. **282**(15),pp.2879–2894.

Ferrer, J.L., Austin, M.B., Stewart C.Jr. Noel, J.P. 2008. Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiology and Biochemistry*. **46**(3),pp.356–370.

Ferrer, J.L., Jez, J.M., Bowman, M.E., Dixon, R.A. and Noel, J.P. 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature structural biology*. **6**(8),pp.775–784.

Fischbach, M.A. and Walsh, C.T. 2006. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chemical reviews*. **106**(8),pp.3468–3496.

Gaisser, S. and Leadlay, P.F. 1998. Sugaring the pill by design. *Nature Biotechnology*. **16**(1),pp.19–20.

Gatzeva-Topalova, P.Z., May, A.P. and Sousa, M.C. 2005. Structure and Mechanism of ArnA: Conformational Change Implies Ordered Dehydrogenase Mechanism in Key Enzyme for Polymyxin Resistance. *Structure*. **13**(6),pp.929–942.

Gavalda, S., Bardou, F., Laval, F., Bon, C., Malaga, W., Chalut, C., Guilhot, C., Mourey, L., Daffé, M. and Quémard, A. 2014. The polyketide synthase Pks13 catalyzes a novel mechanism of lipid transfer in mycobacteria. *Chemistry & Biology*. **21**(12),pp.1660–1669.

Gay, D., You, Y.-O., Keatinge-Clay, A. and Cane, D.E. 2013. Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase. *Biochemistry*. **52**(49),pp.8916–8928.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. 2009. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat Meth*. **6**(5),pp.343–345.

Glaeser, R.M. and Han, B.-G. 2017. Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophysics reports*. **3**(1),pp.1–7.

Glaeser, R.M., Han, B.-G., Csencsits, R., Killilea, A., Pulk, A. and Cate, J.H.D. 2016. Factors that Influence the Formation and Stability of Thin, Cryo-EM Specimens. *Biophysical Journal*. **110**(4),pp.749–755.

Goddard, T.D., Huang, C.C. and Ferrin, T.E. 2007. Visualizing density maps with UCSF Chimera. *Journal of Structural Biology*. **157**(1),pp.281–287.

Gokhale, R.S., Hunziker, D., Cane, D.E. and Khosla, C. 1999. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chemistry & Biology*. **6**(2),pp.117–125.

Goldschmidt, L., Cooper, D.R., Derewenda, Z.S. and Eisenberg, D. 2007. Toward Rational Protein Crystallization: A Web Server for the Design of Crystallizable Protein Variants. *Protein Science*. **16**(8),pp.1569–1576.

Goss, R.J. and Hong, H. 2005. A Novel Fluorinated Erythromycin Antibiotic. *Chemical Communications*. (31),pp.3983–3985.

Goss, R.J., Shankar, S. and Fayad, A.A. 2012. The Generation of 'Unnatural' Products: Synthetic Biology Meets Synthetic Chemistry. *Nat Prod Rep*. **29**(8),pp.870–889.

Green, M.R. and Sambrook, J. 2018. Touchdown Polymerase Chain Reaction (PCR). *Cold Spring Harbor protocols*. **2018**(5),p.pdb.prot095133.

Gregory, M.A., Hong, H., Lill, R.E., Gaisser, S., Petkovic, H., Low, L., Sheehan, L.S., Carletti, I., Ready, S.J., Ward, M.J., Kaja, A.L., Weston, A.J., Challis, I.R., Leadlay, P.F., Martin, C.J., Wilkinson, B. and Sheridan, R.M. 2006. Rapamycin biosynthesis: elucidation of gene product function. *Organic & Biomolecular Chemistry*. **4**(19),pp.3565–3568.

Gupta, S., Lakshmanan, V., Kim, B.S., Fecik, R. and Reynolds, K.A. 2008. Generation of Novel Pikromycin Antibiotic Products Through Mutasynthesis. *ChemBioChem*. **9**(10),pp.1609–1616.

Ha, N.C., Choi, G., Choi, K.Y. and Oh, B.H. 2001. Structure and enzymology of Delta5-3-ketosteroid isomerase. *Current Opinion in Structural Biology*. **11**(6),pp.674–678.

Haff, L.A. 1994. Improved Quantitative PCR Using Nested Primers. *PCR Methods Appl*. **3**(6),pp.332–337.

Haines, A.S., Dong, X., Song, Z., Farmer, R., Williams, C., Hothersall, J., Płoskoń, E., Wattana-Amorn, P., Stephens, E.R., Yamada, E., Gurney, R., Takebayashi, Y., Masschelein, J., Cox, R.J., Lavigne, R., Willis, C.L., Simpson, T.J., Crosby, J., Winn, P.J., Thomas, C.M. and Crump, M.P. 2013. A conserved motif flags acyl carrier proteins for β-branching in polyketide synthesis. *Nat Chem Biol*. **9**(11),pp.685–692.

Hardjasa, A., Ling, M., Ma, K. and Yu, H. 2010. Investigating the Effects of DMSO on PCR Fidelity Using a Restriction Digest-Based Method. *Journal of Experimental Microbiology and Immunology (JEMI)*. **14**,pp.161–164.

Helfrich, E.J.N., Reiter, S. and Piel, J. 2014. Recent Advances in Genome-Based Polyketide Discovery. *Current Opinion in Biotechnology*. **29**,pp.107–115.

Hendrickson, L., Ray Davis, C., Roach, C., Kim Nguyen, D., Aldrich, T., McAda, P.C. and Reeves, C.D. 1999. Lovastatin biosynthesis in Aspergillus terreus: Characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene. *Chemistry & Biology*. **6**(7),pp.429–439.

Herbst, D.A., Huitt-Roehl, C.R., Jakob, R.P., Kravetz, J.M., Storm, P.A., Alley, J.R., Townsend, C.A. and Maier, T. 2018. The structural organization of substrate loading in iterative polyketide synthases. *Nat Chem Biol*. **14**(5),pp.474–479.

Herbst, D.A., Jakob, R.P., Zähringer, F. and Maier, T. 2016. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature*. **531**(7595),pp.533–537.

Holm, L. and Laakso, L.M. 2016. Dali server update. *Nucleic Acids Research*. **44**(W1),pp.W351–W355.

Hong, J., and White, J. D. (2004). The chemistry and biology of rhizoxins, novel antitumor macrolides from *Rhizopus chinensis*. *Tetrahedron*, **60**(27), pp.5653–5681.

Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G. and Lindqvist, Y. 1998. Crystal structure of beta-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. *The EMBO Journal*. **17**(5),pp.1183–1191.

Hubbard, T.J.P., Ailey, B., Brenner, S.E., Murzin, A.G. and Chothia, C. 1999. SCOP: a Structural Classification of Proteins database. *Nucleic Acids Research*. **27**(1),pp.254–256.

Hyde, A.M., Zultanski, S.L., Waldman, J.H., Zhong, Y.-L., Shevlin, M. and Peng, F. 2017. General Principles and Strategies for Salting-Out Informed by the Hofmeister Series. *Organic Process Research & Development*. **21**(9),pp.1355–1370.

Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S. 1999. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proceedings of the National Academy of Sciences*. **96**(17),pp.9509–9514.

Janitz, M. 2011. *Next-Generation Genome Sequencing*. John Wiley & Sons.

Jaravine, V.A. and Orekhov, V.Y. 2006. Targeted Acquisition for Real-Time NMR Spectroscopy. *Journal of the American Chemical Society*. **128**(41),pp.13421–13426.

Jaravine, V.A., Zhuravleva, A.V., Permi, P., Ibraghimov, I. and Orekhov, V.Y. 2008. Hyperdimensional NMR Spectroscopy with Nonlinear Sampling. *Journal of the American Chemical Society*. **130**(39),pp.13182–13182.

Johansson, P., Unge, T., Cronin, A., Arand, M., Bergfors, T., Jones, T.A. and Mowbray, S.L. 2005. Structure of an Atypical Epoxide Hydrolase from *Mycobacterium tuberculosis* Gives Insights into its Function. *Journal of Molecular Biology*. **351**(5),pp.1048–1056.

Jordan, A., Hadfield, J.A., Lawrence, N.J. and McGown, A.T. 1998. Tubulin as a target for anticancer drugs: Agents which interact with the mitotic spindle. *Medicinal Research Reviews*. **18**(4),pp.259–296.

Kabsch, W. 2010. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography*. **66**(2),pp.133–144.

Kallio, P., Sultana, A., Niemi, J., Mäntsälä, P. and Schneider, G. 2006. Crystal Structure of the Polyketide Cyclase AknH with Bound Substrate and Product Analogue: Implications for Catalytic Mechanism and Product Stereoselectivity. *Journal of Molecular Biology*. **357**(1),pp.210–220.

Kang, J., Soog Lee, M. and Gorenstein, D.G. 2005. The Enhancement of PCR Amplification of a Random Sequence DNA Library by DMSO and Betaine: Application to *in vitro* Combinatorial Selection of Aptamers. *Journal of Biochemical and Biophysical Methods*. **64**(2),pp.147–151.

Kang, J. Y., Olinares, P. D. B., Chen, J., Campbell, E. A., Mustaev, A., Chait, B. T., et al. 2017. Structural basis of transcription arrest by coliphage HK022 Nun in an *Escherichia coli* RNA polymerase elongation complex. *eLife*. **6**,213.

Kantardjieff, K.A. and Rupp, B. 2003. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Science*. **12**(9),pp.1865–1871.

Kao, C.M., Luo, G., Katz, L., Cane, D.E. and Khosla, C. 1995. Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase. *Journal of the American Chemical Society*. **117**(35),pp.9105–9106.

Kapur, S., Chen, A.Y., Cane, D.E. and Khosla, C. 2010. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences*. **107**(51),pp.22066–22071.

Kasahara, K., Miyamoto, T., Fujimoto, T., Oguri, H., Tokiwano, T., Oikawa, H., Ebizuka, Y. and Fujii, I. 2010. Solanapyrone synthase, a possible Diels-Alderase and iterative type I polyketide synthase encoded in a biosynthetic gene cluster from *Alternaria solani*. *ChemBioChem*. **11**(9),pp.1245–1252.

Kasinath, V., Faini, M., Poepsel, S., Reif, D., Feng, X.A., Stjepanovic, G., Aebersold, R. and Nogales, E. 2018. Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science*. **359**(6378),pp.940–944.

Keatinge-Clay, A. 2008. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *Journal of Molecular Biology*. **384**(4),pp.941–953.

Keatinge-Clay, A.T. 2007. A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chemistry & Biology*. **14**(8),pp.898–908.

Keatinge-Clay, A.T. 2012. The structures of type I polyketide synthases. *Nat Prod Rep*. **29**(10),pp.1050–1073.

Keatinge-Clay, A.T. and Stroud, R.M. 2006. The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure*. **14**(4),pp.737–748.

Keatinge-Clay, A.T., Maltby, D.A., Medzihradszky, K.F., Khosla, C. and Stroud, R.M. 2004. An antibiotic factory caught in action. *Nature Structural & Molecular Biology*. **11**(9),pp.888–893.

Keatinge-Clay, A.T., Shelat, A.A., Savage, D.F., Tsai, S.-C., Miercke, L.J., O'Connell, J.D., Khosla, C. and Stroud, R.M. 2003. Catalysis, specificity, and ACP docking site of Streptomyces coelicolor malonyl-CoA: ACP transacylase. *Structure*. **11**(2),pp.147–154.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J.E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*. **10**(6),pp.845–858.

Kennedy, J., Auclair, K., Kendrew, S.G., Park, C., Vederas, J.C. and Hutchinson, C.R. 1999. Modulation of Polyketide Synthase Activity by Accessory Proteins During Lovastatin Biosynthesis. *Science*. **284**(5418),pp.1368–1372.

Khare, D., Hale, W.A., Tripathi, A., Gu, L., Sherman, D.H., Gerwick, W.H., Hakansson, K. and Smith, J.L. 2015. Structural Basis for Cyclopropanation by a Unique Enoyl-Acyl Carrier Protein Reductase. *Structure*. **23**(12),pp.2213–2223.

Khosla, C., Tang, Y., Chen, A.Y., Schnarr, N.A. and Cane, D.E. 2007. Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annual Review of Biochemistry*. **76**(1),pp.195–221.

Kim, E., Moore, B.S. and Yoon, Y.J. 2015. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat Chem Biol*. **11**(9),pp.649–659.

Kim, H.J., Ruszczycky, M.W. and Liu, H.-W. 2012. Current Developments and Challenges in the Search for a Naturally Selected Diels-Alderase. *Current Opinion in Chemical Biology*. **16**(1–2),pp.124–131.

Kim, Y., Tanner, K.G. and Denu, J.M. 2000. A continuous, nonradioactive assay for histone acyltransferases. *Analytical Biochemistry*. **280**(2),pp.308–314.

Kingston, D.G.I. 2011. Modern natural products drug discovery and its relevance to biodiversity conservation. *Journal of Natural Products*. **74**(3),pp.496–511.

Scherlach, K., Partida-Martinez, L.P., Hans-Martin Dahse, A., Hertweck, C. 2006. Antimitotic Rhizoxin Derivatives from a Cultured Bacterial Endosymbiont of the Rice Pathogenic Fungus Rhizopus microsporus. *Journal of the American Chemical Society.* **128**,pp.11529–11536).

Klas, K., Tsukamoto, S., Sherman, D.H. and Williams, R.M. 2015. Natural Diels-Alderases: Elusive and Irresistable. *The Journal of Organic Chemistry*. **80**(23),pp.11672–11685.

Kusebauch, B., Busch, B., Scherlach, K., Roth, M. and Hertweck, C. 2010. Functionally distinct modules operate two consecutive alpha,beta--beta,gamma double-bond shifts in the rhizoxin polyketide assembly line. *Angewandte Chemie (International ed. in English)*. **49**(8),pp.1460–1464.

Kusebauch, B., Busch, B., Scherlach, K., Roth, M. and Hertweck, C. 2009. Polyketide-Chain Branching by an Enzymatic Michael Addition. *Angewandte Chemie*. **121**(27),pp.5101–5104.

Kusebauch, B., Scherlach, K., Kirchner, H., Dahse, H.M. and Hertweck, C. 2011. Antiproliferative Effects of Ester- and Amide-Functionalized Rhizoxin Derivatives. *ChemMedChem*. **6**(11),pp.1998–2001.

Kwon, H.-J., Smith, W.C., Xiang, L. and Shen, B. 2001. Cloning and Heterologous Expression of the Macrotetrolide Biosynthetic Gene Cluster Revealed a Novel Polyketide Synthase that Lacks an Acyl Carrier Protein. *Journal of the American Chemical Society*. **123**(14),pp.3385–3386.

Lackner, G., Moebius, N., Partida-Martinez, L. and Hertweck, C. 2011. Complete genome sequence of *Burkholderia rhizoxinica*, an Endosymbiont of Rhizopus microsporus. *Journal of Bacteriology*. **193**(3),pp.783–784.

Lackner, G., Moebius, N., Partida-Martinez, L.P., Boland, S. and Hertweck, C. 2011. Evolution of an endofungal lifestyle: Deductions from the *Burkholderia rhizoxinica* genome. *BMC genomics*. **12**(1),p.210.

Leadlay, P.F. 2014. Structural Biology: Enzyme Assembly Line Pictured. *Nature*. **510**(7506),pp.482–483.

Li, C., Roege, K.E. and Kelly, W.L. 2009. Analysis of the Indanomycin Biosynthetic Gene Cluster from *Streptomyces antibioticus* NRRL 8167. *ChemBioChem*. **10**(6),pp.1064–1072.

Li, C., Wen, A., Shen, B., Lu, J., Huang, Y. and Chang, Y. 2011. FastCloning: a Highly Simplified, Purification-Free, Sequence- and Ligation-Independent PCR Cloning Method. *BMC Biotechnol*. **11**,p.92.

Li, J.W.-H. and Vederas, J.C. 2009. Drug discovery and natural products: end of an era or an endless frontier? *Science*. **325**(5937),pp.161–165.

Li, M.H., Ung, P.M., Zajkowski, J., Garneau-Tsodikova, S. and Sherman, D.H. 2009. Automated Genome Mining for Natural Products. *BMC Bioinformatics*. **10**(1),p.185.

Li, Q., Khosla, C., Puglisi, J.D. and Liu, C.W. 2003. Solution structure and backbone dynamics of the holo form of the frenolicin acyl carrier protein. *Biochemistry*. **42**(16),pp.4648–4657.

Li, Y., Zhang, W., Zhang, H., Tian, W., Wu, L., Wang, S., Zheng, M., Zhang, J., Sun, C., Deng, Z., Sun, Y., Qu, X. and Zhou, J. 2018. Structural Basis of a Broadly Selective Acyltransferase from the Polyketide Synthase of Splenocin. *Angewandte Chemie (International ed. in English).* **57**(20),pp.5823–5827.

Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J. and Tan, S. 2005. Comparison of Affinity Tags for Protein Purification. *Protein Expression and Purification*. **41**(1),pp.98–105.

Liew, C.W., Nilsson, M., Chen, M.W., Sun, H., Cornvik, T., Liang, Z.-X. and Lescar, J. 2012. Crystal structure of the acyltransferase domain of the iterative polyketide synthase in enediyne biosynthesis. *J Biol Chem*. **287**(27),pp.23203–23215.

Linder, M., Johansson, A.J., Olsson, T.S.G., Liebeschuetz, J. and Brinck, T. 2012. Computational design of a Diels-Alderase from a thermophilic esterase: The importance of dynamics. *Journal of Computer-Aided Molecular Design*. **26**(9),pp.1079–1095.

Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., et al. 2015. A new antibiotic kills pathogens without detectable resistance. *Nature*. **517**(7535),pp.455–459.

Lobley, A., Whitmore, L. and Wallace, B.A. 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics (Oxford, England)*. **18**(1),pp.211–212.

Lundqvist, T., Rice, J., Hodge, C.N., Basarab, G.S., Pierce, J. and Lindqvist, Y. 1994. Crystal structure of scytalone dehydratase--a disease determinant of the rice pathogen, *Magnaporthe grisea*. *Structure*. **2**(10),pp.937–944.

Lüttgen, H., Robelek, R., Mühlberger, R., Diercks, T., Schuster, S.C., Köhler, P., Kessler, H., Bacher, A. and Richter, G. 2002. Transcriptional Regulation by Antitermination. Interaction of RNA with NusB Protein and NusB/NusE Protein Complex of *Escherichia coli*. *Journal of Molecular Biology*. **316**(4),pp.875–885.

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

Maier, T., Leibundgut, M. and Ban, N. 2008. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science*. **321**(5894),pp.1315–1322.

Malpartida, F. and Hopwood, D.A. 1984. Molecular cloning of the whole biosynthetic pathway of a *Streptomyces antibioticus* and its expression in a heterologous host. *Nature*. **309**(5967),pp.462–464.

Marblestone, J.G., Edavettal, S.C., Lim, Y., Lim, P., Zuo, X. and Butt, T.R. 2006. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Science*. **15**(1),pp.182–189.

Marsden, A.F., Wilkinson, B., Cortés, J., Dunster, N.J., Staunton, J. and Leadlay, P.F. 1998. Engineering Broader Specificity into an Antibiotic-Producing Polyketide Synthase. *Science*. **279**(5348),pp.199–202.

McChesney, J.D., Venkataraman, S.K. and Henri, J.T. 2007. Plant natural products: Back to the future or into extinction? *Phytochemistry*. **68**(14),pp.2015–2022.

McCoy, A.J. 2007. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica Section D: Biological Crystallography*. **63**(Pt 1),pp.32–41.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. 2007. Phaser crystallographic software. *Journal of applied crystallography*. **40**(Pt 4),pp.658–674.

McDaniel, R., Ebert-Khosla, S., Hopwood, D.A. and Khosla, C. 1995. Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature*. **375**(6532),pp.549–554.

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, Melanie, Betlach, Mary and Ashley, G. 1999. Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel 'Unnatural' Natural Products. *Proceedings of the National Academy of Sciences*. **96**(5),pp.1846–1851.

McDaniel, R., Welch, M. and Hutchinson, C.R. 2005. Genetic approaches to polyketide antibiotics. *Chemical reviews*. **105**(2),pp.543–558.

McMinn, C.L. and Ottaway, J.H. 1975. Studies on the Mechanism of the 2-Oxoglutarate Dehydrogenase System. *Biochemical Society Transactions*. **3**(6),pp.1071–1074.

McMullan, G., Faruqi, A. and Henderson, R. 2016. *The Resolution Revolution: Recent Advances In cryo-EM. Vol. 579*.

Miller, L.H. and Su, X. 2011. *Artemisinin: discovery from the Chinese herbal garden.*

Minami, A., Ose, T., Sato, K., Oikawa, A., Kuroki, K., Maenaka, K., Oguri, H. and Oikawa, H. 2013. Allosteric Regulation of Epoxide Opening Cascades by a Pair of Epoxide Hydrolases in Monensin Biosynthesis. *ACS Chemical Biology*. **9**(2),pp.562–569.

Minowa, Y., Araki, M. and Kanehisa, M. 2007. Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. *Journal of Molecular Biology*. **368**(5),pp.1500–1517.

Molnos, J., Gardiner, R., Dale, G.E. and Lange, R. 2003. A continuous Coupled Enzyme Assay for Bacterial Malonyl–CoA:Acyl Carrier Protein Transacylase (FabD). *Analytical Biochemistry*. **319**(1),pp.171–176.

Moloney, M.G. 2016. Natural Products as a Source for Novel Antibiotics. *Trends in Pharmacological Sciences*. **37**(8),pp.689–701.

Mori, T., Iwabuchi, T., Hoshino, S., Wang, H., Matsuda, Y. and Abe, I. 2017. Molecular basis for the unusual ring reconstruction in fungal meroterpenoid biogenesis. *Nat Chem Biol*. **13**(10),pp.1066–1073.

Motamedi, H., Cai, S.-J., Shafiee, A. and Elliston, K.O. 1997. Structural Organization of A Multifunctional Polyketide Synthase Involved in the Biosynthesis of the Macrolide Immunosuppressant FK506. *European Journal of Biochemistry*. **244**(1),pp.74–80.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*. **67**(Pt 4),pp.355–367.

Nakano, M.M., Corbell, N., Besson, J. and Zuber, P. 1992. Isolation and Characterization of Sfp: a Gene that Functions in the Production of the Lipopeptide Biosurfactant, Surfactin, in *Bacillus subtilis*. *Molecular and General Genetics MGG*. **232**(2),pp.313–321.

Nett, M. 2014. Genome Mining: Concept and Strategies for Natural Product Discovery. *Progress in the Chemistry of Organic Natural Products*. **99**,pp.199–245.

Newman, D.J. and Cragg, G.M. 2012. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *Journal of Natural Products*. **75**(3),pp.311–335.

Nguyen, T.H.D., Tam, J., Wu, R.A., Greber, B.J., Toso, D., Nogales, E. and Collins, K. 2018. Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature*. **557**(7704),pp.190–195.

Nilsson, J., Larsson, M., Ståhl, S., Nygren, P.A. and Uhlén, M. 1996. Multiple affinity domains for the detection, purification and immobilization of recombinant proteins. *Journal of molecular recognition : JMR*. **9**(5-6),pp.585–594.

Noble, A.J., Dandey, V.P., Wei, H., Brasch, J., Chase, J., Acharya, P., Tan, Y.Z., Zhang, Z., Kim, L.Y., Scapin, G., Rapp, M., Eng, E.T., Rice, W.J., Cheng, A., Negro, C.J., Shapiro, L., Kwong, P.D., Jeruzalmi, D., Georges, des, A., Potter, C.S. and

Carragher, B. 2018. Routine single particle CryoEM sample and grid characterization by tomography. *eLife*. **7**,p.32.

Nurizzo, D., Bowler, M.W., Caserotto, H., Dobias, F., Giraud, T., Surr, J., Guichard, N., Papp, G., Guijarro, M., Mueller-Dieckmann, C., Flot, D., McSweeney, S., Cipriani, F., Theveneau, P., Leonard, G.A.2016. RoboDiff: combining a sample changer and goniometer for highly automated macromolecular crystallography experiments. *Acta Crystallographica Section D: Structural Biology*. **72**(8),pp.966–975.

Obajdin, J. 2018. Structural studies of a putative polyketide cyclase IdmH involved in indanomycin biosynthesis. University of Leeds.

Olano, C., Méndez, C. and Salas, J.A. 2010. Post-PKS tailoring steps in natural product -producing actinomycetes from the perspective of combinatorial biosynthesis. *Natural Product Reports*. **27**(4),pp.571–616.

Orekhov, V.Y. and Jaravine, V.A. 2011. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Progress in nuclear magnetic resonance spectroscopy*. **59**(3),pp.271–292.

Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P.S., Keasling, J.D., Reiling, K.K., Renninger, N.S. and Newman, J.D. 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*. **496**(7446),pp.528–532.

Park, H., Kevany, B.M., Dyer, D.H., Thomas, M.G. and Forest, K.T. 2014. A Polyketide Synthase Acyltransferase Domain Structure Suggests a Recognition Mechanism for Its Hydroxymalonyl-Acyl Carrier Protein Substrate A. Surolia, ed. *PLoS One*. **9**(10),p.e110965.

Partida-Martinez, L.P. and Hertweck, C. 2007. A Gene Cluster Encoding Rhizoxin Biosynthesis in *Burkholderia rhizoxina*, the Bacterial Endosymbiont of the Fungus *Rhizopus microsporus*. *ChemBioChem*. **8**(1),pp.41–45.

Partida-Martinez, L.P. and Hertweck, C. 2005. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature*. **437**(7060),pp.884–888.

Patridge, E., Gareiss, P., Kinch, M.S. and Hoyer, D. 2016. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discovery Today*. **21**(2),pp.204–207.

Persson, B., Hedlund, J. and Jörnvall, H. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cellular and Molecular Life Sciences*. **65**(24),pp.3879–3894.

Poepsel, S., Kasinath, V., Nogales, E. 2018. Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nature Structural & Molecular Biology*. **25**(2),pp.154–162.

Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P.R., Jenkins, H.T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., Mcnicholas, S.J., Nicholls, R.A., Noble, M., Pannu, N.S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., Uski, V., Delft, von, F., Waterman, D., Wilson, K., Winn, M. and Wojdyr, M. 2018.

CCP4i2: The new graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D: Structural Biology*. **74**(2),pp.68–84.

Poust, S., Hagen, A., Katz, L. and Keasling, J.D. 2014. Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform. *Current Opinion in Biotechnology*. **30**,pp.32–39.

Quadri, L.E.N., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P. and Walsh, C.T. 1998. Characterization of Sfp, a *Bacillus subtilis* Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. *Biochemistry*. **37**(6),pp.1585–1595.

Reeves, C.D., Murli, S., Ashley, G.W., Piagentini, M., Hutchinson, C.R. and McDaniel, R. 2001. Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations †. *Biochemistry*. **40**(51),pp.15464–15470.

Ridley, C.P., Lee, H.Y. and Khosla, C. 2008. Evolution of polyketide synthases in bacteria. *Proc Natl Acad Sci U S A*. **105**(12),pp.4595–4600.

Robbins, T., Liu, Y.-C., Cane, D.E. and Khosla, C. 2016. Structure and mechanism of assembly line polyketide synthases. *Current Opinion in Structural Biology*. **41**,pp.10–18.

Robichon, C., Luo, J., Causey, T.B., Benner, J.S. and Samuelson, J.C. 2011. Engineering *Escherichia coli* BL21(DE3) Derivative Strains to Minimize *E. coli* Protein Contamination after Purification by Immobilized Metal Affinity Chromatography. *Applied and Environmental Microbiology*. **77**(13),pp.4634–4646.

Robinson, R. 1995. The Structural Relations of Natural Products. *AIBS Bulletin*. **5**(5),pp.11–11.

Rommel, K.R., Li, C. and Kelly, W.L. 2011. Identification of a tetraene-containing product of the indanomycin biosynthetic pathway. *Org Lett*. **13**(10),pp.2536–2539.

Rosenthal, P.B. and Henderson, R. 2003. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology*. **333**(4),pp.721–745.

Rovnyak, D., Frueh, D.P., Sastry, M., Sun, Z.-Y.J., Stern, A.S., Hoch, J.C. and Wagner, G. 2004. Accelerated Acquisition of High Resolution Triple-Resonance Spectra Using Non-Uniform Sampling and Maximum Entropy Reconstruction. *Journal of Magnetic Resonance*. **170**(1),pp.15–21.

Rowe, C.J., Böhm, I.U., Thomas, I.P., Wilkinson, B., Rudd, B.A.M., Foster, G., Blackaby, A.P., Sidebottom, P.J., Roddis, Y., Buss, A.D., Staunton, J. and Leadlay, P.F. 2001. Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chemistry & Biology*. **8**(5),pp.475–485.

Ruan, B., Pong, K., Jow, F., Bowlby, M., Crozier, R.A., Liu, D., Liang, S., Chen, Y., Mercado, M.L., Feng, X., Bennett, F., Schack, von, D., McDonald, L., Zaleska, M.M., Wood, A., Reinhart, P.H., Magolda, R.L., Skotnicki, J., Pangalos, M.N., Koehn, F.E., Carter, G.T., Abou-Gharbia, M. and Graziani, E.I. 2008. Binding of rapamycin analogs to calcium channels and FKBP52 contributes to their neuroprotective activities. *Proc Natl Acad Sci U S A*. **105**(1),pp.33–38.

Scaglione, J.B., Akey, D.L., Sullivan, R., Kittendorf, J.D., Rath, C.M., Kim, E.S., Smith, J.L. and Sherman, D.H. 2010. Biochemical and Structural Characterization of the

Tautomycetin Thioesterase: Analysis of a Stereoselective Polyketide Hydrolase. *Angewandte Chemie*. **122**(33),pp.5862–5866.

Scheres, S.H.W. 2012. RELION: Implementation of a Bayesian Approach to Cryo-Em Structure Determination. *Journal of Structural Biology*. **180**(3),pp.519–530.

Scheres, S.H.W. 2015. Semi-Automated Selection of Cryo-Em Particles in RELION-1.3. *Journal of Structural Biology*. **189**(2),pp.114–122.

Scheres, S.H.W. and Chen, S. 2012. Prevention of overfitting in cryo-EM structure determination. *Nat Meth*. **9**(9),pp.853–854.

Scherlach, K., Busch, B., Lackner, G., Paszkowski, U. and Hertweck, C. 2012. Symbiotic Cooperation in the Biosynthesis of a Phytotoxin. *Angewandte Chemie*. **124**(38),pp.9753–9756.

Schmidt, C. and Robinson, C.V. 2014. A comparative cross-linking strategy to probe conformational changes in protein complexes. *Nat. Protocols*. **9**(9),pp.2224–2236.

Schmidt, T.G.M. and Skerra, A. 2007. The Strep-Tag System for One-Step Purification and High-Affinity Detection or Capturing Of Proteins. *Nat. Protocols*. **2**(6),pp.1528–1535.

Schmitt, I., Partida-Martinez, L.P., Winkler, R., Voigt, K., Einax, E., Dölz, F., Telle, S., Wöstemeyer, J. and Hertweck, C. 2008. Evolution of host resistance in a toxin-producing bacterial–fungal alliance. *The ISME Journal*. **2**(6),pp.632–641.

Schröder, J. 1999. Probing plant polyketide biosynthesis. *Nature structural biology*. **6**(8),pp.714–716.

Schwans, J.P., Sunden, F., Gonzalez, A., Tsai, Y. and Herschlag, D. 2011. Evaluating the Catalytic Contribution from the Oxyanion Hole in Ketosteroid Isomerase. *Journal of the American Chemical Society*. **133**(50),pp.20052–20055.

Schwecke, T., Aparicio, J.F., Molnar, I., Konig, A., Khaw, L.E., Haydock, S.F., Oliynyk, M., Caffrey, P., Cortes, J. and Lester, J.B. 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proceedings of the National Academy of Sciences*. **92**(17),pp.7839–7843.

Serafimov, J.M., Gillingham, D., Kuster, S. and Hilvert, D. 2008. The putative Diels-Alderase macrophomate synthase is an efficient aldolase. *Journal of the American Chemical Society*. **130**(25),pp.7798–7799.

Serafimov, J.M., Lehmann, H.C., Oikawa, H. and Hilvert, D. 2007. Active site mutagenesis of the putative Diels-Alderase macrophomate synthase. *Chemical Communications*. **111**(17),pp.1701–1703.

Sheldrick, G.M. 2010. Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallographica Section D: Biological Crystallography*. **66**(4),pp.479–485.

Shen, B. 2015. A New Golden Age of Natural Products Drug Discovery. *Cell*. **163**(6),pp.1297–1300.

Shen, B. 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*. **7**(2),pp.285–295.

Sheryl Tsai, S.C. and Ames, B.D. 2009. Chapter 2 Structural Enzymology of Polyketide Synthases. *Methods in Enzymology*. **459**,pp.17–47.

Shoolinghin-Jordan, P.M. and Campuzano, I.D.G. 1999. Biosynthesis of 6-Methylsalicylic Acid *In*: *Comprehensive Natural Products Chemistry*. Elsevier, pp. 345–365.

Skinner, S.P., Fogh, R.H., Boucher, W., Ragan, T.J., Mureddu, L.G. and Vuister, G.W. 2016. CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *Journal of biomolecular NMR*. **66**(2),pp.111–124.

Smanski, M.J., Zhou, H., Claesen, J., Ben Shen, Fischbach, M.A. and Voigt, C.A. 2016. Synthetic biology to access and expand nature's chemical diversity. *Nature Reviews Microbiology*. **14**(3),pp.135–149.

Sreerama, N. and Woody, R.W. 2000. Estimation of Protein Secondary Structure From Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. *Analytical Biochemistry*. **287**(2),pp.252–260.

Starcevic, A., Zucko, J., Simunkovic, J., Long, P.F., Cullum, J. and Hranueli, D. 2008. ClustScan : an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Research*. **36**(21),pp.6882–6892.

Stark, H. 2010. GraFix: Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM *In*: *Cryo-EM Part A Sample Preparation and Data Collection*. Methods in Enzymology. Elsevier, pp. 109–126.

Stathopulos, P.B., Scholz, G.A., Hwang, Y.M., Rumfeldt, J.A.O., Lepock, J.R. and Meiering, E.M. 2008. Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Science*. **13**(11),pp.3017–3027.

Staunton, J. and Weissman, K.J. 2001. Polyketide biosynthesis: a millennium review. *Nat Prod Rep*. **18**(4),pp.380–416.

Sultana, A., Kallio, P., Jansson, A., Wang, J.S., Niemi, J., Mäntsälä, P. and Schneider, G. 2004. Structure of the polyketide cyclase SnoaL reveals a novel mechanism for enzymatic aldol condensation. *The EMBO Journal*. **23**(9),pp.1911–1921.

Summers, R.G., Donadio, S., Staver, M.J., Wendt-Pienkowski, E., Hutchinson, C.R. and Katz, L. 1997. Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of *Saccharopolyspora erythraea* that are involved in L-mycarose and D-desosamine production. *Microbiology*. **143**(10),pp.3251–3262.

Sundermann, U., Bravo-Rodriguez, K., Klopries, S., Kushnir, S., Gomez, H., Sanchez-Garcia, E. and Schulz, F. 2012. Enzyme-Directed Mutasynthesis: A Combined Experimental and Theoretical Approach to Substrate Recognition of a Polyketide Synthase. *ACS Chemical Biology*. **8**(2),pp.443–450.

Svensson, O., Gilski, M., Nurizzo, D. and Bowler, M.W. 2017. Recent improvements to the automatic characterization and data collection algorithms on MASSIF-1. *bioRxiv*.p.236596.

Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D., Bowler, M.W. 2015. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallographica Section D: Biological Crystallography*. **71**(8),pp.1757–1767.

Tang, Y., Chen, A.Y., Kim, C.-Y., Cane, D.E. and Khosla, C. 2007. Structural and Mechanistic Analysis of Protein Interactions in Module 3 of the 6-Deoxyerythronolide B Synthase. *Chemistry & Biology*. **14**(8),pp.931–943.

Tang, Y., Kim, C.Y., Mathews, I.I., Cane, D.E. and Khosla, C. 2006. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences*. **103**(30),pp.11124–11129.

Tang, Y., Lee, T.S. and Khosla, C. 2004. Engineered Biosynthesis of Regioselectively Modified Aromatic Polyketides Using Bimodular Polyketide Synthases. Rowena G Matthews, ed. *PLOS Biology*. **2**(2),p.e31.

Tarry, M.J., Haque, A.S., Bui, K.H. and Schmeing, T.M. 2017. X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure*. **25**(5),pp.783–793.e4.

Terwilliger, T.C., Bunkóczi, G., Hung, L.-W., Zwart, P.H., Smith, J.L., Akey, D.L. and Adams, P.D. 2016. Can I solve my structure by SAD phasing? Anomalous signal in SAD phasing. *Acta Crystallographica Section D: Structural Biology*. **72**(3),pp.346–358.

Thomas, I., Martin, C.J., Wilkinson, C.J., Staunton, J. and Leadlay, P.F. 2002. Skipping in a Hybrid Polyketide Synthase: Evidence for ACP-to-ACP Chain Transfer. *Chemistry & Biology*. **9**(7),pp.781–787.

Tian, W., Chen, C., Lei, X., Zhao, J. and Liang, J. 2018. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **350**,p.203.

Tinberg, C.E., Khare, S.D., Dou, J., Doyle, L., Nelson, J.W., Schena, A., Jankowski, W., Kalodimos, C.G., Johnsson, K., Stoddard, B.L. and Baker, D. 2013. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*. **501**(7466),pp.212–216.

Tobert, J.A. 2003. Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. *Nature Reviews Drug Discovery*. **2**(7),pp.517–526.

Trurnit, H.J. 1960. A theory and method for the spreading of protein monolayers. *Journal of Colloid Science*. **15**(1),pp.1–13.

Tsai, S.-C., Lu, H., Cane, D.E., Khosla, C. and Stroud, R.M. 2002. Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry*. **41**(42),pp.12598–12606.

Tsai, S.-C., Miercke, L.J.W., Krucinski, J., Gokhale, R., Chen, J.C.H., Foster, P.G., Cane, D.E., Khosla, C. and Stroud, R.M. 2001. Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: Versatility from a unique substrate channel. *Proceedings of the National Academy of Sciences*. **98**(26),pp.14808–14813.

Tsuruo, T., Oh-hara, T., Iida, H., Tsukagoshi, S., Sato, Z., Matsuda, I., Iwasaki, S., Okuda, S., Shimizu, F. and Sasagawa, K. 1986. Rhizoxin, a macrocyclic lactone antibiotic, as a new antitumor agent against human and murine tumor cells and their vincristine-resistant sublines. *Cancer Research*. **46**(1),pp.381–385.

Tuan, J.S., Weber, J.M., Staver, M.J., Leung, J.O., Donadio, S. and Katz, L. 1990. Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora*

erythraea using a novel actinomycete-*Escherichia coli* cosmid. *Gene*. **90**(1),pp.21–29.

van der Kamp, M.W. 2018 Enlighten protocols. Available from: https://github.com/marcvanderkamp/enlighten.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. and Laue, E.D. 2005. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins*. **59**(4),pp.687–696.

Wagner, S., Klepsch, M.M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K.J., Slotboom, D.J., Persson, J.O. and de Gier, J.-W. 2008. Tuning *Escherichia coli* for Membrane Protein Overexpression. *Proceedings of the National Academy of Sciences*. **105**(38),pp.14371–14376.

Wang, F., Wang, Y., Ji, J., Zhou, Z., Yu, J., Zhu, H., Su, Z., Zhang, L. and Zheng, J. 2015. Structural and Functional Analysis of the Loading Acyltransferase from Avermectin Modular Polyketide Synthase. *ACS Chemical Biology*. **10**(4),pp.1017–1025.

Wani, M.C., Taylor, H.L., Wall, M.E., Coggon, P. and McPhail, A.T. 1971. Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *Journal of the American Chemical Society*. **93**(9),pp.2325–2327.

Watanabe, K., Wang, C.C.C., Boddy, C.N., Cane, D.E. and Khosla, C. 2003. Understanding substrate specificity of polyketide synthase modules by generating hybrid multimodular synthases. *J Biol Chem*. **278**(43),pp.42020–42026.

Waugh, D.S. 2005. Making the most of affinity tags. *Trends in Biotechnology*. **23**(6),pp.316–320.

Webster, J. and Oxley, D. 2012. Protein Identification by MALDI-TOF Mass Spectrometry. *Methods Mol Biol*. **800**,pp.227–240.

Weissman, K.J. 2015a. The structural biology of biosynthetic megaenzymes. *Nat Chem Biol*. **11**(9),pp.660–670.

Weissman, K.J. 2015b. Uncovering the structures of modular polyketide synthases. *Nat Prod Rep*. **32**(3),pp.436–453.

Weissman, K.J. and Leadlay, P.F. 2005. Combinatorial biosynthesis of reduced polyketides. *Nature Reviews Microbiology*. **3**(12),pp.925–936.

Weissman, K.J., Timoney, M., Bycroft, M., Grice, P., Hanefeld, U., Staunton, J. and Leadlay, P.F. 1997. The Molecular Basis of Celmer's Rules: The Stereochemistry of the Condensation Step in Chain Extension on the Erythromycin Polyketide Synthase †. *Biochemistry*. **36**(45),pp.13849–13855.

Wenzel, R.P. 2004. The antibiotic pipeline--challenges, costs, and values. *The New England journal of medicine*. **351**(6),pp.523–526.

Wettstein-Knowles, von, P., Olsen, J.G., McGuire, K.A. and Henriksen, A. 2006. Fatty acid synthesis. Role of active site histidines and lysine in Cys-His-His-type beta-ketoacyl-acyl carrier protein synthases. *FEBS Journal*. **273**(4),pp.695–710.

Whicher, J.R., Dutta, S., Hansen, D.A., Hale, W.A., Chemler, J.A., Dosey, A.M., Narayan, A.R.H., Hakansson, K., Sherman, D.H., Smith, J.L. and Skiniotis, G. 2014.

Structural Rearrangements of a Polyketide Synthase Module During its Catalytic Cycle. *Nature*. **510**(7506),pp.560–564.

Whitmore, L. and Wallace, B.A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res*. **32**(Web Server issue),pp.W668–73.

Wilkinson, B., Foster, G., Rudd, B.A., Taylor, N.L., Blackaby, A.P., Sidebottom, P.J., Cooper, D.J., Dawson, M.J., Buss, A.D., Gaisser, S., Böhm, I.U., Rowe, C.J., Cortés, J., Leadlay, P.F. and Staunton, J. 2000. Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for iterative operation of the erythromycin polyketide synthase. *Chemistry & Biology*. **7**(2),pp.111–117.

Wilkinson, C.J., Frost, E.J., Staunton, J. and Leadlay, P.F. 2001. Chain initiation on the soraphen-producing modular polyketide synthase from *Sorangium cellulosum*. *Chemistry & Biology*. **8**(12),pp.1197–1208.

Williams, D.H., Stone, M.J., Hauck, P.R. and Rahman, S.K. 1989. Why are secondary metabolites (natural products) biosynthesized? *Journal of Natural Products*. **52**(6),pp.1189–1208.

Williamson, M.P. 2013. Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy*. **73**,pp.1–16.

Witkowski, A., Joshi, A.K. and Smith, S. 2002. Mechanism of the β-Ketoacyl Synthase Reaction Catalyzed by the Animal Fatty Acid Synthase †. *Biochemistry*. **41**(35),pp.10877–10887.

Wong, F.T., Jin, X., Mathews, I.I., Cane, D.E. and Khosla, C. 2011. Structure and mechanism of the trans-acting acyltransferase from the disorazole synthase. *Biochemistry*. **50**(30),pp.6539–6548.

Wong, F.T. and Khosla, C. 2012. Combinatorial Biosynthesis of Polyketides - A Perspective. *Current Opinion in Chemical Biology*. **16**(1–2),pp.117–123.

Wu, J., Kinoshita, K., Khosla, C. and Cane, D.E. 2004. Biochemical analysis of the substrate specificity of the beta-ketoacyl-acyl carrier protein synthase domain of module 2 of the erythromycin polyketide synthase. *Biochemistry*. **43**(51),pp.16301–16310.

Xue, Y., Zhao, L., Liu, H.W. and Sherman, D.H. 1998. A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. *Proceedings of the National Academy of Sciences*. **95**(21),pp.12111–12116.

Yadav, G., Gokhale, R.S. and Mohanty, D. 2003. Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. *Journal of Molecular Biology*. **328**(2),pp.335–363.

Yun, Y.S., Lee, T.-H., Nam, G.H., Jang, D.S., Shin, S., Oh, B.-H. and Choi, K.Y. 2003. Origin of the Different pH Activity Profile in Two Homologous Ketosteroid Isomerases. *J Biol Chem*. **278**(30),pp.28229–28236.

Zazopoulos, E., Huang, K., Staffa, A., Liu, W., Bachmann, B.O., Nonaka, K., Ahlert, J., Thorson, J.S., Shen, B. and Farnet, C.M. 2003. *A genomics-guided approach for discovering and expressing cryptic metabolic pathways.*

Zhang, K. 2016. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*. **193**(1),pp.1–12.

Zhang, M.-Q. and Wilkinson, B. 2007. Drug discovery beyond the 'rule-of-five'. *Current Opinion in Biotechnology*. **18**(6),pp.478–488.

Zhang, Y.-M., Hurlbert, J., White, S.W. and Rock, C.O. 2006. Roles of the active site water, histidine 303, and phenylalanine 396 in the catalytic mechanism of the elongation condensing enzyme of *Streptococcus pneumoniae*. *J Biol Chem*. **281**(25),pp.17390–17399.

Zheng, J., Fage, C.D., Demeler, B., Hoffman, D.W. and Keatinge-Clay, A.T. 2013. The Missing Linker: A Dimerization Motif Located within Polyketide Synthase Modules. *ACS Chemical Biology*. **8**(6),pp.1263–1270.

Zheng, J., Gay, D.C., Demeler, B., White, M.A. and Keatinge-Clay, A.T. 2012. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat Chem Biol*. **8**(7),pp.615–621.

Zheng, J., Taylor, C.A., Piasecki, S.K. and Keatinge-Clay, A.T. 2010. Structural and Functional Analysis of A-Type Ketoreductases from the Amphotericin Modular Polyketide Synthase. *Structure*. **18**(8),pp.913–922.

Zheng, Q., Guo, Y., Yang, L., Zhao, Z., Wu, Z., Zhang, H., Liu, J., Cheng, X., Wu, J., Yang, H., Jiang, H., Pan, L. and Liu, W. 2016. Enzyme-Dependent 4+2 Cycloaddition Depends on Lid-like Interaction of the N-Terminal Sequence with the Catalytic Core in Pyrl4. *Cell Chemical Biology*. **23**(3),pp.352–360.

Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y. and Agard, D.A. 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*. **14**(4),pp.331–332.

Zhou, H., Gao, Z., Qiao, K., Wang, J., Vederas, J.C. and Tang, Y. 2012. A fungal ketoreductase domain that displays substrate-dependent stereospecificity. *Nat Chem Biol*. **8**(4),pp.331–333.

Zimmerman, D.E., Kulikowski, C.A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C.-Y., Powers, R. and Montelione, G.T. 1997. Automated Analysis of Protein NMR Assignments Using Methods from Artificial Intelligence. *Journal of Molecular Biology*. **269**(4),pp.592–610.

Zucko, J., Skunca, N., Curk, T., Zupan, B., Long, P.F., Cullum, J., Kessin, R.H. and Hranueli, D. 2007. Polyketide synthase genes and the natural products potential of Dictyostelium discoideum. *Bioinformatics (Oxford, England)*. **23**(19),pp.2543–2549.