**Springer Theses**
Recognizing Outstanding Ph.D. Research

Matthew Jenner

# Using Mass Spectrometry for Biochemical Studies on Enzymatic Domains from Polyketide Synthases

Springer

## Springer Theses

Recognizing Outstanding Ph.D. Research

## Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

## Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

More information about this series at http://www.springer.com/series/8790

Matthew Jenner

# Using Mass Spectrometry for Biochemical Studies on Enzymatic Domains from Polyketide Synthases

Doctoral Thesis accepted by
the University of Nottingham, UK

Springer

*Author*
Dr. Matthew Jenner
School of Chemistry
University of Warwick
Coventry
UK

*Supervisor*
Prof. Neil Oldham
University of Nottingham
Nottingham
UK

ISSN 2190-5053 ISSN 2190-5061 (electronic)
Springer Theses
ISBN 978-3-319-32722-8 ISBN 978-3-319-32723-5 (eBook)
DOI 10.1007/978-3-319-32723-5

Library of Congress Control Number: 2016936964



Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

**Parts of this thesis have been published in the following journal articles:**

***Substrate Specificity in Ketosynthase Domains from trans-AT Polyketide Synthases***

***Matthew Jenner**, Sarah Frank, Annette Kampa, Christoph Kohlhaas, Geoff S. Briggs, Jörn Piel and Neil J. Oldham.*

*Angewante Chemie International Edition, 2013, **52**, 1143–1147.*

***Amino Acid-Accepting Ketosynthase Domain from a trans-AT Polyketide Synthase Exhibits High Selectivity for Predicted Intermediate***

*Christoph Kohlhaas,[†] **Matthew Jenner**,[†] Annette Kampa, Geoff S. Briggs, Jose P. Afonso, Jörn Piel and Neil J. Oldham.*

*Chemical Science, 2013, **4**, 3212–3217.*

***A Close Look at a Ketosynthase from a trans-acyltransferase Modular Polyketide Synthase***

*Darren C. Gay, Glen Gay, Abram J. Axelrod, **Matthew Jenner**, Christoph Kohlhaas, Annette Kampa Neil J. Oldham, Jörn Piel, Adrian T. Keatinge-Clay.*

*Structure, 2014, **22**, 444–451.*

***Acyl Chain Elongation Drives Ketosynthase Substrate Specificity in Polyketide Biosynthesis***

***Matthew Jenner**, Jose P. Afonso, Hannah R. Bailey, Annette Kampa, Sarah Frank, Jörn Piel and Neil J. Oldham*

*Angwante Chemie International Edition, 2015, **54**, 1817–1821.*

***Acyl Hydrolases from trans-AT Polyketide Synthases Target Acetyl Units of Acyl Carrier Proteins***

***Matthew Jenner**, Jose P. Afonso, Christoph Kohlhaas, Petra Karbaum, Sarah Frank, Jörn Piel and Neil J. Oldham Chemical Communications, 2016, **52**,* 5262–5265

# Supervisor's Foreword

This thesis details outstanding work conducted by Matthew Jenner between 2010 and 2014 in the School of Chemistry, University of Nottingham under my supervision. It describes a mass spectrometric investigation of enzymatic domains from *trans*-acyl transferase polyketide synthases (*trans*-AT PKSs). These megasynthases are responsible for the biosynthesis of a wide variety of bioactive natural products such as pederin, psymberin, bacillaene, and kirromycin.

Unlike the more familiar *cis*-AT PKSs, which display collinearity between the domain architecture of the PKS and the structure of the polyketide product, *trans*-AT PKSs exhibit aberrant architecture and often incorporate novel enzymatic domains, resulting in poor biosynthetic assignments. An introduction to polyketide biosynthesis and to the differences between *cis*- and *trans*-PKSs is provided in Chap. 1 of the thesis.

Given that *trans*-AT systems constitute approximately 40 % of all bacterial multimodular PKSs, they represent a major, but poorly characterised, enzyme class that is of high relevance for drug discovery. Against this background, Matthew set out to develop tools that could be used to help understand these intriguing PKSs. Our starting point was a phylogenetic study by Nguyen et al. (*Nat. Biotechnol.* 2008), which predicted that—based on sequence analysis—ketosynthase domains from *trans*-AT PKSs should possess specificity for particular biosynthetic intermediate 'types' (e.g. acetyl starter units, β-hydroxy-, β-keto-, and enoyl-chains). The first objective was to develop a simple functional assay to confirm this proposed specificity. The initial KS acylation step of the enzyme's mechanism was selected as the first point to determine specificity, as this could be easily monitored by measuring the associated mass increase of the isolated KS domains. This work is described in the Chaps. 3 and 4: the first two results and discussion chapters of the thesis. We chose a variety of KSs from the psymberin and bacillaene PKSs, and a set of acyl *N*-acetylcysteamines (acyl SNACs) for use as substrate mimics. Plasmid vectors for the former were kindly supplied by Jörn Piel's group (University of Bonn, then ETH Zurich), as were the SNAC thioesters.

Chapter 3 describes the successful development and application of a acylation assay to investigate the tolerance of KS domains for beta-methyl branched

substrates. Selectivity for non-branched intermediates was seen in KSs that are proposed to process linear acyl chains. Matthew rationalised this observation at the amino acid level by use of homology models, and—in some very nice experiments—employed site-directed mutagenesis to modulate this selectivity and allow acceptance of branched substrates.

Chapter 4 goes on to describe the application of the acylation assay to a KS domain located immediately downstream of the non-ribosomal peptide synthase module within the PKS responsible for the biosynthesis of bacillaene. Here, selectivity for amino-acid-derived SNAC thioesters was seen, as predicted based on the knowledge of the expected substrate intermediate. Once again Matthew rationalised this result by homology modelling, and probed the role of key amino acid residues within the KS. Very pleasingly, his prediction of the importance of an asparagine residue in bonding to the amino-acid-derived substrates was recently confirmed by the Keatinge-Clay group using X-ray crystallography.

Chapter 5 presents a simple method for making acyl-ACPs for use in PKS enzyme assays. This allows the synthesis of more realistic substrate mimics, where the full phosphopantetheine linker chain tethers the acyl chain to the ACP. Matthew went on to use these products to probe the substrate specificity of the acyl hydrolase from the pederin PKS, and demonstrate that its major housekeeping role is probably in targeting unwanted acetyl-ACP, which may be derived from acetyl-CoA during initial activation of the PKS by the promiscuous phosphopantetheine transferases.

In Chap. 6, the development of an assay to measure the selectivity of the KS elongation step is described. This key process follows the initial acylation of the KS active site, which was probed in Chaps. 3 and 4. Thus, with these two assays in hand, Matthew was able to unpick the enzymology of KS domains by studying the selectivity of each step of the KS-catalysed reaction. Interestingly, the elongation step was found to be much more selective than the preceding acylation, and only those substrates that closely mimicked the natural intermediate of the KS were elongated.

In summary, this outstanding thesis represents a significant contribution to our understanding of KS specificity in this intriguing family of PKSs. The assays developed will prove very useful to researchers in the PKS field, and the findings will almost certainly find utility in future PKS engineering efforts.

I would like to thank Matthew for all the hard work he put into his Ph.D. project, for his infectious enthusiasm, and for producing such an excellent thesis.

Nottingham, UK
April 2016

Prof. Neil Oldham

# Abstract

Polyketides form a group of diverse and structurally complex bioactive natural products. Their biosynthesis is directed by multi-domain polyketide megasynthases (PKSs), which extend the acyl chain by a series of condensation and optional reduction steps. Phylogenetic work has shown that, in a particular group of type I systems known as *trans*-AT PKSs, the ketosynthase (KS) domains potentially harbour specificity towards the nature of the first four carbons of the intermediate substrate (e.g. beta-hydroxy, enoyl, methyl-branched). These results suggest a close link between KS evolution and substrate specificity.

This thesis reports studies on the substrate specificity of crucial KS domains from *trans*-AT PKSs. Using a combination of electrospray ionisation-mass spectrometry (ESI-MS) and simple *N*-acetyl cysteamine (SNAC) substrate mimics, the substrate specificity of a range of KS domains from the bacillaene (BaeJ and BaeL) and psymberin (PsyA and PsyD) PKSs, including a KS domain immediately downstream of a non-ribosomal peptide synthase (NRPS) module, have been successfully studied with regard to the initial acylation step of KS-catalysis. In addition, the ability to alter the substrate tolerance of KS domains by simple point mutations in the active site has been successfully demonstrated. A novel method for the synthesis of acyl-acyl carrier proteins (ACP) from SNAC thioesters is also reported. A series of acyl-ACPs have been synthesised using this methodology and successfully used to probe the substrate specificity of both KS domains and the previously uncharacterised acyl hydrolase (AH) domain, PedC.

KS-catalysed chain elongation reactions have also been conducted and monitored by ESI-MS/MS. All KS domains studied exhibited higher substrate specificity at the elongation step than in the preceding acylation. Furthermore, a mechanism of reversible acylation is proposed using the PsyA ACP1-KS1 didomain. The findings reported in this thesis provide important insights into the mechanism of KS specificity and show that mutagenesis can be used to expand the repertoire of acceptable substrates for future PKS engineering efforts.

# Acknowledgements

First and foremost, I would like to convey my utmost thanks to Dr. Neil Oldham for giving me the opportunity to work on this project, and for his constant support, encouragement, and guidance since joining the lab as a project student—It has been a truly enjoyable 5 years!

I also extend my gratitude to Prof. Jörn Piel and members of his group in both Bonn and Zürich for a fruitful collaboration. In particular: Sarah Frank, Christoph Kohlhaas, and Petra Pöplau for the synthesis of numerous compounds—sometimes on demand! I would like to thank Annette Kampa for all her work before and during her visit to Nottingham, and for providing numerous KS constructs. Additionally, thanks to Dr. Anna Vagstad for sending lots of plasmids with extremely detailed advice during my final year.

I would like to specially mention to Dr. Jose Afonso, who taught most of what I now know about molecular biology, and for allowing me to call him the name of any Latin football player I desired. Also, thanks go to Dr. Kleitos Sokratous who not only helped to fix numerous instrument-based issues, but was also a great friend and housemate during his time in Nottingham.

I am extremely appreciative to Prof. Panos Soultanas for allowing me to conduct all my molecular biology work in his lab. Furthermore, I extend my thanks to Dr. Geoff Briggs and Matt Green for their constant support and advice whilst working in CBS. I also want to thank all the past and present members of the Oldham, Soultanas, and Searle groups for making the lab such a great place to work, in particular to Jon Hopper, Lucy Roach, Richard Elms, Dan Scott, Lucio Manzi, Nkazi Tshuma, Liz Morris, Jenny Adlington, Alex Slater, Juliet Morgan, Jed Long, Vasillis Paschalis, and Sarah Northall.

Many thanks go to Matt, Alex, and Dan from the MonoQs, who have provided endless laughs and a welcome musical distraction from work. I would also like to give special thanks to Christina for her continuous patience, love, and reassurance throughout my studies. Huge thanks to Michael and Kathryn for making my holiday periods at home so entertaining—a time of year I always look forward to. Finally, I would like to dedicate this thesis to my parents on the year of their 30th wedding anniversary as a token of my gratitude for all their love, support, and encouragement during my studies, without which, none of this would have been possible.

# Contents

# Abbreviations

| | |
|---|---|
| A-Domain | Adenylation Domain |
| ACP | Acyl Carrier Protein |
| ADP | Adenosine Diphosphate |
| AL | Acyl Ligase |
| Ala | Alanine |
| AMP | Adenosine Monophosphate |
| Arg | Arginine |
| Asn | Asparagine |
| AT | Acyltransferase |
| ATP | Adenosine Triphosphate |
| C-Domain | Condensation Domain |
| CID | Collision-Induced Dissociation |
| CoA | Coenzyme A |
| CR | Crotonase |
| CRM | Charge Residue Model |
| Cy | Cyclase |
| Cys | Cysteine |
| DCC | *N-N'*-Dicyclohexylcarbodiimide |
| DMSO | Dimethylsulphoxide |
| DNA | Deoxyribonucleic Acid |
| e.q. | Equivalent |
| ECH | Enoyl-CoA Dehydratase |
| EDTA | Ethylenediaminetetraacetic Acid |
| ER | Enoylreductase |
| ESI | Electrospray Ionisation |
| FAS | Fatty Acid Synthase |
| Gln | Glutamine |
| Glu | Glutamic Acid |
| Gly | Glycine |
| GNAT | GCN5-related *N*-acetyltransferase |
| His | Histidine |

| | |
|---|---|
| HMGS | 3-hydroxy-3-methylglutaryl-CoA |
| IEM | Ion Evaporation Model |
| IPTG | Isopropylthio-β-galactoside |
| KR | Ketoreductase |
| KS | Ketosynthase |
| LB | Luria Broth |
| LC | Liquid Chromatography |
| Leu | Leucine |
| Lys | Lysine |
| Met | Methionine |
| MS/MS | Tandem Mass Spectrometry |
| MS | Mass Spectrometry |
| MT | Methyltransferase |
| NMR | Nuclear Magnetic Resonance |
| NRPS | Non-Ribosomal Peptide Synthase |
| Phe | Phenylalanine |
| PKS | Polyketide Synthase |
| PPant | Phosphopantetheine |
| PPTase | 4'-Phosphopantetheinyl Transferase |
| Pro | Proline |
| Ser | Serine |
| SNAC | *N*-Acetyl Cysteamine |
| TE | Thioesterase |
| TFA | Trifluoroacetic Acid |
| Thr | Threonine |
| TOF | Time of Flight |
| Trp | Tryptophan |
| Tyr | Tyrosine |
| Val | Valine |

# Chapter 1
# Introduction

## 1.1 Natural Products

The realm of natural products and their derivatives has provided the most successful source of bioactive drug molecules for generations [1]. The medicinal use of natural products can be dated back to 2600 B.C. in Mesopotamia, where the earliest records of natural products are documented on clay tablets, including simple oils such as *Commiphora myrrha* (myrrh) to treat colds and inflammation [2]. Since then, many bioactive compounds have been discovered including salicin (the precursor to aspirin (**1**), from *Salix alba*), morphine (**2**) (from *Papaver somniferum*) and the anti-malarial quinine (**3**) (from *Cinchona succirubra*). However, arguably the most notable natural product discovery is that of penicillin (**4**) from the fungus *Penicillium notatum*, which was credited to Alexander Fleming in 1929, but with major contributions from Howard Florey, Edward Penley Abraham, and Ernst Chain. Since this breakthrough, the introduction of antibiotics such as tetracycline (**5**) and kanamycin (**6**) has decreased the mortality rate from bacteria-induced diseases drastically (Fig 1.1) [3].

Despite the positive impact antibiotics have had on human health, their misuse combined with global travel has allowed strains of pathogenic bacteria to acquire resistance to these compounds, requiring the need for novel antibiotic agents [4]. Since the early 80's natural products or their derivatives have comprised ~40 % of new chemical entities, and represent 70 % of current antibiotic compounds [5]. Despite this, in recent years pharmaceutical drug discovery programmes have tended to focus upon combinatorial libraries of small fragment molecules. It is estimated that <10 % of the worlds biodiversity has been investigated to date for potential biological activity, therefore a metaphorical chemical vault of natural products remains undiscovered. Accessing this chemical diversity remains the current challenge for researchers.



M. Jenner, *Using Mass Spectrometry for Biochemical Studies on Enzymatic Domains from Polyketide Synthases*, Springer Theses,
DOI 10.1007/978-3-319-32723-5_1

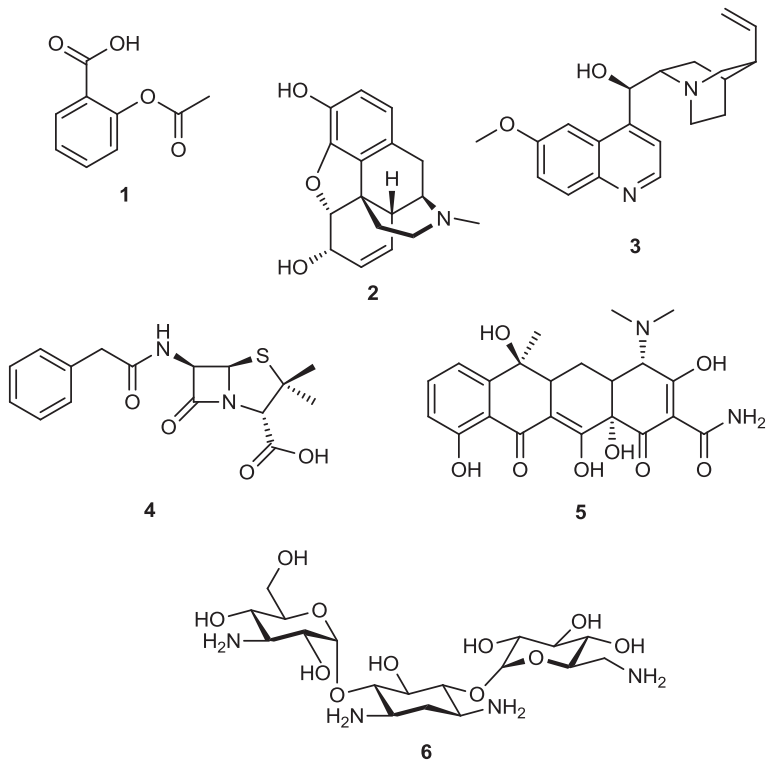



**Fig. 1.1** Aspirin (**1**), Morphine (**2**), Quinine (**3**), Penicillin G (**4**), Tetracycline (**5**), Kanamycin (**6**)

### *1.1.1 Polyketides and Non-ribosomal Peptides*

Polyketides (PK), non-ribosomal polypeptides (NRP) and their hybrids (PK-NRP) represent a diverse class of natural products that exhibit a wide range of pharmaceutical activities (Fig. 1.2). Such biological activity includes: antibiotic (Erythromycin A (**7**)), immunosuppressant (FK-506 (**8**)), antifungal (Amphotericin B (**9**)), antitumour (Geldanmycin (**10**)) and hypolipidemic agents (Lovastatin (**11**)). With such an extensive range of therapeutic biological activities, natural product research is constantly working to discover novel polyketide structures capable of enhancing human health.

Although many of these highly sought-after natural products are, in principle, available from sources in nature, obtaining sufficient quantities to characterise their potential therapeutic activity is often difficult [6]. Moreover, efficient chemical synthesis of these complex compounds is extremely challenging. Investigating the biosynthetic mechanisms responsible for such attractive compounds has allowed the chemistry of polyketide-producing enzymes to be harnessed for the production of novel, biochemically bioactive natural products [7, 8].

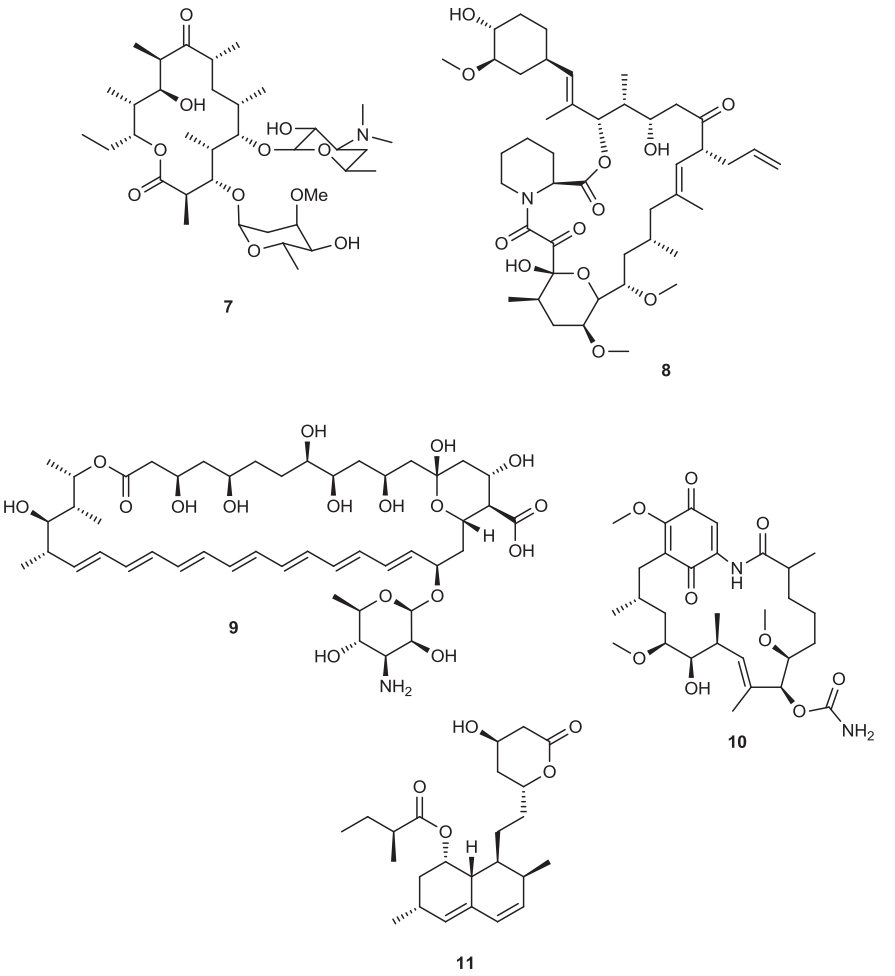
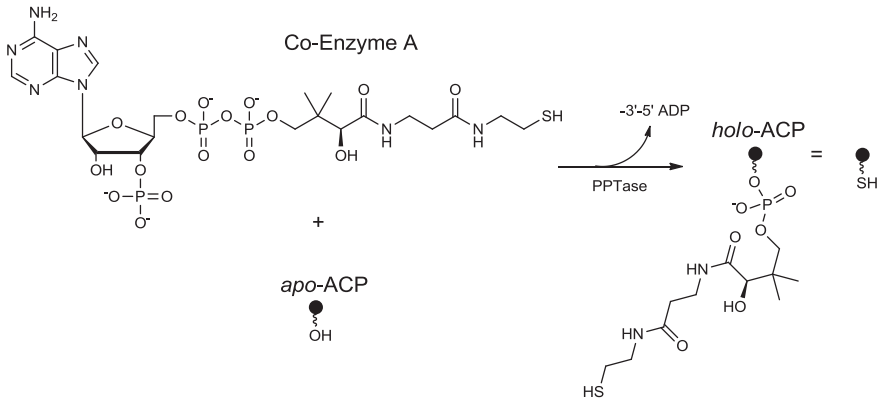



**Fig. 1.2** Erythromycin A (**7**), FK-506 (**8**), Amphotericin B (**9**), Geldanmycin (**10**), Lovastatin (**11**)

This thesis reports efforts to better understand the enzymology of key proteins involved in polyketide biosynthesis, and how understanding the specificity of these enzymes can assist the discovery of novel natural products.

## 1.2 Polyketide Synthases

Polyketide synthases (PKSs) are large protein complexes responsible for the biosynthesis of an array of complex, biologically active compounds, many of which are employed in medicine. Several classes of PKSs exist, each with an architectural variation on a common mechanistic theme.




**Scheme 1.1** Post-translational attachment of phosphopantetheine from co-enzyme A to a conserved serine residue on an *apo*-ACP, catalysed by a PPTase

### *1.2.1 Polyketide Biosynthesis*

Despite the structural diversity of polyketides, the building blocks of these compounds are simple acyl-CoAs, and the biosynthetic logic is closely associated with that of fatty acid synthases (FASs) [9, 10]. In both polyketide and fatty acid biosynthesis, the growing chain is covalently tethered to an acyl carrier protein (ACP) via a phosphopantetheine (PPant) moiety. The PPant chain, derived from CoA, is attached post-translationally to a conserved serine residue on the *apo*-ACP by a phosphopantetheinyl transferase (PPTase), yielding the mature *holo*-ACP (Scheme 1.1) [11, 12].

The *holo*-ACP, in conjunction with an acyltransferase (AT) and a ketosynthase (KS), forms the minimal set of enzymatic domains required for a single chain elongation reaction. This set of enzymatic domains is often termed a module. Polyketide biosynthesis is initiated by the loading of a starter unit, derived from an acyl-CoA, onto the PPant arm of the ACP [12, 13]. In fatty acid biosynthesis the starter unit is typically limited to acetyl, whereas PKSs can utilise a variety of primer units including acetyl-, propionyl-, butyryl and other variants [10]. The role of the AT domain is to catalyse the transfer of a malonyl-derived extender unit from a CoA-thioester onto the thiol of the ACP phophopantetheinyl moiety (Fig. 1.3).

The elongating carbon-carbon bond formation is then achieved by Claisen condensation chemistry catalysed by the KS domain. Further optional modification at the β-keto position is directed by the presence of ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains, generating β-hydroxyl, olefinic and fully saturated intermediates respectively (Fig. 1.4) [12, 14, 15].

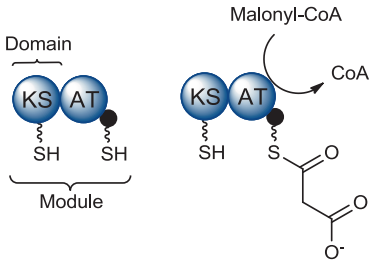



**Fig. 1.3** The 'minimal module' showing the enzymatic domains required for a single cycle of chain elongation (*left*). The role of the AT domain is to selectively load malonyl-derived units onto the thiol of the ACP (*right*)

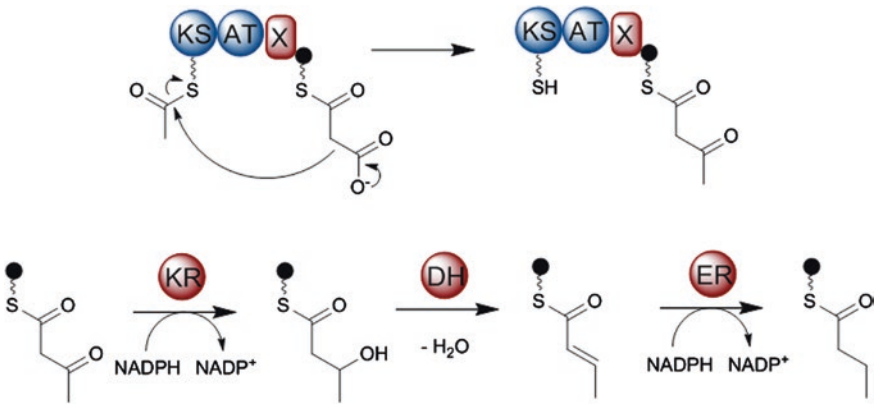



**Fig. 1.4** The Claisen condensation between an acetyl and malonyl unit catalysed by a KS domain (*top*). The X represents optional reductive domains which can alter the functionality at the β-keto position, producing β-hydroxyl, olefinic and fully saturated intermediates

The growing polyketide chain passes from the ACP to the KS of the following downstream module, where further elongation and modification is carried out. The polyketide chain is therefore translocated from module to module until the final polyketide structure is assembled. Following the final elongation cycle, a thioesterase (TE) domain catalyses hydrolysis or lactonisation, which releases the polyketide chain from the PKS (Fig. 1.5) [16].

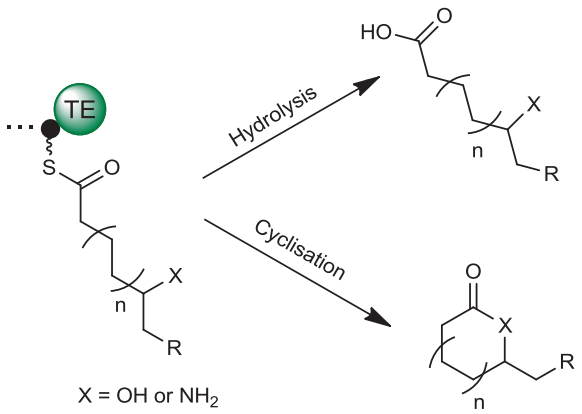



**Fig. 1.5** Reactions catalysed by a TE domain. Hydrolytic cleavage releases the product as a free acid, whereas the cyclisation reaction can produce lactone and lactam rings of varying sizes

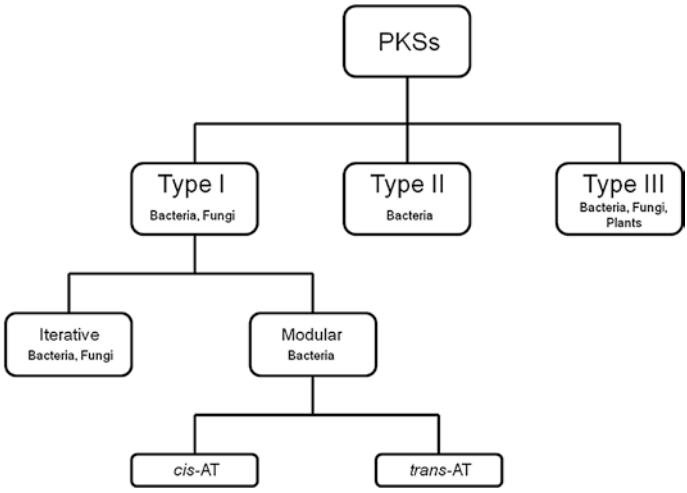



**Fig. 1.6** Tree diagram showing the different types of PKSs. Further subdivisions are shown for type I PKS systems

The released product is often merely a precursor to the bioactive compound, with the activity of tailoring enzymes required to produce the final product. Examples of post-PKS modifications include glycosylations, methylations and hydroxylations.

### *1.2.2 Polyketide Synthase Classification*

As previously mentioned, PKSs can be grouped into different 'classes' on the basis of their molecular architecture and to some extent their mode of action. The major divisions are into types I, II and III (Fig. 1.6) [17, 18]. Type I PKSs, which can be further subdivided, are primarily found in bacteria and fungi, whereas Type II PKSs are exclusive to bacteria. Type III PKSs are predominantly observed in plants, however some examples are known in bacteria and fungi.

#### 1.2.2.1 Type I PKSs

As highlighted in Fig. 1.6, the class of type I PKSs can be further sub-divided based purely on enzymatic architecture. The iterative type I PKSs are found in both bacteria and fungi, whereas the modular type I PKSs are exclusive to bacteria.

Iterative Type I PKSs

Iterative type I PKSs employ a single module to carry out repetitive elongation steps. The extent of β-keto processing is dictated by the presence of reductive domains within the module. Therefore, within this class of PKSs products are referred to as non-reducing (NR), partially reducing (PR) and highly reducing (HR). Although confined to a single module, the level of reduction may not be the same for each round of elongation, as some reductive domains are used optionally [17]. The example shown in Fig. 1.7 depicts the biosynthesis of compactin (**12**). In this instance five units of malonyl-CoA are required to build the carbon backbone. The KR and DH domains are used in every iteration, whereas the ER is utilised only twice. The condensation domain catalyses a Diels-Alder reaction forming the cyclised core, and the methyl-transferase (MT) domain is used to install the methyl branch of the α-methylbutyryl ester moiety [19].

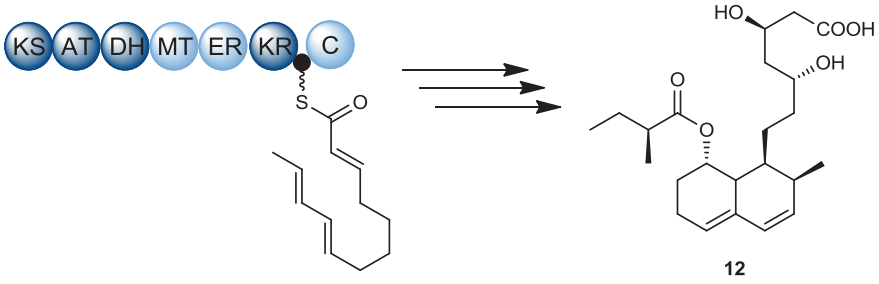



**Fig. 1.7** Biosynthesis of compactin (**12**) from an iterative type I PKS. *Light blue* domains are not used for every iteration of the biosynthesis

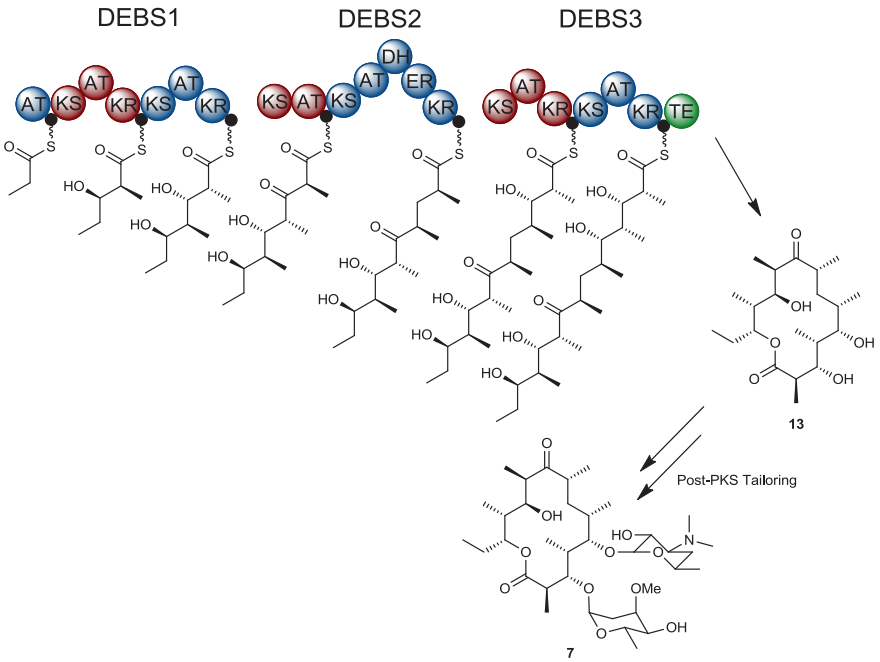



**Fig. 1.8** Biosynthesis of erythromycin A (**7**) by the DEBS system, an example of a modular type I PKS. The three DEBS proteins harbour six elongating modules and a loading module. The TE domain catalyses a macrolactonisation releasing 6-dEB (**13**), which is further modified to yield (**7**)

## Modular Type I PKSs

Modular type I, or non-iterative PKSs, have been likened to molecular assembly lines, where multiple sequential modules are harnessed to produce their polyketide product. Here, each module harbours the enzymatic domains required to catalyse a single cycle of chain elongation and optional reductions [20]. A complete modular type I PKS is usually comprised of two or more large polypeptide chains complexed together, leading to overall molecular weights on the MDa scale. The 6-deoxyerythronolide B synthase (DEBS) system, responsible for the biosynthesis of erythromycin A, is a comprehensively studied example of such a PKS, and is highlighted in Fig. 1.8 [21]. In this system, the biosynthesis is initiated by the loading of a propionyl unit onto the first ACP by a loading module comprised of an AT and an ACP. The subsequent AT domains load methylmalonyl extender units onto the ACPs for chain extensions. After six successive chain elongations with varying degrees of reductive processing, a TE domain catalyses the macrolactonisation of the nascent polyketide chain resulting in the release of 6-deoxyerythronolide B (6-dEB) (**13**), the precursor to the antibiotic erythromycin A (**7**) [22].

### The Principle of Colinearity

The multi-modular nature of PKSs and associated sequential modification of the polyketide intermediates in modular type I PKS systems allows effective prediction of the final product from knowledge of the domain order. This direct correlation between domain organisation and polyketide product is referred to as colinearity [23]. Post-genomic techniques, and the propensity for the modular type I PKS genes to arrange into distinct modules, have allowed the construction of biosynthetic schemes from genetic sequences alone [9, 24]. The colinearity exhibited between PKS architecture and polyketide structure has greatly assisted many of the breakthroughs in this field.

### *cis*- and *trans*-AT PKSs

Relatively recently a class of modular type I PKSs have been discovered which do not conform to the textbook colinearity rules [15]. In these particular systems, the domain organisation rarely correlates with the polyketide product due to the high diversity of non-canonical modules that often contain novel enzymatic domains. The most noticeable feature of these clusters, however, is the absence of integral AT domains within the PKS. Instead, the AT activity is supplied by a reduced number of free-standing enzymatic domains servicing every module in the entire cluster (Fig. 1.9) [25]. These systems are therefore termed '*trans*-AT' PKSs, of which the pederin PKS is an example (Fig. 1.10) [26]. In contrast, the 6-dEB system is an example of a '*cis*-AT' PKS, where the AT domains are integral to the PKS, with an AT domain within each module (Fig. 1.8).

Phylogenetic studies have shown that the *cis*- and *trans*-AT PKSs have evolved independently from FASs and, although both systems appear very similar, the *trans*-AT PKSs harbour far more enzymatic functionality than their *cis*-AT counterparts [27, 28]. As a result, many more module variations exist for *trans*-AT systems. Eight module variants are known for *cis*-AT PKSs: KS-AT-ACP, KS-AT-KR-ACP, KS-AT-DH-KR-ACP, KS-AT-DH-KR-ER-ACP and their MT-based analogues. In contrast, over 50 module architectures have already been identified in *trans*-AT PKSs [15].

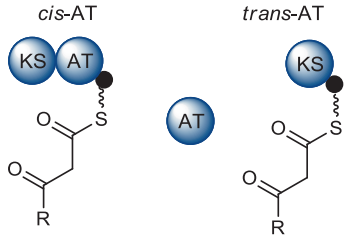



**Fig. 1.9** The module structure for *cis*- and *trans*-AT PKSs systems

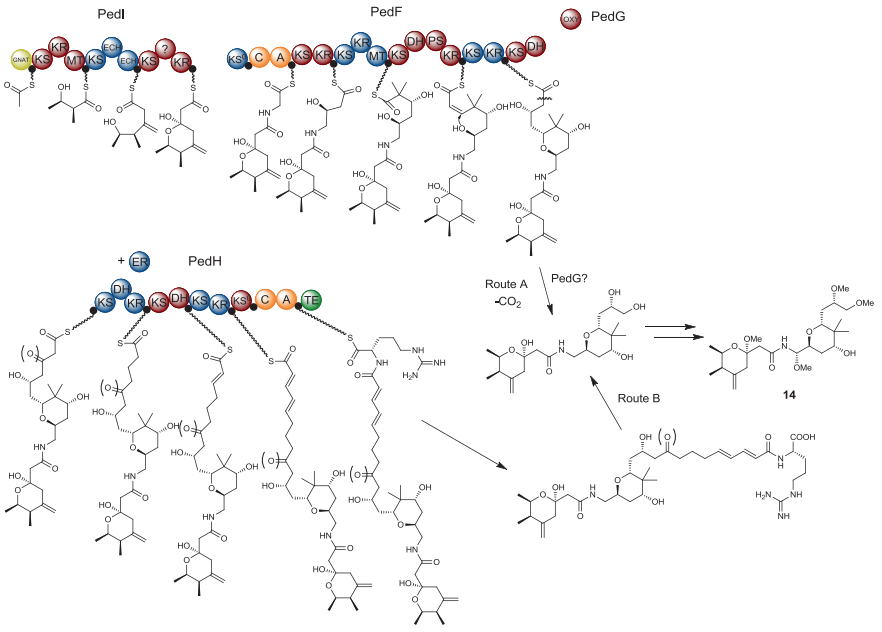



**Fig. 1.10** The proposed biosynthesis of pederin (**14**), an example of a *trans*-AT PKS system

A feature that is archetypal of *trans*-AT PKSs is the presence of non-elongating KS domains ($KS^0$), which usually lack a crucial catalytic histidine residue required for elongation. These KS domains are often located within modules split over two proteins, and within the domain sequence KS-KR-ACP-$KS^0$//DH-ACP that generates *Z*-configured double bonds [15, 27].

## KS Specificity-Based Assignment of *trans*-AT PKSs

The *trans*-AT PKSs exhibit aberrant architecture and often incorporate novel enzymatic domains, resulting in poor biosynthetic assignment using colinearity rules. Recent phylogenetic work by Nguyen et al. established that KS domains from *trans*-AT systems harbour, at the sequence level, specificity towards their predicted biosynthetic intermediates [14]. This finding offers the exciting potential for KS-based PKS-to-product assignment for *trans*-AT systems. The extent of evolutionary-based substrate specificity is predicted to extend as far as the β-position for each biosynthetic intermediate; and, although the KS specificity method works remarkably well in silico, it lacks functional testing [29].

The KS substrate-specificity method for assigning novel PKS clusters was tested in the same study for the structural prediction of a novel polyketide with a single amide unit, known as thailandamide A (**15**). In this study, the colinearity method and the KS substrate-specificity method were both used to assign the

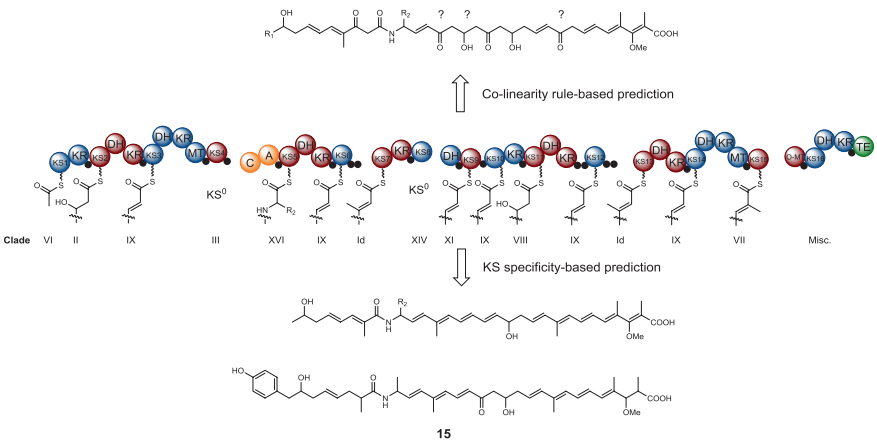



**Fig. 1.11** Comparison of colinearity and KS-specificity approaches to predicting the structure of thailandamide A (**15**). The KS-specificity method generates a better prediction for *trans*-AT systems. Reproduced from [14]

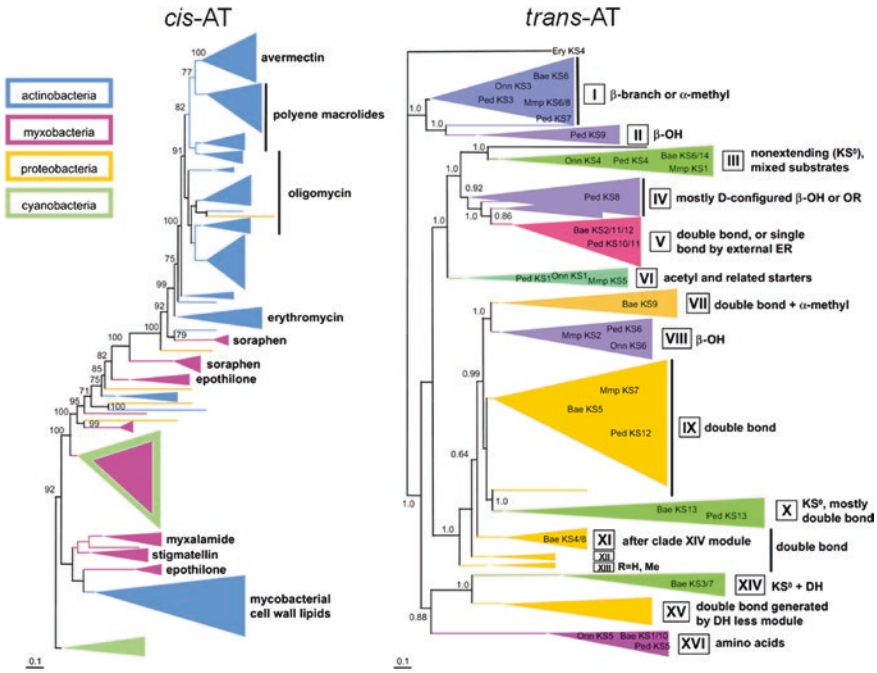



**Fig. 1.12** Bayesian phylograms of KS domains from *cis*-AT and *trans*-AT PKS systems. The KS domains from *cis*-AT PKSs cluster with respect to their biosynthetic product, whereas the *trans*-AT KS domains cluster with respect to their individual predicted substrate. Adapted from [29]

cluster. As shown in Fig. 1.11, the use of colinearity rules resulted in an incorrect structure with numerous errors, as the method is unable to account for the aberrant enzymatic architecture and mechanisms. However, using the KS substrate-specificity method, a configuration close to the NMR determined structure was obtained.

The Bayesian phylograms shown in Fig. 1.12 highlight the substrate-specific clades of *trans*-AT KS domains. On the left is the cladogram resulting from the same bioinformatic treatment of the KS domains from *cis*-AT PKSs, constructed by Dittmann and co-workers [30]. Here, it was observed that the KS domains from *cis*-AT PKSs group according to their biosynthetic pathway; in stark contrast to the KS domains from *trans*-AT PKSs. From this result it has been inferred that *cis*- and *trans*-AT PKSs arose via different evolutionary pathways. The *cis*-AT PKSs are thought to have derived from the duplication of entire lengths DNA encoding complete modules. In contrast, the *trans*-AT PKSs are believed to be a mosaic of DNA fragments, acquired from horizontal gene transfer events [14]. Horizontal or lateral gene transfer is defined as the transfer of genetic material between different organisms, and is extremely common in bacteria. The presence of multiple phyla within even the smallest enzymatic clades in the *trans*-AT cladogram (Fig. 1.12) is highly indicative of this evolutionary mechanism [31, 32].

### 1.2.2.2 Type II PKSs

The type II PKSs are found exclusively in bacteria, and adopt an iterative mechanism. However, unlike iterative type I PKSs, the catalytic domains are all expressed on different genes. Characteristic of this particular PKS group is the 'minimal PKS', comprised of two KS domains; $KS_\alpha$ and $KS_\beta$, where $KS_\beta$ is often

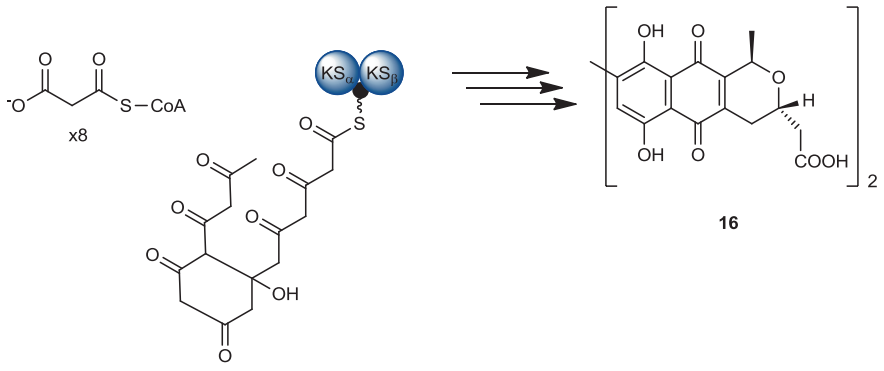



**Fig. 1.13** Biosynthesis of actinorhodin (**16**), an example of an iterative type II PKS. During the biosynthesis eight malonyl-CoA units are sequentially condensed to produce half of the dimeric actinorhodin molecule

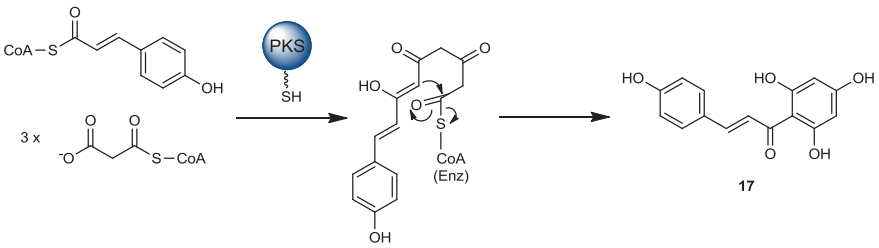



**Fig. 1.14** Biosynthesis of naringenin chalcone (**17**), an example of a type III PKS. This PKS system uses 4-coumaroyl-CoA as a starter unit, and condenses three malonyl-CoA units to produce the final product (**17**)

referred to as the 'chain length factor' (CLF) dictating the length of the polyketide chain [33]. Optional domains such as ketoreductases (KR), cyclises (CYC) and aromatases (ARO) govern the modification and folding of the intermediates during biosynthesis. Figure 1.13 depicts the biosynthesis of actinorhodin (**16**) via successive condensation of eight malonyl-CoA units. The action of optional tailoring domains yields the final product **16** [34].

### 1.2.2.3 Type III PKSs

Type III PKSs are all-in-one systems which are able to select the specific starter unit, direct the elongation and catalyse the necessary cyclisation reactions, usually generating small phenol or napthol rings. Although originally thought to be found in plants alone, examples of such systems have been identified in fungi and bacteria [35]. Figure 1.14 shows the biosynthesis of naringenin chalcone (**17**), a precursor to flavonoid and aureusidin biosynthesis, by selection of 4-coumaroyl-CoA as the starter unit and subsequent elongation with three malonyl units. Cyclisation of the polyketide chain yields the final product [36].

### 1.2.2.4 Non-ribosomal Peptide Synthases

Much like modular type I PKSs, non-ribosomal peptide synthases (NRPS) form giant modular megasynthases responsible for the production of many secondary metabolites in bacteria and fungi. However, whereas PKSs use malonyl-derived units to construct complex polyketide chains, NRPSs produce metabolites by linking amino acid units together. The substrate pool for NRPSs is vast, in addition to the 20 proteinogenic L-amino acids; NRPs are also constructed from non-proteinogenic amino acids. Approximately 500 amino acid derivatives have been found in nature, of which the 20 proteinogenic amino acids comprise just 4 % [37]. Once post-NRPS modifications and hybrid systems have been considered (see Sect. 1.2.2.5), the vast structural diversity of NRPSs becomes apparent.

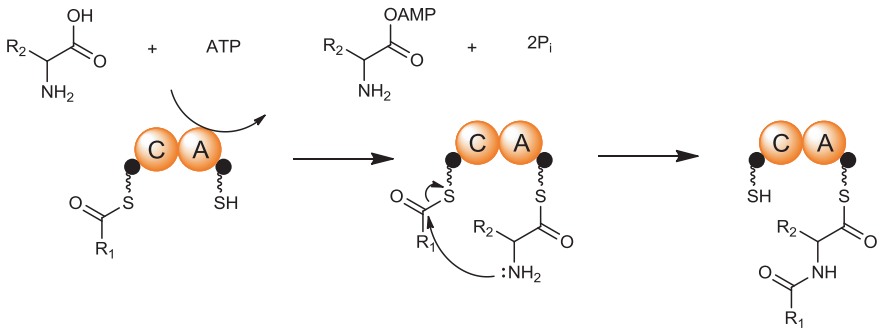



**Fig. 1.15** Mechanism of NRPS domains. The A domain selectively adenylates an amino acid using ATP, followed by condensation of two amino acid units to form a peptide bond

The biochemical logic of NRP biosynthesis is very similar to that of PKSs, utilising a carrier protein-tethered thioester to activate the acyl group of each monomer unit and transfer the elongated chain between catalytic domains. Whilst the monomer units in PKS assembly lines are acyl-CoA thioesters, the monomer units for NRPSs are proteinogenic and non-proteinogenic amino acids. Since there is no evidence for aminoacyl-CoAs in biological systems, aminoacyl-AMP mixed anhydrides are used instead [13].

A minimal NRPS module consists of an adenylation domain (A), condensation domain (C) and a peptidyl-carrier protein (PCP). In the first instance, the substrate-specific adenylation domain activates the carboxyl region of the amino acid with ATP, forming the mixed acyl-phosphoric acid anhydride with AMP, followed by loading onto the phosphopantheteine moiety of the PCP. The condensation domain subsequently catalyses the nucleophilic attack of the amino group of the previously activated amino acid, to the carbonyl of the tethered acyl group from the previous module [38, 39]. This results in the formation of a new peptide bond between the two units (Fig. 1.15).

Optional domains can be incorporated into a module for further modification to the intermediate. For example, a cylisation (Cy) domain can be used to produce thiazoline or oxazoline intermediates, which can be further modified by an oxidation (Ox) domain to yield the thiazole and oxazole products [40, 41]. Epimierisation (E) domains are also a common feature, which catalyse the stereochemical inversion of L-amino acids to D-amino acids [42]. Finally, following the last module, a thioesterase (TE) domain either hydrolyses or cyclises the intermediate to yield the NRP product [16].

### 1.2.2.5 Hybrid NRPS-PKS Systems

Despite the enzymatic differences between NRPS and PKSs, the underlying tenet of covalently tethered thioesters and the biosynthetic logic of assembly, allow the two

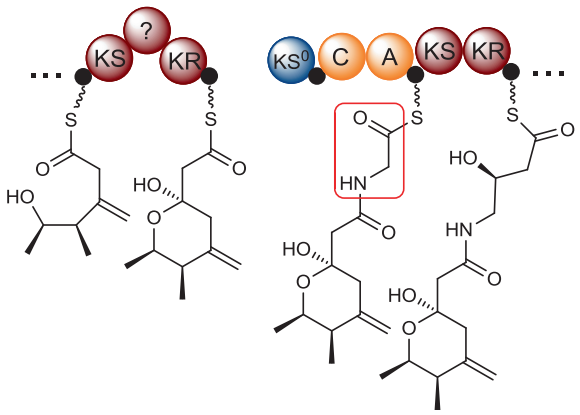



**Fig. 1.16** Partial biosynthesis of pederin (**14**) showing the integrated NRPS module, which incorporates a glycine unit into the growing polyketide chain, highlighted in *red*

systems to be integrated. Such clusters are referred to as NRPS-PKS hybrids, and are responsible for the biosynthesis of many bioactive compounds such as the immunosuppressant FK-506 (**8**) and the anti-tumour compound epothiolone [43, 44]. It is worth noting that NRPS modules are a regular occurrence in *trans*-AT PKSs, another factor that increases the structural diversity in these systems [27]. An example of this is shown in Fig. 1.16, which shows a part of the biosynthesis of pederin (**14**), where the NRPS module incorporates a glycine unit into the polyketide chain [26].

### 1.2.2.6 Bacillaene: Properties, Structure and Biosynthesis

The antibiotic activity of bacillaene (**18**) was first identified in the mid-90s from extracts of a *Bacillus subtilis* strain. Although the mechanism of action involves the inhibition of protein synthesis, the actual target and precise mechanism of action are still unknown [45]. Due to the highly unstable nature of **18**, the compound has only recently been fully characterised. The *pksX* gene cluster from *Bacillus subtilis* 168, attributed to the biosynthesis of **18,** is ~80 kb in size. A highly analogous gene cluster is also found in the genome of *Bacillus amyloliquefaciens* FZB 42 termed the *bae* PKS [46, 47]. Insights into the various intermediates of the Bae PKS were assigned using a strain of *Bacillus amyloliquefaciens* FZB 42 with the terminal TE domain deleted. From the extract of this strain, 13 stalled intermediates were detected which had been hydrolytically released from the PKS. This work allowed the functional elucidation of many aberrant features of *trans*-AT PKSs [48, 49].

The gene cluster consisting of BaeJ, BaeL, BaeM, BaeN and BaeR produces dehydrobacillaene. It harbours two NRPS modules, one in BaeJ and the other located in BaeN, which incorporate glycine and alanine amino acids respectively (Fig. 1.17). The cluster also encodes for three free-standing AT domains, hence the assignment of the cluster as a *trans*-AT PKS. One of the *trans*-acting AT domains, BaeE, is fused to an ER domain that is thought to act upon the dehydrated

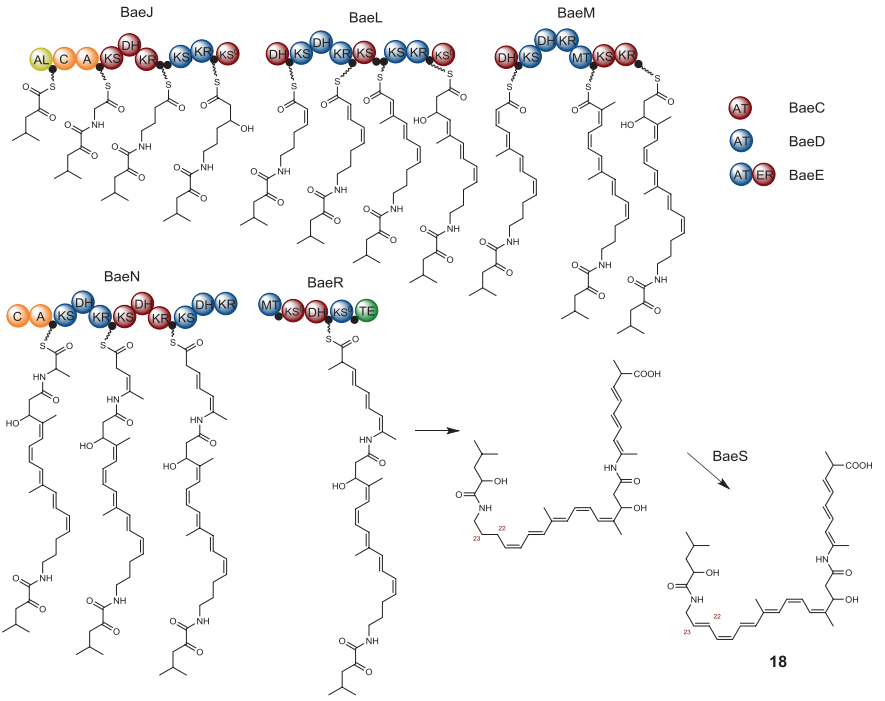



**Fig. 1.17** Proposed biosynthesis of bacillaene (**18**)

intermediate of module 2 [15]. BaeB belongs to the acyl-hydrolase (AH) clade of AT domains, responsible for the hydrolytic release of stalled intermediates, however no such activity was observed during in vitro studies [50].

Similar to many *trans*-AT PKSs systems, the *bae* PKS exhibits many unusual modules and enzymatic domains. Possibly the most intriguing modification is the installation of a β-methyl branch between KS5 and KS6 in the BaeL protein. Here, the transformation known as β-branching commences with the decarboxylation of malonyl-ACP by a non-elongating KS, to yield acetyl-ACP. A 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) then catalyses an aldol addition between the acetate unit and the growing polyketide chain, followed by dehydration and decarboxylation by two enoyl-CoA dehydratases (ECH), outlined in Fig. 1.18 [51].

Ultimately, the released dehydrobacillaene product is oxidised by BaeS, a cytochrome P450 enzyme, to yield bacillaene, **18**.

### 1.2.2.7 Psymberin: Properties, Structure and Biosynthesis

Psymberin (**19**) is a cytotoxic compound isolated from the sea sponge *Psammocinia* aff. *Bulbosa* by Crews and co-workers in 2004 [52]. The compound has been shown to have a cytostatic effect against solid tumors, with the dihydroisocoumarin region

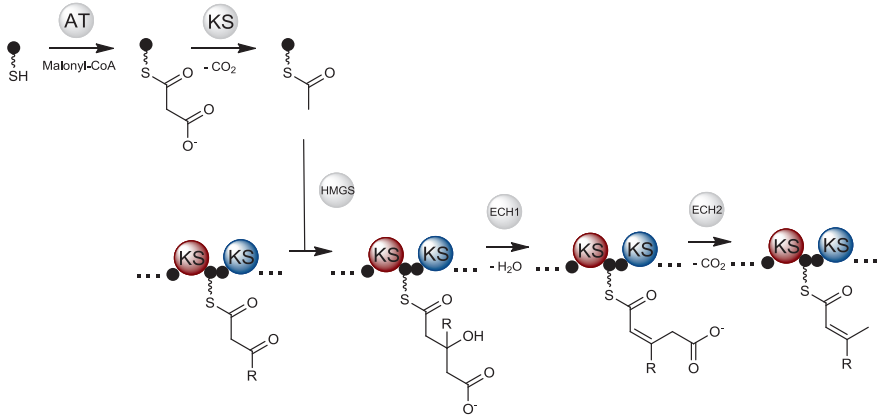



**Fig. 1.18** Proposed mechanism of β-branching during bacillaene biosynthesis. The same enzymatic scheme can be applied to form an exomethylene branch

of the molecule essential for anti-tumour function [53]. Despite the source organism being uncultivable, a metagenomic approach employed by the Piel lab has enabled the prediction of a biosynthetic scheme for the natural product [54].

The *trans*-AT PKS cluster consists of two genes, *psyA* and *psyD* which between them, encode for thirteen modules. The PsyA protein commences with a GCN5-releated *N*-acetyltransferase (GNAT) domain, which catalyses the decarboxylative loading of malonyl-CoA onto the phosphopantetheine arm of the downstream ACP, initiating psymberin biosynthesis. This catalytic activity is almost exclusively found in *trans*-AT PKS clusters, despite some examples from *cis*-AT systems. The first module in PsyA contains two putative crotonase (CR) domains, responsible for installing a β-methylene moiety. PsyD contains ten modules, one of which is a glycine-incorperating NRPS module, and is therefore has the highest reported number of PKS modules in a single protein (Fig. 1.19) [54]. Module seven harbours a pyran synthase (PS) domain, which catalyses the formation of a cyclic ether. Recently the pederin PS domain was experimentally demonstrated to possess this activity [55].

### *1.2.3 Structure and Mechanisms of PKS Domains*

PKSs harbour an array of catalytic domains which, mechanistically, are relatively well understood. Within this section the structural and catalytic features of core PKS domains will be discussed.

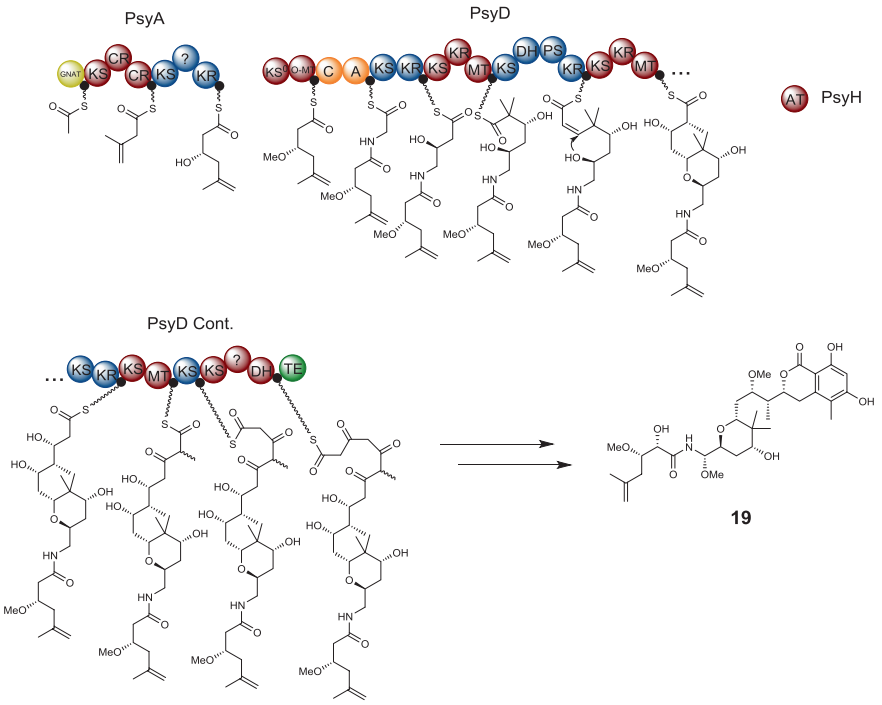



**Fig. 1.19** The proposed biosynthesis of psymberin (**19**)

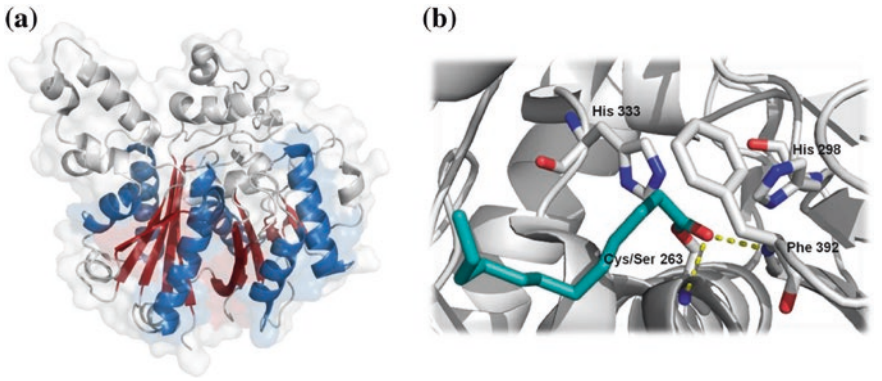



**Fig. 1.20** Structural analysis of the ketosynthase domain. **a** Crystal structure of the *E.coli* FAS KS showing the conserved thiolase fold in *blue* and *red*. Structure rendered in PyMol from PDB: 1EK4. **b** The active site of dodecanoyl-FAS KS, showing the Cys-His-His catalytic triad. The carbonyl of the acyl chain is stabilised by two backbone amides from Ser263 and Phe392, forming the oxyanion hole

### 1.2.3.1 The Ketosynthase Domain

The ~430 residue ketosynthase (KS) domain catalyses the carbon-carbon bond forming Claisen condensation reaction required for the chain elongation step in polyketide and fatty acid biosynthesis [56]. All published crystal structures of KS domains from both FASs and PKSs exhibit a highly conserved thiolase fold, forming a five-layered central core to the enzyme [57, 58]. The conserved fold of the KS domain can be simplified to a 2α-5β-2α-5β-2α arrangement, which in essence is two five-stranded β-sheets sandwiched by three sets of α-helices (Fig. 1.20a). KS domains are generally dimeric and their interface regions contribute significantly to the dimerisation of the type I PKSs (see Sect. 1.2.4). The majority of KS domains employ a Cys-His-His catalytic triad active site, located in conserved TA**C**SSS, EA**H**GTG and KSNIG**H**T motifs [56].

Although the mechanism of action for both the transthioesterification and condensation reactions has not been conclusively elucidated, insights into both reactions have been gained from crystal structures of a FAS KS acylated with various fatty acid chains (PDB:1EK4) and of *E.coli* FabF with platensimycin (PDB:2GFX) [59, 60]. The acylated FAS KS structures were obtained by mutation of the active site Cys to a Ser which bind the acyl substrate more efficiently than the WT, probably due to the increased stability of an ester over a thioester. Here, decanoyl and dodecanoyl-CoAs (CoA harbours a phosphopantetheine and can therefore bind in the KS active site) were soaked into the crystals allowing an acylated KS structure to be obtained. From the structures obtained the role of the oxyanion hole in stabilising the acyl-KS was revealed, however, the crucial decarboxylation step was postulated to release $CO_2$, not $HCO_3^-$, which has since been shown to be the eliminated $C_1$ unit due to the presence of an activated water molecule in the KS active site [61, 62].

In the case of the platesimycin-FabF complex, the active site Cys was mutated to a glutamine residue, effectively mimicking an acylated state. The crystal

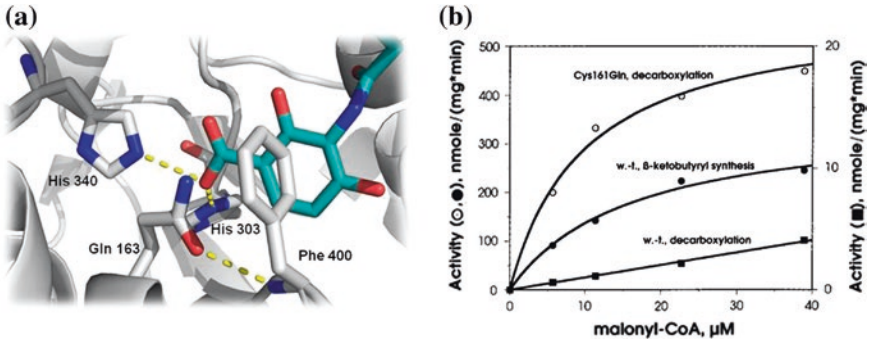



**Fig. 1.21** **a** Structure of the platesimycin-FabF complex showing the active site Cys mutated to a Gln, effectively mimicking an acylated state. The terminal carbonyl of platesimycin can be seen to form contacts with the catalytic His residues. Rendered in PyMol from PDB: 2GFX **b** Enzymatic plot showing that Cys → Gln KS mutants are potent decarboxylase enzymes. Taken from [61]

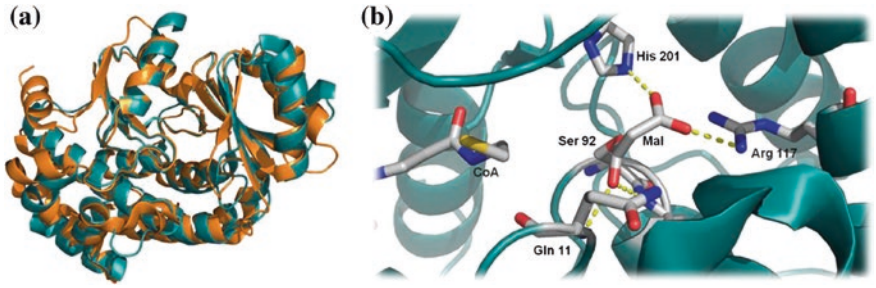



**Fig. 1.22** Structural analysis of the acyltransferase domain. **a** Structure overlay of PKS (*orange*) and FAS (teal) AT domains. The structural similarity indicates the high homology between the systems. **b** The FAS AT domain with malonate bound to the active site serine. The carbonyl sits in an oxyanion hole formed by two backbone amides, and an important set of interactions from His201 and Arg117 stabilise the malonate unit. Rendered in PyMol from PDB: 2G2Z

structure revealed that the carbonyl of the glutamine side chain was hydrogen bonded to the two backbone amides in the oxyanion hole in agreement with the 1EK4 structure [60]. In addition, the terminal carboxyl of platesimycin is bound by the two histidine residues (Fig. 1.21a). Using this structural data in conjunction with biochemical data, it has been proposed that only one of the histidine residues is catalytically important in the activation of the carboxylate extender unit and subsequent stabilisation of the enolate tautomer. The other histidine residue is thought to act as a general base, activating a water molecule allowing nucleophilic attack upon the carboxylate [63].

KS domains act as decarboxylases when the active site Cys is mutated to a Gln, or in an acylated state (Fig. 1.22b) [61]. It is believed that the co-ordination of the carbonyl in the oxyanion hole aids the organisation of residues in the active site for effective decarboxylation and elongation [64]. The by-product of the decarboxylation reaction is, however, bicarbonate not carbon dioxide as previously postulated.

The current KS mechanistic hypothesis is as follows (Scheme 1.2):

*I - Acylation*

The transthioesterification reaction commences when an ACP-tethered acyl chain enters the substrate binding channel of the KS. The binding channel of the KS domain is split into two distinct regions; the substrate-binding and the phosphopantetheine-binding areas. Although the amino acid sequence for the phosphopantetheine-binding region is extremely well conserved, even in FAS KS domains, the intricacies of the substrate-binding pocket are poorly understood. The $pK_a$ of the active-site cysteine is lower than usual for a thiol due to the positive charge generated by the helix-dipole, upon which it resides. This increases the nucleophilicity of the thiol, allowing effective attack of the carbonyl. The resulting negative charge on the thioester carbonyl is stabilised by an oxyanion hole, constructed from two backbone amides, one of which is from the active-site cysteine. The other residue is located on a hairpin turn, and in type II FAS KSs this is a phenylalanine residue, often referred to as the ‘gatekeeper’ residue [63]. The tetrahedral

intermediate then collapses releasing the phosphopantetheine-ACP, leaving the growing polyketide chain attached to the KS.

*II - Decarboxylation*

The next step involves the decarboxylation of the chain extender unit (malonyl or methylmalonyl), which is supplied by the upstream ACP. Upon binding, a water molecule is activated by one of the catalytic histidine residues, facilitating the attack at the malonyl C3 carboxylate. This releases $HCO_3^-$, and the resulting enolate is stabilised by a conserved histidine.

*III - Condensation*

The final step involves the attack on the acyl-thioester by the enolate, with the tetrahedral intermediate stabilised by the oxyanion hole. The β-keto product is then formed following collapse of the tetrahedral intermediate. The elongated intermediate is now free to diffuse out of the active site, albeit thethered to the receiving ACP, and downstream for further modification.

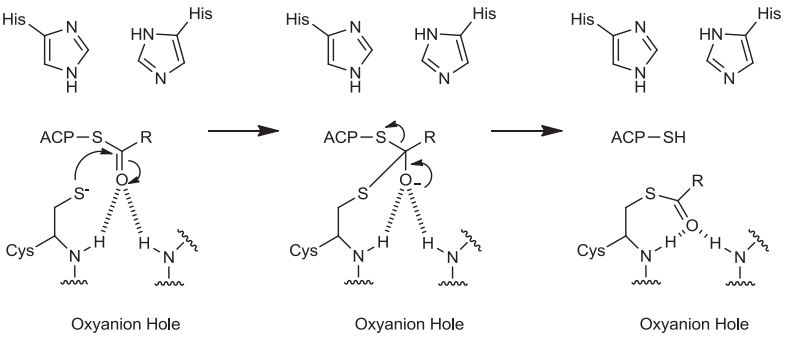
Experimental data suggests that KS domains from *cis*-AT PKSs have the ability to be selective for their substrates [65]. The EryKS2 domain was observed to be acylated most rapidly by the (2*S*,3*R*)-2-methyl-3-hydroxy-pentanoyl-SNAC mimic of the native intermediate, and a $K_m$ value of 5.3 ± 0.9 mM was obtained. Possibly a more important observation is that when the KS domain is acylated with an unnatural intermediate no condensation product is observed. This suggests that although that active site Cys can be acylated by aberrant intermediates, it is unable to catalyse the elongation reaction due to constraints within the active site. A possible explanation for this may be that acylation with the correct intermediate facilitates the organisation of residues required for an effective condensation. The converse would therefore be true for a KS acylated with a non-native intermediate [64]. In addition, the KS domain from the metazoan FAS has been shown to harbour preference towards fully reduced acyl chains—a trait which allows correctly saturated fatty acids to be generated. Analysis of the binding channel of FAS KS domains reveals a high proportion of hydrophobic residues in the acyl region of the binding pocket, promoting non-polar interactions [66].

#### 1.2.3.2 Loading and Non-elongating Ketosynthase Domains

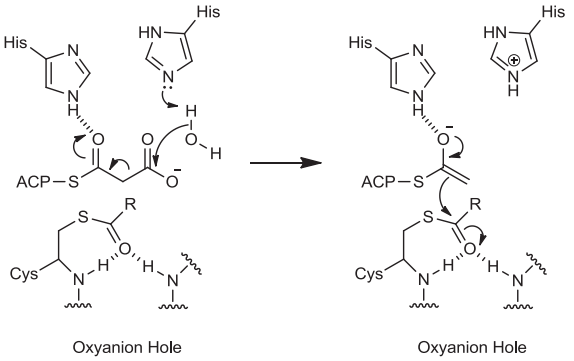
PKSs employ subtle variants of KS domains to fulfil specific functions within the biosynthetic cluster. Many of the loading modules from *cis*-AT PKSs utilise a KS domain with a glutamine residue in place of the active site cysteine, and are therefore termed $KS^Q$ domains [67]. It has been shown that alkylation of standard elongating KS domains with iodoacetamide structurally converts the cysteine residue into a moiety with similar chemical characteristics of a $KS^Q$ domain [68].

Another fascinating variant on the KS architecture is the non-elongating KS domain, termed $KS^0$. These $KS^0$ domains are a common feature of *trans*-AT PKSs, which lack a crucial active site histidine required for chain elongation. It

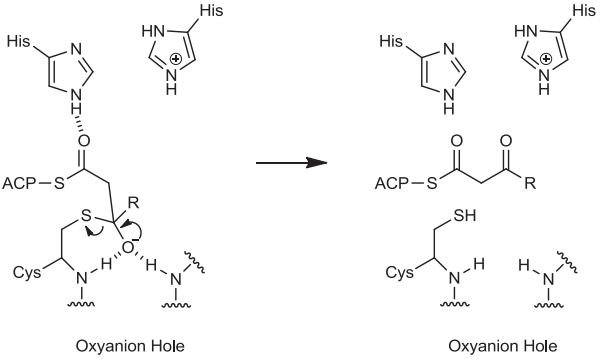
*I - Acylation*




*II - Decarboxylation*




*III - Condensation*




**Scheme 1.2** Proposed mechanism of action for ketosynthase domains

is postulated that they merely function to pass the intermediates onto the next module, or to allow modification from a downstream domain without any chain elongation occurring. However, the exact purpose of these domains is still poorly understood [15, 69].

### 1.2.3.3 The Acyltransferase Domain

Acyl-transferase (AT) domains have a highly conserved architecture from FASs to PKSs, consisting of a hydrolase core domain, and a smaller ferredoxin-like subdomain. Together, these domains form an active-site channel for substrate binding [57, 70]. Functionally, AT domains catalyse the transfer of acyl derivatives, most commonly malonyl or methylmalonyl, from CoA to the phophopantetheine chain of an ACP. The catalytic mechanism involves a Ser-His dyad, with the catalytic serine located in the well conserved GH**S**XG-motif [71].

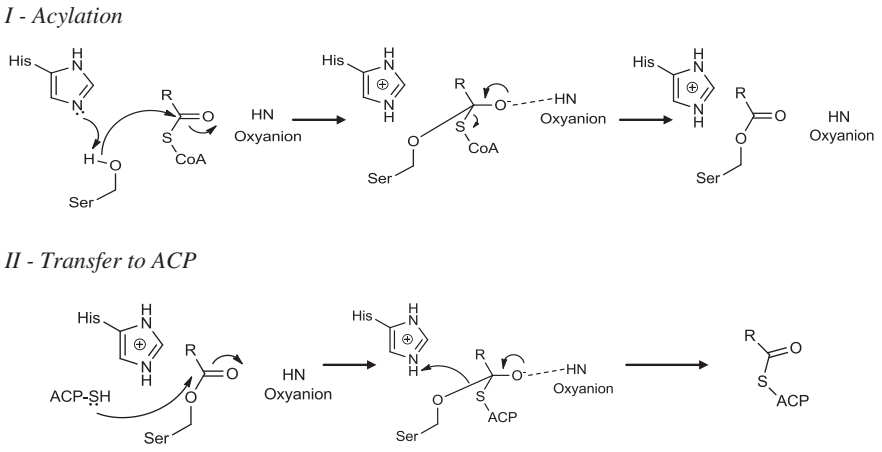
The mechanism of acylation and subsequent transfer to the ACP is thought to proceed as follows:

*I - Acylation*

The catalytic mechanism proceeds via nucleophilic attack, by the active site serine, at the thioester carbonyl of the acyl-CoA moiety. The resulting tetrahedral intermediate is stabilised by an oxanion hole generated by backbone amides, before collapse releases the CoA moiety allowing it to diffuse away from the active site.

*II - Transfer to ACP*

The acyl-loaded AT then allows docking of the phosphopantetheine chain of an ACP. The nucleophilic thiol of the phosphopantetheine chain attacks the ester




**Scheme 1.3** Proposed mechanism of an acyltransferase domain

carbonyl, resulting in another tetrahedral intermediate, again stabilised by the oxyanion hole. Re-protonation of the catalytic serine by the active site histidine allows the release of the acyl group onto the phosphopantetheine chain, which can then diffuse from the active site attached to the ACP (Scheme 1.3).

In type I modular PKSs, the specificity of the loading AT is often different to that of the extending ATs. The ATs associated with the loading module often exhibit a relaxed substrate tolerance, and can employ acetyl-, propionyl-, isopropionyl-, isobutyryl-, crotonyl- and phenylacetyl-CoA. However, the AT domains responsible for chain extension appear to have strict specificity, governed by the nature of specific motifs, which dictates the nature of the acyl unit loaded [72, 73]. The AT domains from *trans*-AT clusters exhibit substrate specificity akin to the extender unit ATs, therefore their substrate specificity can be predicted from the sequence alone [74]. In the case of PKSs with *trans*-acting AT domains, where the intrinsic position of the domain within the polypeptide chain is lost, it was proposed that a short sequence (~110aa) C-terminal to the KS domains serves as an AT-docking domain (ATd). Thought to be remnants of functional ATs, these well-conserved sequences are thought to allow the *trans*-acting AT to dock, and adopt their regular position within the module [25, 75].

#### 1.2.3.4 The Acyl Carrier Protein

Acyl carrier proteins (ACPs) are structurally simple proteins, consisting of a four-helix bundle that is stabilised by a hydrophobic core. An '*apo*-ACP' is post-translationally modified by a phosphopantetheinyl transferase enzyme (PPTase), which catalyses the transfer of a 4'-phosphopanteine moiety from co-enzyme A to a conserved serine residue located in the motif (D/E)xGxD**S**L, producing the mature '*holo*-ACP' [76, 77]. The ~18Å long phosphopantethiene moiety is an important addition to the ACP, as it allows the intermediates to be covalently attached to the PKS via a labile thioester linkage. In addition, it provides the flexibility required (2 nm) to reach active sites of further accessory domains within a given module [10, 56].

The conserved serine residue, which is post-translationally modified to incorporate a phosphopantetheine chain, is located at the N-terminal side of helix II. Helix II also acts as the 'recognition helix' due to its overall negative charge, and is believed to facilitate docking to positively charged regions on AT and KS domains [78]. An intriguing feature observed in type I PKS ACPs is the presence of a short N-terminal helix—termed 'helix 0'. The helix was first observed in the NMR structure of EryACP2 (PDB: 2JU2), and it has been postulated to have a role in ACP docking to downstream KS domains (Fig. 1.23a) [79]. Experiments carried out by the Khosla laboratory showed that ACPs transfer acyl groups preferentially to the KS domain immediately downstream. This suggests that interactions between helix II of ACPs and KS domains could be specific for given pairs of these domains [80].

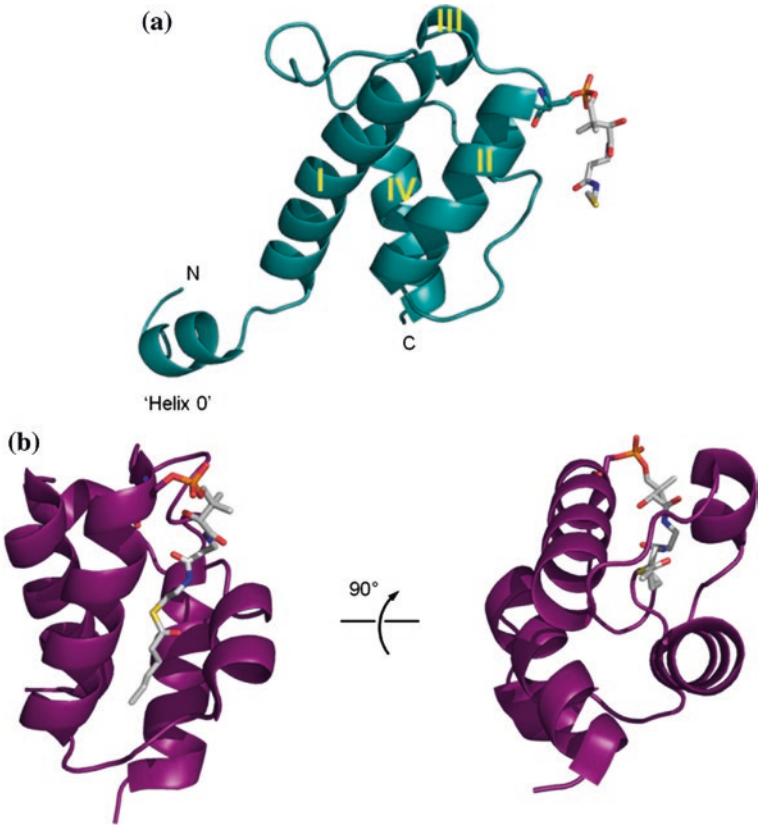



**Fig. 1.23** Structural analysis of the ACP domain with phosphopantetheine chain attached to the conserved Ser residue of the ACP. **a** Showing the four core helicies and the N-terminal 'helix 0'. **b** Orientations of acyl-PPant chain sequestered in the helical core of a type II FAS ACP. Rendered in PyMol from PDB: 2JU2 and 2FAC

Interestingly, crystal structures of ACPs from type II FAS indicate that the fatty acid acyl chain is sequestered between the helices of the ACP domain (PDB: 2FAC and 2FAE), highlighted in Fig. 1.23b [81]. However, this is not believed to be the case for ACPs from type I PKSs, highlighted by the NMR structure of the acyl-ACP from the curacin cluster (PDB: 2LIW) [82]. Although type I PKS intermediates may not bind in the helical cavity, interactions may indeed occur between the acyl chain and the sides of the helix bundle. This hypothesis has been somewhat supported by the observation that epimerisation of α-substituted, β-ketothioesters is prevented when tethered to the ACP [83].

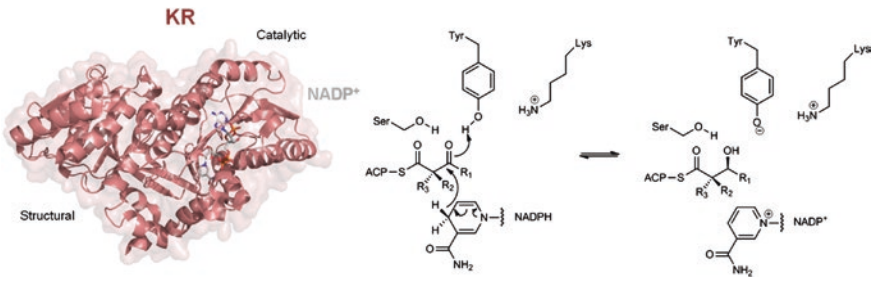



**Fig. 1.24** Crystal structure of the AmpKR2 domain. The structural and catalytic domains are labelled (*left*). The proposed catalytic mechanism of a KR domain is shown, with hydride attack and donation of a proton from a tyrosine residue generating the β-hydroxyl moiety. Rendered in PyMol from PDB: 3MJS

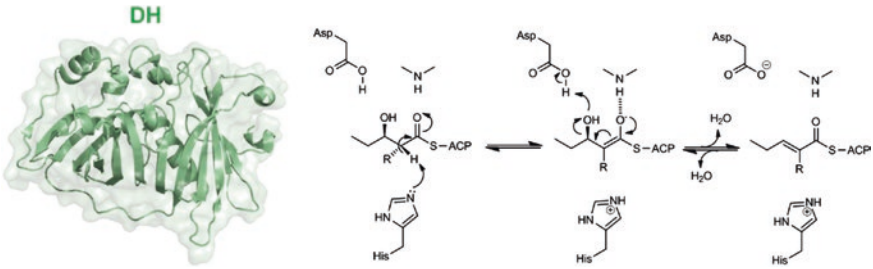



**Fig. 1.25** Crystal structure of the EryDH4. The proposed catalytic mechanism of a DH domain is shown, with abstraction of the C–H proton by a histidine and proton donation from an aspartic acid residue allowing the promoting the release of $H_2O$. Rendered in PyMol from PDB:3EL6

### 1.2.3.5 The Ketoreductase, Dehydratase and Enoylreductase Domains

The ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains are the most common 'optional' domains found within PKS modules. As previously described in Sect. 1.2.1, the presence of a KR domain within a module generates a β-hydroxyl intermediate, the addition of a DH will produce an α, β-unsaturated product, with the fully saturated intermediate obtained by addition of an ER domain.

The KR domain utilises NADPH to reduce the β-keto moiety of the growing polyketide chain, often in a stereospecific manner. The domain itself forms a Rossman-like fold, and conserved tyrosine, serine and lysine residues aid in catalysis [84]. Upon binding of the β-keto substrate, the pro-4*S* hydride of NADPH attacks the carbonyl, followed by abstraction of a proton from the catalytic tyrosine, resulting in a β-hydroxyl intermediate. KR domains are divided into three groups with regard to their catalytic activity. The A-type KRs generate an L-configured hydroxyl group, and the B-type KRs are responsible for producing

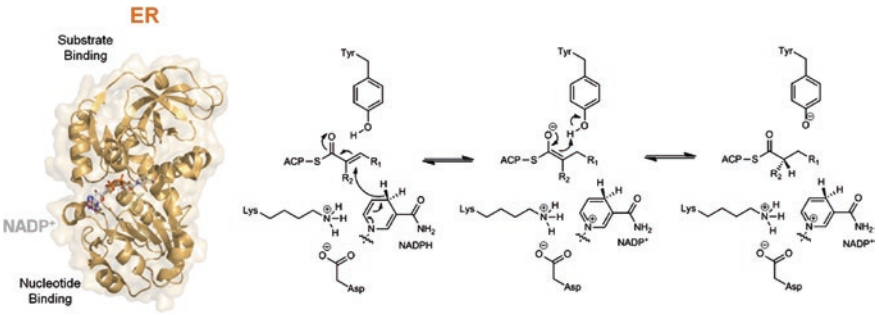



**Fig. 1.26** Crystal structure of the SpnER2. The substrate and nucleotide binding domains are labelled (*left*). The proposed catalytic mechanism of a ER domain, where a conserved tyrosine residue can provide a proton to allow full saturation of the intermediate. Rendered in PyMol from PDB: 3SLK

the D-configuration. The third type, C-type, are catalytically inactive [85, 86] (Fig. 1.24).

Dehydration of the β-hydroxy acyl chain is achieved by the DH domain, yielding an α, β-unsaturated moiety from an E1cB-elimination mechanism (Fig. 1.25). The proposed mechanism for dehydration proceeds via donation of a proton from an aspartic acid residue in the HPALL**D** motif, followed by removal of an α-proton by a catalytic histidine, generating the unsaturated product [87, 88]. The monomeric DH unit is ~280 residues in length, and forms a dimeric unit across the centre of type I PKSs (see Sect. 1.2.4).

The ER domain is ~310 residues in length, and catalyses the ultimate reduction of the biosynthetic intermediate. The current view regarding the reaction mechanism is that it proceeds by Michael addition. The pro-4*R* hydride of NADPH attacks the β-carbon forming an enolate intermediate within the binding pocket. The subsequent step is then dependent upon the residues available in the active site [89]. A recent crystal structure of spinosyn ER2 (SpnER2) revealed that a tyrosine in the active site could provide a proton to produce the fully saturated intermediate [90]. However, in cases where the tyrosine is not present, a conserved lysine residue could fulfil the same role (Fig. 1.26).

#### 1.2.3.6 The Thioesterase Domain

As briefly explained in Sect. 1.2.1, the ~270 residue TE domain catalyses either a hydrolysis or an intermolecular cyclisation reaction to terminate the biosynthesis of polyketides and fatty acids, releasing the final product from the assembly line. The TE domain exhibits an α/β-hydrolase fold and is dimeric in modular type I PKS systems. The mechanism of both hydrolytic cleavage and macrolactonisation commences with the transfer of the polyketide chain from the final ACP onto the active site serine of the TE, forming an acyl-TE intermediate. The hydrolytic

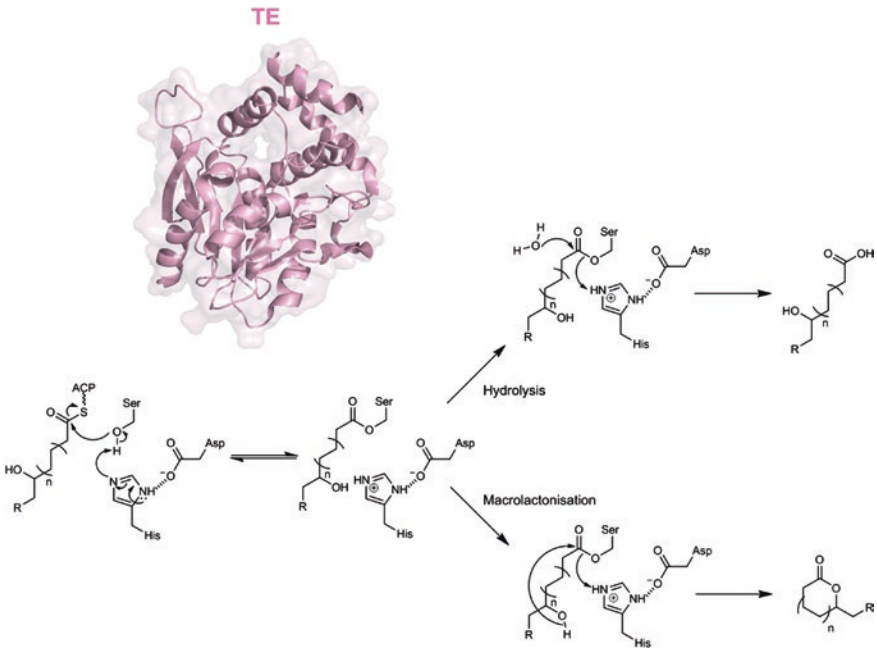



**Fig. 1.27** Crystal structure of the EryTE domain. The proposed hydrolytic and cyclisation mechanisms of a TE domain are shown. Rendered in PyMol from PDB: 1KEZ

cleavage scenario relies on the activation of a water molecule by a conserved histidine residue, promoting nucleophilic attack at the carbonyl of the thioester. In contrast, the macrolactonisation reaction usually proceeds when a secondary alcohol/amine has been installed early in the biosynthesis. This can be de-protonated by an active site histidine and, in the environment of the binding pocket, attack at the thioester promoted, releasing the cyclised product (Fig. 1.27) [91, 92].

### *1.2.4 Architechture of the 'Megasynthase'*

In recent years the structural understanding of type I PKSs has developed significantly. In 2002 electron microscopy revealed the quaternary structure of the dimeric metazoan FAS, appearing to form a sideways 'H' organisation (Fig. 1.28a) [93]. However, whether the homodimer formed in the FAS was organised in a head-to-head or head-to-tail fashion remained unknown until 2006 when a 4.5 Å crystal structure of the porcine FAS was solved (PDB: 2VZ9), which confirmed the head-to-head model for domain organisation of the FAS (Fig. 1.28b) [94]. Around the same time, crystal structures of a KS-AT didomain and a KR domain from the

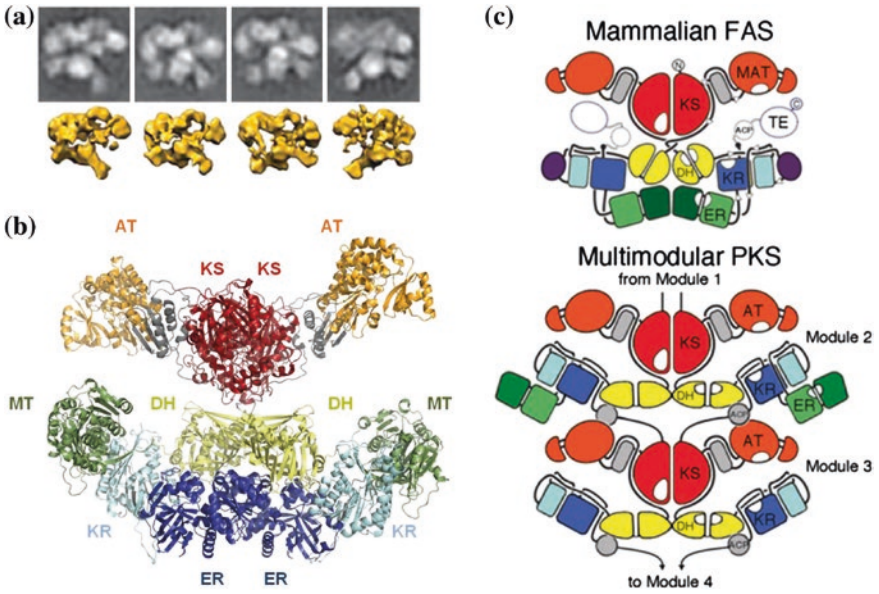



**Fig. 1.28** **a** Electron microscopy of metazoan FAS adopting 'H' conformation. **b** The 4.5 Å crystal structure of the porcine FAS. **c** Schematic representation comparing the models for FAS and modular PKSs

erythromycin PKS emerged, which provided insights into the sequence boundaries of domains within PKSs (PDB: 2HG4 and 2FR0 respectively) [57, 84].

The architecture of modular type I PKSs has only recently been deciphered, with crystal structures now available for KS, AT, KR, DH and ER domains. Although the type I PKSs share a large amount of structural similarity with the FAS, there are some clear differences between the two types of megasynthases. Notable differences are that the ER domain in type I PKSs is monomeric, compared to the dimeric ER which forms the base of the FAS. In addition the type I PKS DH dimer is linear, and the KR/ER interface is isolated, contrary to the FAS structure. These differences are highlighted in Fig. 1.28c [56].

### *1.2.5 In Vitro Techniques for Studying PKSs*

In order to fully understand the complex enzymatic processes carried out by PKS domains, in vitro studies are often carried out. Although in vivo studies are desirable, some of the organisms responsible for producing novel natural products are currently unculturable under laboratory conditions, placing the emphasis on in vitro studies to investigate such biosynthetic systems.

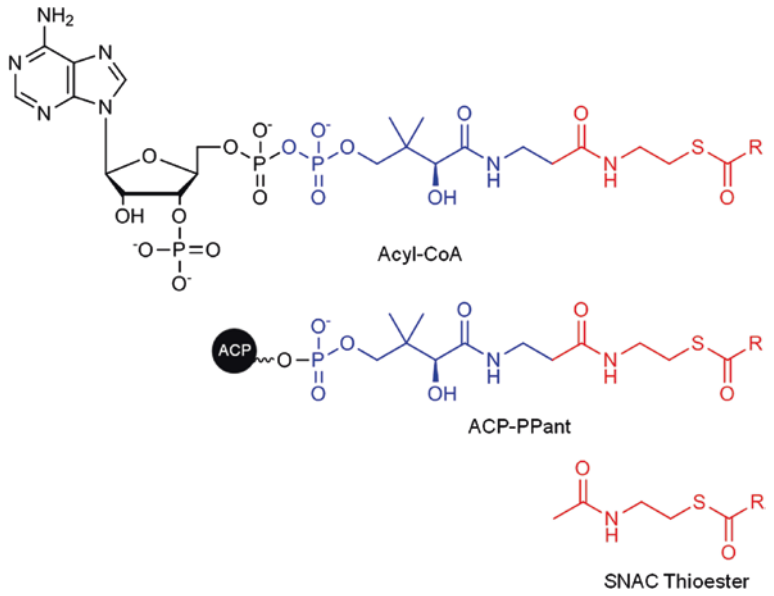



**Fig. 1.29** Structural comparison between acyl-CoA, ACP-phosphopantetheine and an *N*-acetylcysteamine (SNAC) thioester

In vitro experiments require production of recombinant proteins and an appropriate substrate probe. In the native system, acyl-CoAs are the substrates of AT domains, whereas ACP-tethered acyl-phosphopantetheine is used to deliver substrates to all other catalytic domains. A range of simple acyl-CoAs are commercially available, and can be used to probe the enzymology of PKS domains (Fig. 1.29). However, several studies have successfully demonstrated that by simply mimicking the terminal region of the phosphopantetheine chain using *N*-acetylcysteamine, enzymatic activity of PKS domains can be monitored [65, 95]. A simple coupling reaction between *N*-acetylcysteamine and the free acid of the desired acyl group can be used to generate substrate *N*-acetylcysteamine (SNAC) thioesters mimics of PKS intermediates.

The use of SNAC thioesters has many advantages. Firstly, the commercially available *N*-acetylcysteamine is far cheaper than CoA, making the synthesis of a range of substrate probes financially viable. In addition, the SNAC thioester offers reduced functionality, removing the potential side reactions during synthesis. Finally, the SNAC thioesters are far more stable than acyl-CoAs as labile phosphoester bonds are no longer present.

An improvement upon the SNAC thioester is likely to be an acyl-ACP, which are also used as substrate probes for PKS domains. The fundamental drawback to acyl-ACPs is that their synthesis is expensive and time consuming. The current methodology requires enzymatic transfer of the acyl-PPant group of an acyl-CoA to an *apo*-ACP using a phosphopantetheinyl transferase (PPTase) enzyme [96].

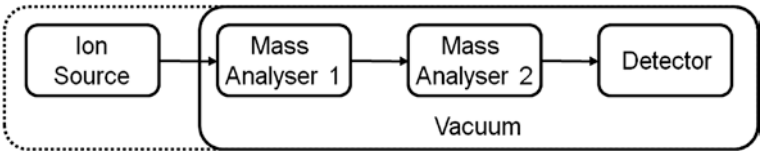



**Fig. 1.30** Schematic of a typical mass spectrometer with two mass analysers separating the ion source from the detector. The *dashed line* represents an optional vacuum region dependent upon the ionisation method

This approach suffers from several disadvantages. Firstly, it requires production of *apo*-ACP from a strain deficient in the 4'-phosphopantetheinyl transferase *sfp* gene. The subsequent synthesis of acyl-ACPs is often slow and requires either production or purchase of the PPTase enzyme. Furthermore, attachment of more elaborate acyl groups requires access to synthetically challenging and expensive acyl-CoAs.

## 1.3 Mass Spectrometry

Mass spectrometry (MS) is a powerful analytical technique used across many scientific disciplines. In practice, a mass spectrometer is used to measure the molecular mass of an analyte molecule via generation of gas-phase ions and subsequent separation by their mass-to-charge ratio (*m/z*). The first mass spectrometer was devised by J.J. Thomson in 1912, which he used to measure mass spectra for a range of gases. A notable improvement upon this initial work was conducted by Francis Aston in 1922 allowing measurement of isotopes from non-radioactive elements [97, 98]. Since this early pioneering work, the field of MS has developed as a vital analytical technique for analysis of small molecules up to large protein complexes.

Although many variations of MS instruments are available, all consist of three core components for function; the ion source, mass analyser(s) and the detector (Fig. 1.30). The mass analysers must be held under vacuum, however depending upon the nature of the ionisation process, the ion source may be under vacuum or at atmospheric pressure [99].

### *1.3.1 Ionisation Methods*

Many ionisation techniques have been developed capable of producing gas-phase ions from samples in solid, liquid and gas phases. Examples include electron ionisation (EI), chemical ionisation (CI), matrix-assisted laser desorbtion ionisation

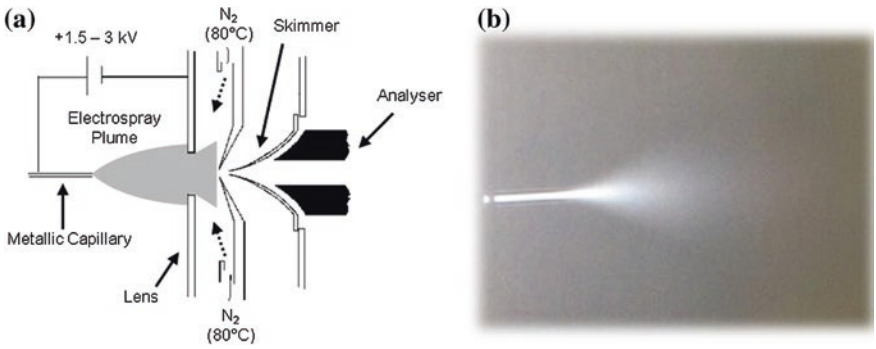



**Fig. 1.31** **a** Plan of a typical electrospray ionisation source operating in positive ion mode. Adapted from [99]. **b** Electrospray ionisation plume photographed from a Thermo LTQ Ion Trap source

(MALDI) and electrospray ionisation (ESI). However, for the purposes of this thesis only ESI will be described in detail due to its extensive use within this work.

#### 1.3.1.1 Electrospray Ionisation

The technique of electrospray ionisation (ESI) has proved a highly valuable tool for studying proteins and their complexes in recent years. The fusion of mass spectrometry with the development of gentle ionisation methods such as ESI, have allowed larger, more delicate protein complexes to be analysed. ESI has the capability to ionise and desolvate biomolecules from solution, into the gas phase without input of excessive internal energy to the molecular ions. This soft ionisation technique can allow the preservation of covalent and non-covalent modifications to proteins [100, 101].

An electrospray plume is generated upon application of a strong electric field to a flow (1–20 μL min) of typically volatile liquid, through a conductive capillary. The tip of the capillary is held at a high potential difference (1.5–3 kV), producing an electric field in the region of $10^6$ V $m^{-1}$. The application of a strong electric field causes the liquid to deform into a Taylor cone due to the build-up of charge at the capillary tip, and results in ejection of a fine jet of charged liquid (Fig. 1.31b) [102]. The stability of the droplets in the electrospray plume is dictated by two opposing forces; the surface tension of the droplet which increases cohesion of the droplet holding it together, and coulombic repulsion which acts to break up the droplet due to the repulsion of like charges on the droplet surface [101, 103].

The electric field pushes the droplets through an evaporation chamber, where a flow of heated inert gas (often $N_2$), may be used to speed up the evaporation process. As the solvent evaporates, the charge density increases at droplet surface as the droplet size is reduced. Coulomb repulsion between the surface charges increases until the repulsive force exceeds the surface tension maintaining the

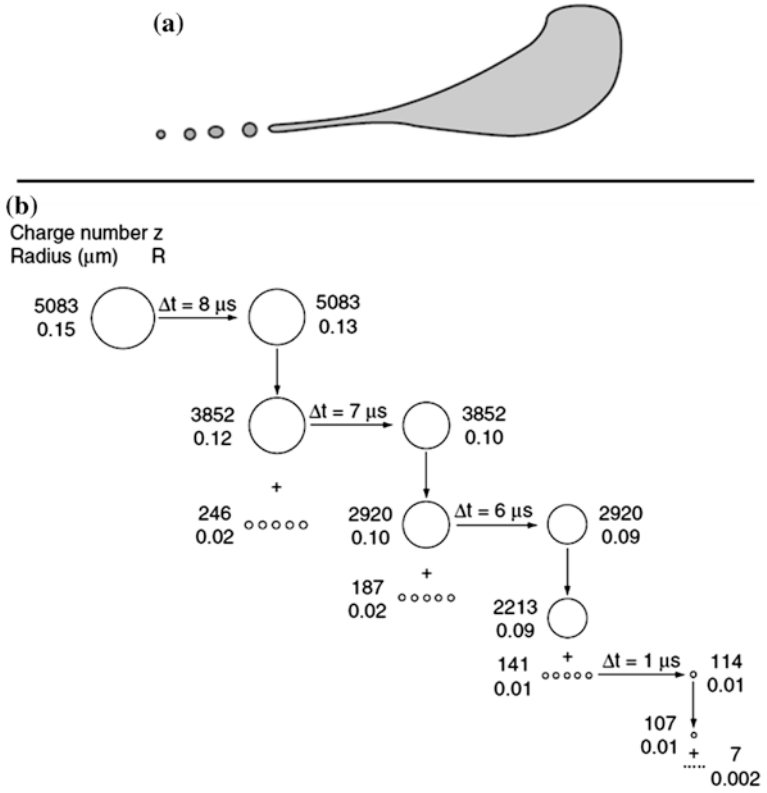



**Fig. 1.32** **a** Mechanical decomposition of a droplet from an electrospray source. Taken from [99]. **b** Scheme of droplet evolution during the electrospray process as a result of solvent evaporation and fission. Adapted from [103]

droplet. This point is defined at the Rayleigh limit, where coulombic fission drives the dispersion of the jet into a fine plume of charged droplets [104]. The Rayleigh equation (Eq. 1.1) defines the relationship between the droplet charge ($Q_{Ry}$), the droplet radius (R), the surface rension ($\gamma$) and the electrical permittivity of the environment ($\varepsilon_0$).

$$Q_{Ry} = 8\pi(\varepsilon_0 \gamma R^3)^{1/2} \tag{1.1}$$

The premise of coulombic fisson taking place upon reaching the Rayleigh limit was widely believed to be the driving force behind the production of a fine plume of electrically charged droplets. However, work published in 1994 by Gomez and Tang, produced images of heptane droplets dividing from an ESI source [105]. Figure 1.32a depicts one of their images, showing that the droplet can be mechanically deformed, which effectively decreases the coulombic repulsion required for

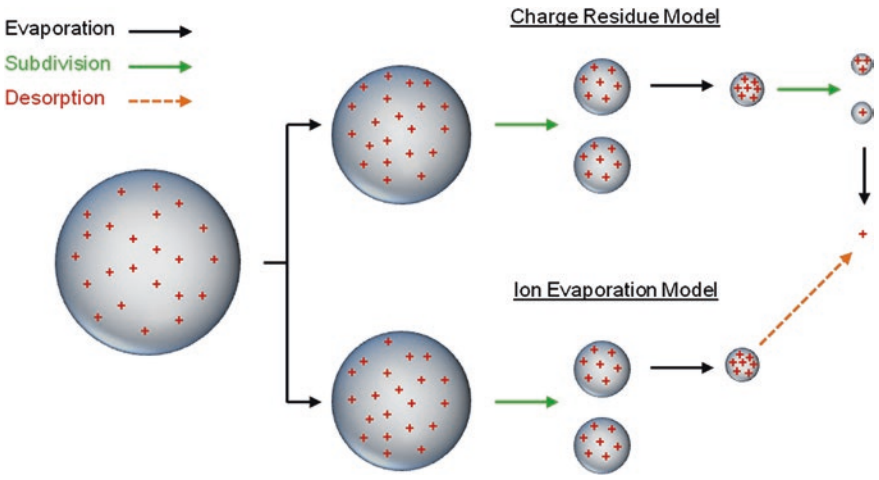



**Fig. 1.33** Summary of CRM and IEM for formation of gas phase ions from a highly charged droplet. Adapted from [103]

coulombic fission. Ejection of smaller droplets from the tail of the parent droplet occurs as a result, with the daughter droplets carrying approximately 2 % of the mass and 15 % of the charge (Fig. 1.32b). The parent droplet will continue to shrink due to evaporation of solvent producing further generations of daughter droplets. The mechanism for the formation of gas-phase ions from charged droplets is yet to be elucidated, and is the subject of much debate. Currently, two prevailing theories have been proposed; the charged-residue model (CRM) and the ion evaporation model (IEM) [106].

The Charged Residue Model (CRM), presented by Dole et al. in 1968, proposes that coulombic fission proceeds until no further evaporation can occur, where each droplet contains a single analyte molecule, with the charges on the droplet transferred directly to the analyte [107]. In contrast, Iribarne and Thompson put forward the Ion Evaporation Model (IEM) in 1976, where the authors propose that the charged analyte ions are ejected from the droplet surface upon reaching a given radii, defined as $R_E$, before reaching the Rayleigh instability limit, defined as $R_R$ [108]. Figure 1.33 depicts simplistic mechanisms for the two models. Although both mechanisms appear plausible, evidence suggests that the IEM mechanism is favoured when the analyte is relatively small, whereas larger analytes such as proteins proceed via the CRM model.

### 1.3.1.2 Nano-Electrospray Ionisation

Nano-Electrospray Ionisation (nESI) is a variant on ESI which allows much lower flow rates, and therefore allows reduced sample consumption. The principle was developed in the mid-90s, with the primary aim of coupling to HPLC systems;

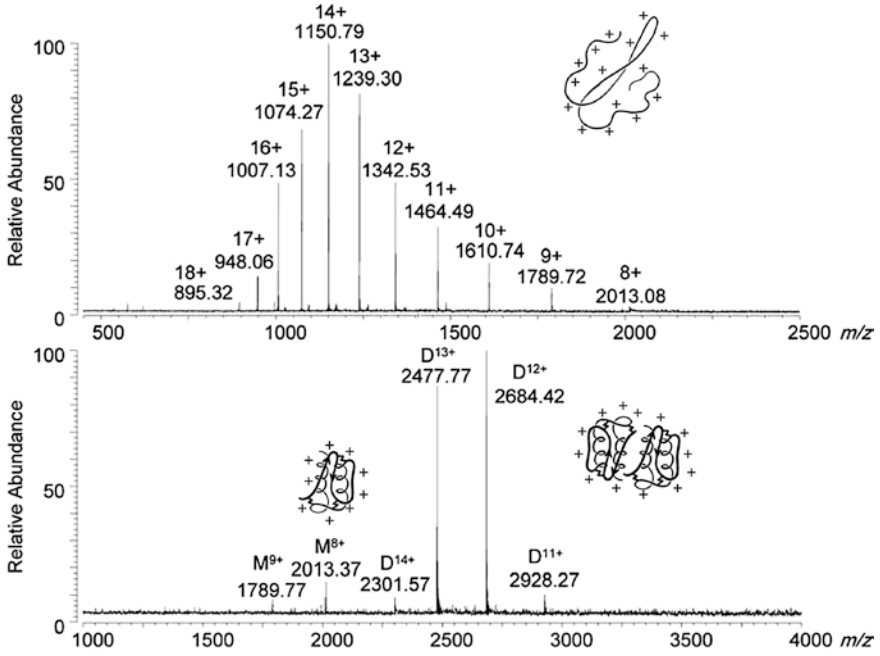



**Fig. 1.34** Positive electrospray mass spectra of *E.coli* ecotin protein sprayed from denaturing conditions, $H_2O$/MeCN, 1 % acetic acid (*top*) and non-denaturing conditions, 10 mM $NH_4OAc$ (*bottom*). Denaturing conditions yield a wide charge state envelope and measurement of the 16 kDa monomeric unit. Non-denaturing conditions allow preservation of protein structure, thus a 32 kDa dimeric (D) is observed, with a small amount of monomer observed (M). Tight charge state envelopes indicate a folded state as only surface residues are protonated. Adapted from [103]

and ~500 nL/min flow rates were reported which were able to sustain an electrospray plume [109]. Although commonly used in conjunction with HPLC analysis, nESI is also used in its static form for the analysis of large proteins and complexes to reduce the amount of sample required for MS analysis [110, 111]. The nESI emitters are constructed from glass capillaries pulled to a fine tip, and coated in a thin layer of conductive metal (often gold). Static nESI emitters are usually mounted on an (*x*, *y*, *z*) adjustable platform allowing suitable positioning to the MS inlet.

### 1.3.1.3 Analysing Proteins by ESI-MS

Protein ions observed in positive ESI-MS spectra are quasimolecular due to the addition of multiple protons to the surface of the protein. This is often represented as $[M + nH]^{n+}$ when assigning spectral peaks. The number of charges on a given protein is dependent upon the number of accessible ionisable residues. In the case

of positive ionisation, basic residues are the targets for protonation. One of the most noteable features of an ESI-MS spectrum is the appearance of a clustered series of peaks corresponding to a series of multiply charged ions [99]. However, the charge state distribution is often dependent upon the folded state of the protein.

Figure 1.34 depicts the ESI-MS spectra of the dimeric *E.coli* ecotin protein sprayed from $H_2O$/MeCN, 1 % acetic acid (top) and 10 mM $NH_4OAc$, pH 7.5 (bottom). In this example, the denaturing solution has promoted unfolding of the protein and protonation of all exposed basic residues, promoted by the presence of acid. This results in a wide charge state envelope ranging from 8+ to 18+, with the monomeric molecular weight of 16 kDa measured. In contrast, the non-denaturing buffer, 10 mM $NH_4OAc$, pH 7.5, allows the native fold and quaternary structure of the protein to be maintained. In this case, the dimeric (D) ecotin is predominantly observed with a molecular mass of 32 kDa. Despite the increased mass, the average charge state is 12+, indicating that only solvent exposed residues on the surface of the protein are protonated [103].

Generally, non-denaturing conditions are employed when protein subunit stoichiometry or non-covalent ligand binding are being measured, as the native protein fold is often required for such functionality. Whereas denaturing conditions offer increased signal intensity and reduced peak adduction, ideal for an accurate mass measurement.

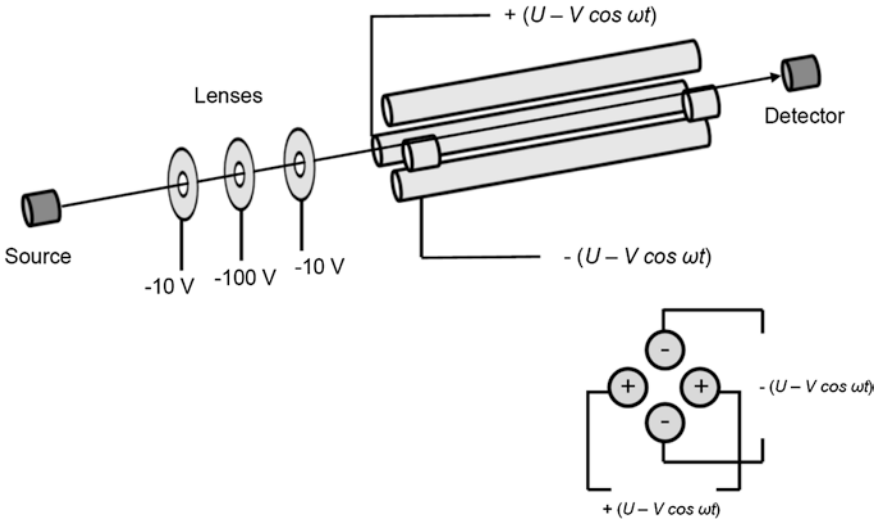



**Fig. 1.35** Schematic representation of a quadrupole mass analyser and pre-focussing lenses. A circuit diagram of the four poles is also shown. Reproduced from [99]

### *1.3.2 Mass Analysers*

Post ionisation and desolvation, analyte ions are separated according to their *m/z* ratio by a mass analyser. The analyte ions are usually separated using static or dynamic electric/magnetic fields. An exception to this would be time-of-flight which does not utilise a magnetic/electric field, except if a reflection is used, which employs a static electric field. Although many mass analysers and their variants have been developed, only the principles of a quadrupole and time-of-flight mass analysers will be described in this section in accordance with the techniques used in later chapters.

#### 1.3.2.1 Quadrupole Analysers

The fundamental principles of a quadrupole mass analyser were described in 1953 by Paul and Steinwegen [112]. The device itself consists of four rods in a parallel arrangement, either of circular or hyperbolic cross-sections, and utilises the principle of stable ion trajectories within an oscillating electric field to separate *m/z* ratios. For example a positive ion entering a quadrupole will be attracted to a negative rod, however if the polarity of the rod switches to positive before the ion has chance to discharge on it, then the ion will change its path direction.

Figure 1.35 shows the principle and a schematic of a quadrupole. As an ion travels forward on the *z*-axis, it is subjected to an oscillating electric field, which continually forces the ion towards the central cavity of the analyser. The electric field is generated by DC and RF voltages applied to the four electrodes along the *x*- and *y*-axes (producing a 2D field), allowing the ions to travel down through the field-free *z*-axis [99].

The 2D quadrupole potential, $\Phi_0$, is defined by the following equations:

$$\Phi_0 = +(U - V\cos\omega t), \tag{1.2}$$

$$-\Phi_0 = -(U - V\cos\omega t) \tag{1.3}$$

$$\Phi_0(x, y) = \frac{(x^2 - y^2)}{r_0^2} \tag{1.4}$$

where $\omega$ is the angular frequency, $U$ is the DC voltage applied (500 to 2000 V), $V$ is the RF voltage applied (0 to ± 3000 V) and $r_0$ is the quadrupole internal radius.

Although the ions travelling through the quadrupole maintain a constant velocity along the *z*-axis, they are subjected to small accelerations and decelerations via the *x*- and *y*-axis, as a result of the electric field applied. The motion of the ions in the *x* and *y* directions whilst travelling through the quadrupole are determined using the following equations:

$$-A\frac{\partial\Phi_0}{\partial x} = \frac{-2x}{r_0^2}(U - V\cos\omega t) \tag{1.5}$$

$$-A\frac{\partial \Phi_0}{\partial y} = \frac{2y}{r_0^2}(U - V\cos\omega t) \tag{1.6}$$

where $A$ is the amplitude of the quadrupole field.

The motion of an ion in a quadrupole field along the $x$ and $y$ axis can be described by second order Mathieu equations:

$$\frac{d^2x}{d\xi^2} + (a_x - 2q_x\cos 2\xi)x = 0 \tag{1.7}$$

$$\frac{d^2y}{d\xi^2} + \left(a_y - 2q_y\cos 2\xi\right)y = 0 \tag{1.8}$$

where the variable $a_x$, $a_y$, $q_x$ and $q_y$ are defined as:

$$a_u = a_x = -a_y = \frac{8zeU}{mr_0^2\omega^2} \tag{1.9}$$

$$q_u = q_x = -q_y = \frac{4zeV}{mr_0^2\omega^2} \tag{1.10}$$

$$\xi = \frac{\omega t}{2} \tag{1.11}$$

In order to obtain a stable ion trajectory along the $x$ and $y$ axis, $r_0 \neq 0$ for both equations. When $r_0 = 0$, the ion will hit the electrode and discharge. Therefore

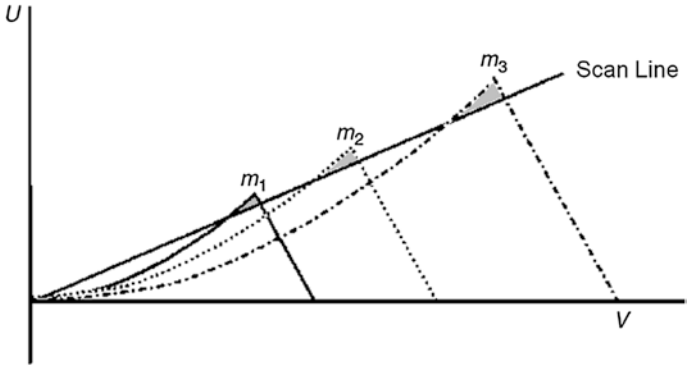



**Fig. 1.36** Plot of ion stability as a function of $U$ and $V$ for ions with increasing masses, $m_1 > m_2 > m_3$. The scan *line* is shown intersecting the stability areas for each ion by changing $U$ linearly as a function of $V$. Modified from [94]

solutions to the Mathieu equations can be denoted as bounded ($r_0 \neq 0$) or unbounded ($r_0 = 0$).

Considering the aforementioned variables of the Mathieu equation, it is possible to deduce that:

$$U = a_u \frac{m}{z} \frac{\omega^2 r_0^2}{8e} \tag{1.12}$$

$$V = q_u \frac{m}{z} \frac{\omega^2 r_0^2}{4e} \tag{1.13}$$

where $u$ can be $x$ or $y$.

Equations 1.12 and 1.13 govern the behaviour of $m/z$ values with respect to $U$ and $V$. Figure 1.36 show this relationship graphically.

The scan line, shown above, is the result of the linear change of $U$ as a function of $V$. A stable ion trajectory for ions $m_1$, $m_2$ and $m_3$ through the quadrupole is achieved when the scan line passes through their respective stability regions (shaded grey). A scan line with a steeper gradient will result in higher resolution, but will suffer from decreased sensitivity. Removal of the DC voltage ($U = 0$) excludes the ability to select ions of a particular $m/z$, allowing transmission of all ions through the quadrupole.

#### 1.3.2.2 Time-of-Flight Analysers

A common mass analyser used for large biological molecules is the time-of-flight (TOF) analyser, due to its theoretically unlimited $m/z$ range. First proposed by W. E. Stephens in 1946, the principle of TOF is elegantly simple; separation of ions by their velocity post acceleration [113]. Ions are accelerated through a field-free flight tube, with nominally the same amount of kinetic energy applied to each ion. It therefore follows that different masses will have different velocities, with smaller ions reaching the detector earlier than larger ions.

In the first instance, ions are accelerated by a source region, which pulses packets of ions into the flight tube using a high electric field. At this point all the ions receive the same amount of kinetic energy, $E_k$, given by Eq. 1.14.

$$E_k = \frac{mv^2}{2} = zeV_{acl} \tag{1.14}$$

where $m$ is the mass of the ion, $v$ is the velocity of the ion, $z$ is the charge on the ion, $e$ is the elementary unit of charge ($1.602 \times 10^{-19}$ C) and $V_{acl}$ is the accelerating voltage applied. Therefore, it follows that the velocity of an ion, $v$, travelling in the flight tube can be calculated using the following equation:

$$v = \sqrt{\left(\frac{2zeV_{acl}}{m}\right)} \tag{1.15}$$

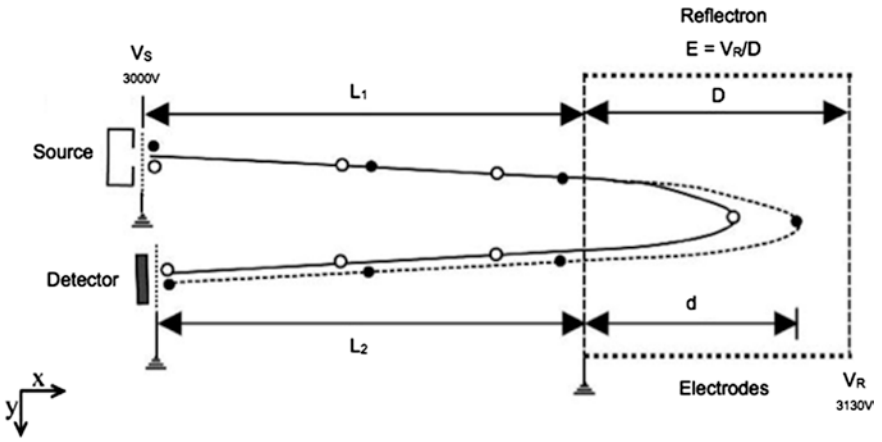



**Fig. 1.37** Schematic representation of a TOF analyser with a reflectron. ● = ions with $m$ mass and with $q$ $E_k$. ○ = ions with $m$ mass and $<q$ $E_k$. The ○ ions reach the reflectron later than ●, but will not penetrate as far into the electric field, the difference in $E_k$ is therefore corrected by the longer flight path of ●. Reproduced from [99]

Using Eq. 1.15, it becomes clear that the higher the mass of a given ion, the slower the velocity. Therefore a high mass ion will hit the detector later than a low mass ion. Equation 1.16 defines the time, $t$, taken for a given ion to travel the length, $L$, of a flight tube and reach the detector.

$$t = \frac{L}{v} \tag{1.16}$$

Substitution of $v$ from Eq. 1.15 and subsequent rearrangement yields:

$$\frac{m}{z} = \left(\frac{2eV_{acl}}{L^2}\right)t^2 \tag{1.17}$$

Using Eq. 1.17, the $m/z$ value of any given ion subjected to TOF can be calculated.

A limitation of TOF, in its linear form, is poor resolution; a result of ions with the same $m/z$ values arriving over a broad range of times. A notable improvement to TOF resolution has come from the use of a reflectron. A reflectron utilises a static electric field to reverse the direction of the ion beam. The improved resolution is obtained by focusing ions with differing kinetic energies, but the same $m/z$. Consider two ions with the same $m/z$, but with different kinetic energies. The ion travelling faster will pass deeper into the electric field, thus delaying the reverse of direction allowing the slower ion to effectively catch up (Fig. 1.37) [99, 103].

The motion of ions in a reflectron-assisted TOF analyser can be described as follows. Ions of a given $m/z$ ratio and kinetic energy are pulsed towards the reflectron with a constant velocity, $v_i$. The ions penetrate the reflectron electric field until a depth, $d$, is reached when the ion velocity is zero, $v_0$. At this point, the

electric field acts to reverse the direction of the ions, expelling them into the field-free region of the flight tube with the same velocity as they entered, $v_i$. Due to the deceleration, the average velocity of the ion in the reflectron is $v_i/2$, and the total flight distance to the detector given by $2d$. It can therefore be inferred that the total time an ion spends in the reflectron, $t_R$, electric field is given by:

$$t_R = \frac{4d}{v_i} \tag{1.18}$$

Using Fig. 1.37 as a guide, $L_1$ and $L_2$ can be defined as the distances between the source and reflectron, and vice versa. Therefore, the amount of time an ion spends in the field-free flight tube, $t_F$, is given by:

$$t_F = \frac{L_1 + L_2}{v_i} \tag{1.19}$$

The total time an ion spends in the reflectron-assisted TOF analyser, $t_T$, is the sum of Eqs. 1.18 and 1.19:

$$t_T = \frac{(L_1 + L_2 + 4d)}{v_i} \tag{1.20}$$

Substitution of $(L_1 + L_2 + 4d)$ for $L$, and subsequent rearrangement of the linear TOF Eq. 1.17 yields:

$$\frac{m}{z} = \left(\frac{2eV_{acl}}{(L_1 + L_2 + 4d)^2}\right) t_T^2 \tag{1.21}$$

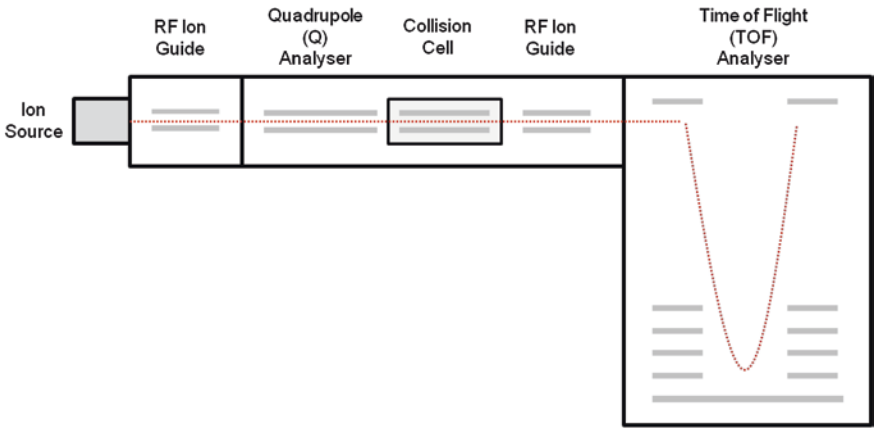



**Fig. 1.38** Example of a hybrid mass analyser instrument. A schematic of standard Q-TOF instrument, which is equipped with quadrupole (Q) and Time of Flight (TOF) mass analysers, thus classified as a Q-TOF. Separating the two mass analysers is a collision cell for tandem MS

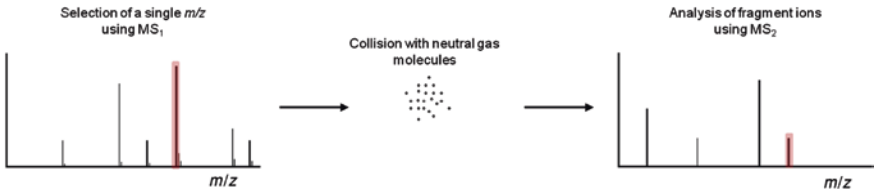



**Fig. 1.39** Diagrammatic representation of a tandem MS experiment by CID. The selected precursor ion is highlighted in *red*. Adapted from [99]

From which *m/z* ratios can be calculated from reflectron-assisted TOF analysers. Given the significant improvement in resolution that the reflectron offers, most commercial TOF-mass spectrometers are fitted with a reflectron.

### 1.3.2.3 Hybrid Mass Analyser Instruments

A multitude of hybrid mass spectrometers have been developed which utilise several mass analysis techniques within a single instrument. These instruments allow better characterisation of analytes often due to the ability to perform tandem mass spectrometry experiments (see Sect. 1.3.2.4). Arguably the most common hybrid instrument is the quadrupole time-of-flight (Q-TOF), which is comprised of a quadrupole filter situated in front of a collision cell and followed by a TOF analyser (Fig. 1.38). This particular instrument set-up allows the quadrupole to function in two modes; the RF only mode which allows all ions to pass through the quadrupole and onto the TOF analyser, or, by applying the DC voltage the quadrupole can resolve single ions of a given *m/z*, which can be further activated in the collision cell before TOF analysis.

### 1.3.2.4 Tandem Mass Spectrometry

Tandem mass spectrometry, commonly abbreviated to MS/MS or $MS^2$, is a method which usually applies a dissociation process to analyse ions of selected *m/z* values. Because of this, a minimum of two mass analysers are required for this technique. The most common application of MS/MS is isolation of a parent ion, $m_p$, of interest with the first analyser, which then either spontaneously dissociates or is activated to yield fragment ions, $m_f$, and neutral fragments, $m_n$, detected by the second analyser.

$$m_p^+ \xrightarrow{Activation} m_f^+ + m_n \tag{1.22}$$

Tandem mass spectrometry analysis can be carried out in space or time. The former requires two or more mass analysers coupled together, a quadrupole time-of-flight (QTof) or a triple quad (QqQ), for example. Equally, tandem mass

spectrometry can be performed over time using ion trap devices such as a quadrupole-ion trap (QIT) or an orbitrap.

### Collision Induced Dissociation

The principle of tandem MS requires selection and fragmentation of a given ion by the first mass analyser ($MS_1$), and subsequent detection by the second ($MS_2$) (Fig. 1.39). A commonly employed method of fragmentation is collision induced dissociation (CID), where analyte ions are accelerated into a chamber of inert gas molecules (e.g. $N_2$ or Ar). Collisions with the gas molecules converts collisonal kinetic energy into internal energy, promoting fragmentation of the ion assuming the energy is sufficient. The amount of input energy altered by dictating the amount of acceleration applied to the ions upon entering the CID chamber. The energy transferred to an ion is distributed equally prior to fragmentation, thus using this method; often the weakest bonds are preferentially cleaved. In order to detect fragmented ions post CID, the collision cell must be positioned between $MS_1$ and $MS_2$. A common arrangement is shown in Fig. 1.38, where the CID cell (labelled 'trap') is situated between a quadrupole ($MS_1$) and a TOF ($MS_2$) analyser [99].

## References


1. B.B. Mishra, V.K. Tiwari, Natural products: an evolving role in future drug discovery. Eur. J. Med. Chem. **46**, 4769–4807 (2011)
2. G.M. Cragg, D.J. Newman, Biodiversity: a continuing source of novel drug leads. Pure Appl. Chem. **77**, 7–24 (2005)
3. B. Spellberg, J.H. Powers, E.P. Brass, L.G. Miller, J.E. Edwards, Trends in antimicrobial drug development: implications for the future. Clin. Infect. Dis. **38**, 1279–1286 (2004)
4. M. Leeb, Antibiotics: a shot in the arm. Nature **431**, 892–893 (2004)
5. S.G. Van Lanen, B. Shen, Microbial genomics for the improvement of natural product discovery. Curr. Opin. Microbiol. **9**, 252–260 (2006)
6. D.J. Newman, G.M. Cragg, Natural products as sources of new drugs over the last 25 years. J. Nat. Prod. **70**, 461–477 (2007)
7. D.J. Newman, G.M. Cragg, K.M. Snader, The influence of natural products upon drug discovery. Nat. Prod. Rep. **17**, 215–234 (2000)
8. C. Hertweck, Hidden biosynthetic treasures brought to light. Nat. Chem. Biol. **5**, 450–452 (2009)
9. C. Hertweck, The Biosynthetic Logic of Polyketide Diversity. Angew. Chem. Int. Ed. **48**, 4688–4716 (2009)
10. S. Smith, S.C. Tsai, The type I fatty acid and polyketide synthases: a tale of two megasynthases. Nat. Prod. Rep. **24**, 1041–1072 (2007)
11. J.N. Copp, B.A. Neilan, The phosphopantetheinyl transferase superfamily: phylogenetic analysis and functional implications in cyanobacteria. Appl. Environ. Microbiol. **72**, 2298–2305 (2006)
12. C. Khosla, Structures and mechanisms of polyketide synthases. J. Org. Chem. **74**, 6416–6420 (2009)

13. M.A. Fischbach, C.T. Walsh, Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. **106**, 3468–3496 (2006)
14. T. Nguyen et al., Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nat. Biotechnol. **26**, 225–233 (2008)
15. J. Piel, Biosynthesis of polyketides by trans-AT polyketide synthases. Nat. Prod. Rep. **27**, 996–1047 (2010)
16. L. Du, L. Lou, PKS and NRPS release mechanisms. Nat. Prod. Rep. **27**, 255–278 (2010)
17. J. Staunton, K.J. Weissman, Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. **18**, 380–416 (2001)
18. B. Shen, Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr. Opin. Chem. Biol. **7**, 285–295 (2003)
19. Y. Abe et al., Molecular cloning and characterization of an ML-236B (compactin) biosynthetic gene cluster in Penicillium citrinum. Mol. Genet. Genomics **267**, 636–646 (2002)
20. K.J. Weissman, Introduction to polyketide biosynthesis. Methods Enzymol.: Complex Enzym. Microb. Nat. Prod. Biosynth. Part B: Polyketides Aminocoumarins Carbohydr. **459**, 3–16 (2009)
21. L. Katz, The DEBS paradigm for type I modular polyketide synthases and beyond. Methods Enzymol.: Complex Enzym. Microb. Nat. Prod. Biosynth. Part B: Polyketides Aminocoumarins Carbohydr. **459**, 113–142 (2009)
22. C. Khosla, Y. Tang, A.Y. Chen, N.A. Schnarr, D.E. Cane, Structure and mechanism of the 6-deoxyerythronolide B synthase. Ann. Rev. Biochem. **76**, 195–221 (2007)
23. D.A. Hopwood, Genetic contributions to understanding polyketide synthases. Chem. Rev. **97**, 2465–2497 (1997)
24. G. Yadav, R.S. Gokhale, D. Mohanty, Towards prediction of metabolic products of polyketide synthases: an In Silico analysis. PLoS Comput. Biol. **5**, e1000351 (2009)
25. Y.Q. Cheng, G.L. Tang, B. Shen, Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. Proc. Natl. Acad. Sci. USA **100**, 3149–3154 (2003)
26. J. Piel, A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. Proc. Natl. Acad. Sci. USA **99**, 14002–14007 (2002)
27. T. Gulder, M. Freeman, J. Piel, The catalytic diversity of multimodular polyketide synthases: natural product biosynthesis beyong textbook assembly rules. Top. Curr. Chem. 1–53 (2011)
28. J. Piel, D.Q. Hui, N. Fusetani, S. Matsunaga, Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. Environ. Microbiol. **6**, 921–927 (2004)
29. T. Hochmuth, J. Piel, Polyketide synthases of bacterial symbionts in sponges—evolution-based applications in natural products research. Phytochemistry **70**, 1841–1849 (2009)
30. H. Jenke-Kodama, A. Sandmann, R. Muller, E. Dittmann, Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evolut. **22**, 2027–2039 (2005)
31. J.P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**, 754–755 (2001)
32. A. Ginolhac et al., Type I polyketide synthases may have evolved through horizontal gene transfer. J. Mol. Evolut. **60**, 716–725 (2005)
33. C. Hertweck, A. Luzhetskyy, Y. Rebets, A. Bechthold, Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. Nat. Prod. Rep. **24**, 162–190 (2007)
34. S. Okamoto, T. Taguchi, K. Ochi, K. Ichinose, Biosynthesis of actinorhodin and related antibiotics: discovery of alternative routes for quinone formation encoded in the *act* gene cluster. Chem. Biol. **16**, 226–236 (2009)
35. D. Yu, F. Xu, J. Zeng, J. Zhan, Type III polyketide synthases in natural product biosynthesis. IUBMB Life **64**, 285–295 (2012)
36. M.B. Austin, A.J.P. Noel, The chalcone synthase superfamily of type III polyketide synthases. Nat. Prod. Rep. **20**, 79–110 (2003)

37. C.T. Walsh, R.V.O. Brien, C. Khosla, Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. Angew. Chem. Int. Ed. **52**, 7098–7124 (2013)
38. R. Finking, M.A. Marahiel, Biosynthesis of nonribosomal peptides. Ann. Rev. Microbiol. **58**, 453–488 (2004)
39. S.A. Sieber, M.A. Marahiel, Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. Chem. Rev. **105**, 715–738 (2005)
40. H.W. Chen, S. O'Connor, D.E. Cane, C.T. Walsh, Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. Chem. Biol. **8**, 899–912 (2001)
41. T.L. Schneider, B. Shen, C.T. Walsh, Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. Biochemistry **42**, 9722–9730 (2003)
42. D.B. Stein, U. Linne, M.A. Marahiel, Utility of epimerization domains for the redesign of nonribosomal peptide synthetases. FEBS J. **272**, 4506–4520 (2005)
43. H. Motamedi, A. Shafiee, The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. Eur. J. Biochem. **256**, 528–534 (1998)
44. L. Tang et al., Cloning and heterologous expression of the epothilone gene cluster. Science **287**, 640–642 (2000)
45. P.S. Patel et al., Bacillaene, a novel inhibitor of prokaryotic protein-synthesis produced by *Bacillus subtilis* - Production, taxonomy, isolation, physicochemical characterisation and biological activity. J. Antibiot. **48**, 997–1003 (1995)
46. C. Scotti et al., A *Bacillus subtilis* large ORF coding for a polypeptide highly similar to polyketide synthases. Gene **130**, 65–71 (1993)
47. P.D. Straight, M.A. Fischbach, C.T. Walsh, D.Z. Rudner, R. Kolter, A singular enzymatic megacomplex from Bacillus subtilis. Proc. Natl. Acad. Sci. USA **104**, 305–310 (2007)
48. J. Moldenhauer, X.H. Chen, R. Borriss, J. Piel, Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. Angew. Chem. Int. Ed. **46**, 8195–8197 (2007)
49. J. Moldenhauer et al., The final steps of bacillaene biosynthesis in Bacillus amyloliquefaciens FZB42: direct evidence for beta, gamma dehydration by a trans-acyltransferase polyketide synthase. Angew. Chem. Int. Ed. **49**, 1465–1467 (2010)
50. K. Jensen et al., Polyketide proofreading by an acyltransferase-like enzyme. Chem. Biol. **19**, 329–339 (2012)
51. C.T. Calderone, W.E. Kowtoniuk, N.L. Kelleher, C.T. Walsh, P.C. Dorrestein, Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis. Proc. Natl. Acad. Sci. U.S.A. **103**, 8977–8982 (2006)
52. R.H. Cichewicz, F.A. Valeriote, P. Crews, Psymberin, a potent sponge-derived cytotoxin from *Psammocinia* distantly related to the pederin family. Org. Lett. **6**, 1951–1954 (2004)
53. X. Jiang, N. Williams, J.K. De Brabander, Synthesis of psymberin analogues: probing a functional correlation with the pederin/mycalamide family of natural products. Org. Lett. **9**, 227–230 (2007)
54. K.M. Fisch et al., Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. Nat. Chem. Biol. **5**, 494–501 (2009)
55. P. Pöplau, S. Frank, B.I. Morinaka, J. Piel, An Enzymatic domain for the formation of cyclic ethers in complex polyketides. Angew. Chem. Int. Ed. **50**, 13215–13218 (2013)
56. A.T. Keatinge-Clay, The structures of type I polyketide synthases. Nat. Prod. Rep. **29**, 1050–1073 (2012)
57. Y.Y. Tang, C.Y. Kim, D.E. Mathews, II, Cane, C. Khosla, The 2.7-angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc. Natl. Acad. Sci. USA **103**, 11124–11129 (2006)
58. G. Pappenberger et al., Structure of the human fatty acid synthase KS-MAT didomain as a framework for inhibitor design. J. Mol. Biol. **397**, 508–519 (2010)

59. J.G. Olsen, A. Kadziola, P. von Wettstein-Knowles, M. Siggaard-Andersen, S. Larsen, Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. Structure **9**, 233–243 (2001)
60. J. Wang et al., Platensimycin is a selective FabF inhibitor with potent antibiotic properties. Nature **441**, 358–361 (2006)
61. A. Witkowski, A.K. Joshi, Y. Lindqvist, S. Smith, Conversion of a beta-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. Biochemistry **38**, 11643–11650 (1999)
62. A. Witkowski, A.K. Joshi, S. Smith, Mechanism of the beta-ketoacyl synthase reaction catalyzed by the animal fatty acid synthase. Biochemistry **41**, 10877–10887 (2002)
63. Y.-M. Zhang, J. Hurlbert, S.W. White, C.O. Rock, 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**, 17390–17399 (2006)
64. K. Watanabe, C.C.C. Wang, C.N. Boddy, D.E. Cane, C. Khosla, Understanding substrate specificity of polyketide synthase modules by generating hybrid multimodular synthases. J. Biol. Chem. **278**, 42020–42026 (2003)
65. J.Q. Wu, K. Kinoshita, C. Khosla, D.E. Cane, 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**, 16301–16310 (2004)
66. J.E. Nixon, G.R. Putz, J.W. Porter, Synthesis of triacetic acid lactone by pigeon liver fatty acid synthase complex. J. Biol. Chem. **243**, 5471–5478 (1968)
67. P.F. Long et al., Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. Mol. Microbiol. **43**, 1215–1225 (2002)
68. G.B. Kresze, L. Steber, F. Lynen, D. Oesterhelt, Reaction of yeast fatty-acid synthetase with iodoacetamide. 3. Malonyl-coenzyme-a decarboxylase as product of reaction of fatty-acid synthetase with iodoacetamide. Eur. J. Biochem. **79**, 191–199 (1977)
69. D. Song et al., Alternative method for site-directed mutagenesis of complex polyketide synthase in Streptomyces albus JA3453. Acta Biochim. Biophys. Sinica **40**, 319–326 (2008)
70. L. Serre, E.C. Verbree, Z. Dauter, A.R. Stuitje, Z.S. Derewenda, The *Escherichia coli* malonyl-CoA-acyl carrier protein transacylase at 1.5-angstrom resolution—crystal-structure of a fatty-acid synthase component. J. Biol. Chem. **270**, 12961–12964 (1995)
71. C. Oefner, H. Schulz, A. D'Arcy, G.E. Dale, Mapping the active site of Escherichia coli malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography. Acta Crystallogr. Sect. D-Biol. Crystallogr. **62**, 613–618 (2006)
72. A.F.A. Marsden et al., Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. Science **263**, 378–380 (1994)
73. Y.Q. Cheng, J.M. Coughlin, S.K. Lim, B, Shen, in Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates. Methods in Enzymology, vol. 459 (Elsevier Academic Press Inc., New York, 2009), pp. 165–186
74. G. Yadav, R.S. Gokhale, B. Mohanty, Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. J. Mol. Biol. **328**, 335–363 (2003)
75. F.T. Wong, X. Jin, I.I. Mathews, D.E. Cane, C. Khosla, Structure and mechanism of the *trans*-acting acyltransferase from the disorazole synthase. Biochemistry **50**, 6539–6548 (2011)
76. R.H. Lambalot et al., A new enzyme superfamily—the phosphopantetheinyl transferases. Chem. Biol. **3**, 923–936 (1996)
77. B.N. Wu, Y.M. Zhang, Z. Jie, C.O. Rock, Key residues responsible for acyl carrier protein (ACP) and beta-ketoacyl-acyl carrier protein reductase (FabG) interaction. J. Biol. Chem. **278**, 52935–52943 (2004)
78. K.J. Weissman, H. Hong, B. Popovic, F. Meersman, Evidence for a protein-protein interaction motif on an acyl carrier protein domain from a modular polyketide synthase. Chem. Biol. **13**, 625–636 (2006)

79. V.Y. Alekseyev, C.W. Liu, D.E. Cane, J.D. Puglisi, C. Khosla, Solution structure and proposed domain-domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. Protein Sci. **16**, 2093–2107 (2007)
80. S.S. Chandran, H.G. Menzella, J.R. Carney, D.V. Santi, Activating hybrid modular interfaces in synthetic polyketide synthases by cassette replacement of ketosynthase domains. Chem. Biol. **13**, 469–474 (2006)
81. A. Roujeinikova et al., Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates. J. Mol. Biol. **365**, 135–145 (2007)
82. A. Busche et al., Characterisation of molecular interactions between ACP and halogenase domains in the Curacin A polyketide synthase. ACS Chem. Biol. **7**, 377–385 (2012)
83. R. Castonguay, W. He, A.Y. Chen, C. Khosla, D.E. Cane, Stereospecificity of ketoreductase domains of the 6-deoxyerythronolide B synthase. J. Am. Chem. Soc. **129**, 13758–13769 (2007)
84. A.T. Keatinge-Clay, R.M. Stroud, The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. Structure **14**, 737–748 (2006)
85. A.T. Keatinge-Clay, A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem. Biol. **14**, 898–908 (2007)
86. J. Zheng, C.A. Taylor, S.K. Piasecki, A.T. Keatinge-Clay, Structural and functional analysis of a-type ketoreductases from the amphotericin modular polyketide synthase. Structure **18**, 913–922 (2010)
87. A. Keatinge-Clay, Crystal structure of the erythromycin polyketide synthase dehydratase. J. Mol. Biol. **384**, 941–953 (2008)
88. D.L. Akey et al., Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. Structure **18**, 94–105 (2010)
89. B. Persson, J. Hedlund, H. Jornvall, The MDR superfamily. Cell. Mol. Life Sci. **65**, 3879–3894 (2008)
90. J. Zheng, D.C. Gay, B. Demeler, M.A. White, A.T. Keatinge-Clay, Divergence of multimodular polyketide synthases revealed by a didomain structure. Nat. Chem. Biol. **8**, 615–621 (2012)
91. B. Chakravarty, Z.W. Gu, S.S. Chirala, S.J. Wakil, F.A. Quiocho, Human fatty acid synthase: structure and substrate selectivity of the thioesterase domain. Proc. Natl. Acad. Sci. USA **101**, 15567–15572 (2004)
92. S.C. Tsai et al., Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. Proc. Natl. Acad. Sci. USA **98**, 14808–14813 (2001)
93. J. Brink et al., Quaternary structure of human fatty acid synthase by electron cryomicroscopy. Proc. Natl. Acad. Sci. USA **99**, 138–143 (2002)
94. T. Maier, S. Jenni, N. Ban, Architecture of mammalian fatty acid synthase at 4.5 angstrom resolution. Science **311**, 1258–1262 (2006)
95. N.A. Schnarr, A.Y. Chen, D.E. Cane, C. Khosla, Analysis of covalently bound polyketide intermediates on 6-deoxyerythronolide B synthase by tandem proteolysis-mass spectrometry. Biochemistry **44**, 11836–11842 (2005)
96. R.J. Cox et al., Post-translational modification of heterologously expressed Streptomyces type II polyketide synthase acyl carrier proteins. FEBS Lett. **405**, 267–272 (1997)
97. J.J. Thompson, *Rays of positive electricity and their application to chemical analysis* (Longmans, London, 1913)
98. G. Squires, Francis Aston and the mass spectrograph. Dal. Trans. 3893–3899, (1998)
99. E. Hoffman, V. Stroobant, *Mass Spectrometry: Principles and Applications* (Wiley, England, 2007)
100. J.A. Loo, Studying noncovalent protein complexes by electrospray ionization mass spectrometry. Mass Spectrom. Rev. **16**, 1–23 (1997)

101. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Electrospray ionization for mass-spectrometry of large biomolecules. Science **246**, 64–71 (1989)
102. G. Taylor, Disintegration of water drops in electric field. Proc. R. Soc. A **280**, 383–397 (1964)
103. R.B Cole, Electrospray and MALDI Mass Spectrometry: Fundamentals, Instrumentation, Practicalities, and Biological Applications: Fundamentals, Instrumentation, and Applications, 2nd edn (Wiley-Blackwell, UK, 2010)
104. J.H. Gross, Mass Spectrometry: A Textbook. 2nd edn (Springer, Berlin, 2004)
105. A. Gomez, K.Q. Tang, Charge and fission of droplets in electrostatic sprays. Phys. Fluids **6**, 404–414 (1994)
106. P. Kebarle, U.H. Verkerk, Electrospray: from ions in solution to ions in the gas phase, what we know now. Mass Spectrom. Rev. **28**, 898–917 (2009)
107. M. Dole, L.L. Mack, R.L. Hines, Molecular beams of macroions. J. Chem. Phys. **49**, 2240 (1968)
108. J.V. Iribarne, B.A. Thomson, Evaporation of small ions from charged droplets. J. Chem. Phys. **64**, 2287–2294 (1976)
109. M. Wilm, M. Mann, Analytical properties of the nanoelectrospray ion source. Anal. Chem. **68**, 1–8 (1996)
110. B.T. Ruotolo et al., Evidence for macromolecular protein rings in the absence of bulk water. Science **310**, 1658–1661 (2005)
111. C. Uetrecht et al., High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. Proc. Natl. Acad. Sci. USA **105**, 9216–9220 (2008)
112. W.Paul, H. Steinwedel, vol. 8 (Zeitschrift fur Naturforschung, 1953), pp. 448–450
113. W.C. Wiley, I.H. McLaren, Time-of-flight mass spectrometer with improved resolution. Rev. Sci. Instrum. **26**, 1150–1157 (1955)

# Chapter 2
# Materials and Methods

## 2.1 Materials

### *2.1.1 Instruments*

#### 2.1.1.1 Protein Production, Purification and Mutagenesis

Cell lines were transformed using a **MiniPulser electroporation system** (BioRad), and growth plates were incubated in a **WTC incubator** (Binder). Cell lines were grown in a **Series 25 orbital incubator** (New Brunswick), and optical densities of cell cultures were determined using a **UV 1101 Biotech photometer** (WPA). Cell culture centrifugations were carried out in an **Evolution RC centrifuge** (Sorvall), using either a SLC-3000 or SS-34 rotor. Protein concentration centrifugations were conducted in a **CS-15R centrifuge** (Beckman). Cell lysis was performed on a **Soniprep 150 sonicator** (Sanyo), and proteins were purified using a **HiTrap® Chelating Column** (GE Healthcare) attached to an **ÄKTA Prime** (Amersham Pharmacia Biotech). Protein samples were analysed by SDS-PAGE using a **Mini-Protean protein electrophoresis system** (BioRad), and gel images were taken in a **G:Box gel imager** (Syngene). Polymerase chain reactions were conducted in a **Techgene thermocycler** (Techne), with DNA and protein concentrations determined by UV using a **Nanodrop800** (Thermo Scientific).

#### 2.1.1.2 Mass Spectrometry

Waters SYNAPT HDMS

All mass spectrometry measurements described within this study were performed using a **Waters SYNAPT HDMS**, which is a hybrid quadrupole-ion mobility-orthogonal acceleration TOF instrument (oa-TOF). This instrument has multiple



M. Jenner, *Using Mass Spectrometry for Biochemical Studies on Enzymatic Domains from Polyketide Synthases*, Springer Theses,
DOI 10.1007/978-3-319-32723-5_2

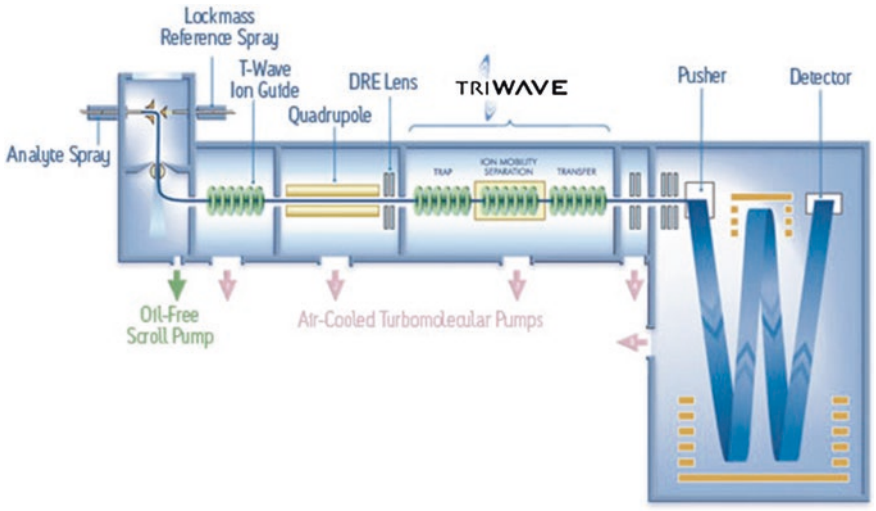



**Fig. 2.1** Schematic of a Waters SYNAPT HDMS mass spectrometer. Image taken from Waters website

ionisation capabilities, of which ESI and nano-ESI were utilised during this study. The instrument schematic shown in Fig. 2.1 highlights the position of various components within the mass spectrometer. The quadrupole is located immediately after the first ion guide, and has two modes of operation. In RF mode, the quadrupole transfers all ions into the Triwave region of the instrument. However, when DC voltages are applied to the roads, the quadrupole can be used as a mass filter (as described in Sect. 1.2.3.1). The resolving power of the quadrupole can be manually altered using the MassLynx software to generate wide/narrow isolation windows.

Collision induced activation was performed in the trap region of the instrument, which is a 10 cm cell consisting of stacked ring electrodes, and utilises an RF current to focus the ions. The cell is filled with an inert buffer gas, in this case Argon, to promote the collisional activation process. The kinetic energy applied to ions in this region is the product of the charge state and the voltage applied to accelerate the ion into the trap. This can be controlled by varying the potential difference between the first ion guide and the trap (Fig. 2.1).

### Nanospray Capillaries

NanoESI capillaries were prepared from **borosilicate glass capillaries (OD = 1.00 mm, ID = 0.78 mm, length = 10 cm)** (Warner), using a **Flaming/Brown P-97 micropipette puller** (Sutter Instruments). The pulled tip was then coated in gold for electrical conductivity using a vacuum evaporator built in-house (see Sect. 2.2.9).

## *2.1.2 Buffers and Reagents*

### 2.1.2.1 Reagents

**Ultrapure water** (18.2 MΩ·cm), was obtained from a Millipore water purification system and was used for preparation of sample solutions. **Analytical Grade Acetonitrile (MeCN)** (Fischer Scientific) and **Trifluoroacetic Acid (TFA)** (Sigma-Aldrich) were used for all ZipTip buffers (see Sect. ''Denaturing MS—ZipTip Buffers''). The **coenzyme A derivatives (acetyl-, butyryl- and malonyl)** used in assays were purchased from Sigma-Aldrich. ***N*-acetylcysteamine (SNAC) thioesters** were provided by the Piel lab (ETH, Zurich), and stock solutions were solubilised in **dimethyl sulfoxide (DMSO)** (Fischer Scientific). **Iodoacetamide** (Sigma-Aldrich) was used for all alkylation reactions.

### 2.1.2.2 Transformation and Expression Media

**Luria Agar (LA)** (Sigma-Aldrich) was used to grow cell colonies prior to transformations**. Millers Modification Luria Broth (LB)** (Sigma-Aldrich) was used for all cell cultures of plasmid and protein synthesis. Cell culture additives such as **Kanamycin (Kan)** (Sigma-Aldrich), **Chloramphenicol (Cm)** (Sigma-Aldrich), **Tetracycline (Tet)** (Sigma-Aldrich) and **Isopropyl-1-thio-β-D-galactopyranoside (IPTG)** (Melford) were re-suspended as concentrated stock solutions before dilution into cell culture media. Components of buffers were purchased from either Sigma Aldrich or Fischer Scientific.

### 2.1.2.3 Expression Vectors

All expression vectors were supplied by Professor Jörn Piel (ETH, Zurich). Each vector contains the cDNA sequence for the protein domain of interest. Expression conditions are detailed in Table 2.1.

### 2.1.2.4 Protein Purification Buffers and Reagents

All reagents in this section were purchased from Fischer Scientific or Sigma Aldrich. The compositions of buffers used for various aspects of protein purification are listed below:

**Re-suspension/Binding Buffer**: 25 mM Tris HCl, 500 mM NaCl, pH 7.6
**Eluting Buffer**: 25 mM Tris HCl, 500 mM NaCl, 500 mM Imidazole, pH 7.6
**Stripping Buffer**: 25 mM Tris HCl, 500 mM NaCl, 50 mM EDTA, pH 7.6
**Nickel Sulphate**: 100 mM $NiSO_4$
**SDS-PAGE Running Buffer**: 25 mM Tris HCl, 192 mM Glycine, 0.1 % SDS
**TAE Buffer**: 40 mM Tris HCl, 20 mM Acetic acid, 1 mM EDTA

**Table 2.1** Expression conditions for recombinant proteins

| Construct | Plasmid | Resistance | Inducer | Expression Temp (°C) |
|---|---|---|---|---|
| BaeJ KS1 (N-Term $His_8$-Tag) | pHis8-BaeKS1 | Kan | 1mM IPTG | 16 |
| BaeJ KS2 (N-Term $His_8$-Tag) | pHis8-BaeKS2 | Kan | 1mM IPTG | 16 |
| BaeL KS5 (N-Term $His_8$-Tag) | pHis8-BaeKS5 | Kan | 1mM IPTG | 21 |
| PsyA KS1 (N-Term $His_8$-Tag) | pHis8-PsyKS1 | Kan | 1mM IPTG | 16 |
| PsyA KS2 (N-Term $His_8$-Tag) | pHis8-PsyKS2 | Kan | 1mM IPTG | 16 |
| PsyD $KS3^0$ (N-Term $His_8$-Tag) | pHis8-PsyKS3 | Kan | 1mM IPTG | 16 |
| PsyA ACP3 (N-Term $His_8$-Tag) | pHis8-ACP3 | Kan | 0.75mM IPTG | 16 |
| PedC AT (C-Term $His_6$-Tag) | pET-24a-PedC | Kan | 1mM IPTG | 18 |
| GroES-GroEL-Tig | pG-Tf2 | Cm | 112μM Tet | 18 |
| PedD AT (N-Term $His_8$-Tag) | pHis8-PedD | Kan | 1mM IPTG | 16 |
| PsyA ACP1-KS1 (N-Term $His_6$-Tag) | pET28b-ACPKS1 | Kan | 1mM IPTG | 16 |
| PPT Svp (C-Term $His_6$-Tag) | pQE-70-Svp | Kan | 1mM IPTG | 16 |

Regions highlighted in grey indicate co-expression

SDS-PAGE gels were made at 12.5 and 15 % depending upon the size of protein being analysed. Table 2.2 details the amounts of each reagent required.

#### 2.1.2.5 Sample Preparation for Mass Spectrometry Analysis

Non-denaturing MS

All non-denaturing MS measurements were conducted in aqueous $NH_4OAc$ (Fischer Scientific) at various concentrations ranging from 5–500 mM.

**Table 2.2** Volumes of reagents required for SDS-PAGE gels

| | 12.5 % resolving | 15 % resolving | 4 % stacking |
|---|---|---|---|
| $mQH_2O$ | 3.4 mL | 2.8 mL | 3.1 mL |
| 40 % Acrylamide | 2.4 mL | 3.0 mL | 0.5 mL |
| 1.5 M Tris, pH 8.8 | 2.0 mL | 2.0 mL | 1.25 mL |
| 10 % SDS | 80 μL | 80 μL | 50 μL |
| 10 % APS | 80 μL | 80 μL | 50 μL |
| TEMED | 8 μL | 8 μL | 5 μL |
| Total vol. | 8 mL | 8 mL | 5 mL |

Denaturing MS—ZipTip Buffers

The composition of buffers used for Zip-Tip preparation and subsequent MS-analysis is detailed below:

**Equilibrium Buffer**: 0.1 % TFA in $mQH_2O$
**Elution Buffer**: 20:80 $mQH_2O$:MeCN, 0.1 % TFA.

## *2.1.3 Consumables*

### 2.1.3.1 Protein Expression and Purification

Electrocompetent cells were transformed using **MicroPulser electroporation cuvettes** (Bio-Rad). Purification of His-Tag fusion proteins was conducted using a **Hi-Trap Chelating column** (GE Healthcare), and any subsequent gel filtration steps were carried out on **Superdex 75/200 columns** (GE Healthcare) depending upon protein size. Proteins were concentrated using **Vivaspin 20 ultrafiltration spin filters** (Sartorius) with a 5 kDa or 30 kDa cut-off depending upon the size of the protein.

### 2.1.3.2 Mass Spectrometry

The **nanoESI capillaries** were prepared from borosilicate glass tubes 1.0 mm outer and 0.78 mm inner diameter (Warner Instruments). The pulled capillaries were then coated in gold using **99 % gold wire** (VWR International) using a home-built vacuum evaporator. Samples were loaded into capillaries using **GELoader tips** (Eppendorf). Sample preparation for denaturing-MS was conducted using **$C_{18}$ and $C_4$ ZipTip** (Millipore) pipette tips.

## 2.2 Methods

### *2.2.1 Protein Expression and Purification*

#### 2.2.1.1 Transformations

Electrocompetent *Escherichia coli* XL-1 Blue and BL-21 (DE3) were used for plasmid synthesis and protein synthesis respectively. Aliquots (100 μL) of electrocompetent cells were thawed on ice, and 20–100 ng of plasmid was added and mixed. Cells were kept on ice for 5 min before transferring to an electroporation cuvette, where an electrical pulse is applied to the cells. Immediately following electroporation, cells were transferred to 900 μL of LB medium and incubated at 37 °C with agitation for 1 h. Typically, 100 μL of cells were then plated onto LA plates enriched with the appropriate antibiotic (30 μg/mL Kanamycin or Chloramphenicol). Plates were incubated overnight at 37 °C, with transformation of the cells confirmed by presence of colonies, which were picked and grown for plasmid synthesis (XL-1 Blue) or protein synthesis (BL-21 (DE3)).

#### 2.2.1.2 Protein Overexpression

A single colony of transformed *Escherichia coli* BL-21 (DE3) was selected and added to LB media (10 mL) containing kanamycin (30 μg/mL) and incubated overnight at 37 °C. The resulting pre-culture was added to LB media (1 L) containing kanamycin (30 μg/mL), followed by incubation at 37 °C (see Table 2.1). Protein expression was induced by addition of IPTG (1 mM) when the optical density of the culture reached 0.6, and expression was allowed to proceed overnight at an optimized temperature (see Table 2.1). Following expression, cells were collected by centrifugation (4,000 x*g*, 15 min, 4 °C) and re-suspended in buffer (25 mM Tris-HCl, 500 mM NaCl, pH 7.6) at 5 mL/L of growth.

#### 2.2.1.3 Purification of PHis8-Fusion Proteins

*Escherichia coli* BL-21 (DE3) cells containing overexpressed pHis8-fusion proteins were lysed by sonication. The lysate was then centrifuged (37,000 x*g*, 30 min, 4 °C). The resulting supernatant was loaded onto a HiTrap® Chelating Column (GE Healthcare), pre-loaded with 100 mM $NiSO_4$ and equilibrated in re-suspension buffer. An initial wash of 5 % eluting buffer (25 mM Tris-HCl, 500 mM NaCl, 25 mM Imidazole, pH 7.6) was applied to remove contaminating proteins. A gradient of 5–100 % of eluting buffer was applied over a volume of 60 mL to elute the pHis8-fusion domain. Presence of protein in fractions was confirmed by SDS-PAGE, and an additional gel filtration step (Superdex 75/200 GE Healthcare) was applied in cases where excessive contamination was observed.

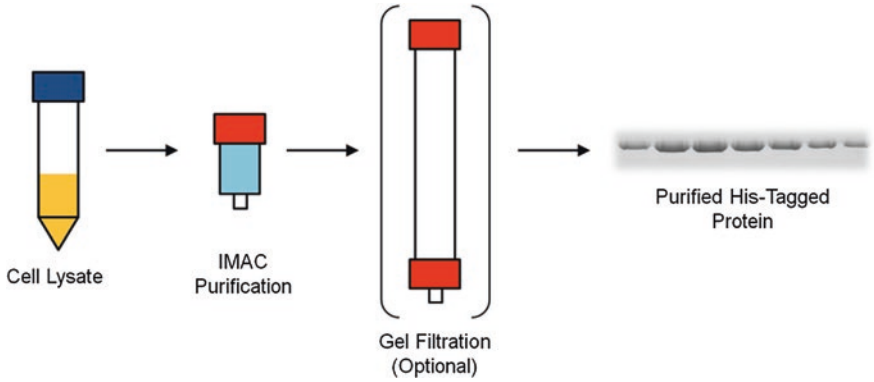



**Fig. 2.2** Principle of using His-Tag proteins and IMAC chromatography columns to purify proteins. The matrix of the HiTrap® Chelating Column forms a tetradentate complex with $Ni^{2+}$. His-Tagged proteins can then provide the final two bonds to form the stable hexadentate complex, and immobilising the protein on the column

Protein-containing fractions were pooled and concentrated using a Viva-Spin centrifugal concentrator at an appropriate MWCO (5 kDa for ACP, 30 kDa for KS). Addition of 10 % glycerol (v/v) to the resulting concentrate of pure protein allowed snap-freezing in liquid $N_2$, followed by storage at −80 °C (Fig. 2.2).

#### 2.2.1.4 Site-Directed Mutagenesis

BaeJ KS1 (N206A), (M268A), (L450A) and BaeL KS5 (M237A)

The mutant KS1 and KS5 genes were constructed using the Phusion (New England Biolabs) mutagenesis procedure. The entire expression plasmid, pHis8-BaeKS1, was amplified as a linear product using the primers in Table 2.3, which had been previously phosphorylated using polynucleotide kinase (New England Biolabs). The PCR product was digested with DpnI to remove the template DNA, gel purified from a TAE-Agarose gel using a Nucleospin® gel purification kit (Macherey-Nagel), ligated overnight using T4 Ligase (New England Biolabs) and transformed into *E. coli* strain XL1-Blue. The presence of the correct mutation was confirmed by sequencing.

BaeJ KS1 (C207A) and PedD(R97Q)

Site directed mutagenesis for the BaeJ KS1 (C207A) and PedD(R97Q) mutants were performed using a Stratagene Quikchange II kit following manufacturers protocol, using pHis8-BaeKS1 and pHis8-PedD respectively as a templates. Oligonucleotide primers used are shown below in Table 2.4.

**Table 2.3** Oligonucleotide primers for Bae KS1 and KS5 mutants

| Mutant | Primers | Mutation |
|---|---|---|
| KS1(N206A) | KS1N206A_For (5'-TGCTCATCTTCGTTAATCGGCCTGC-3') KS1N206A_Rev (5'- TGCGGCGTGGACAAAATAGCTCGG-3') | AAC → GCA (Asn → Ala) |
| KS1(M268A) | KS1M268A_For (5'-GCCATCGGCGGTGAAGGGG-3') KS1M268A_Rev (5'-GCCGTCAGCATCCGCGTCAAACG-3') | ATG → GCC (Met → Ala) |
| KS1(L450A) | KS1L450A_For (5'-GCCGGCGGAACAAACACACACG-3') KS1L450A_Rev (5'-GCCGAAAGAGCTTAAAGCCATGCGG-3') | CTT → GCC (Leu → Ala) |
| KS5(M237A) | KS5MA_For (5'-TGCTCTTCTTCTTTAACAGCCATTCATCTGG-3') KS5MA_Rev (5'- TGCCGTATCCAGAGCCATGCTCG-3') | ATG → GCA (Met → Ala) |

**Table 2.4** Oligonucleotide primers for Bae KS1(C207A) and PedD(R97Q) mutants

| Mutant | Primers | Mutation |
|---|---|---|
| KS1(C207A) | KS1C207A_For (5'- CTATTTTGTCCACGCCAAC**GC** CTCATCTTCGTTAATCGGC-3') KS1C207A_Rev (5'- GCCGATTAACGAAGATGAG**GC** GTTGGCGTGGACAAAATAG-3') | AAC → GCA (Cys → Ala) |
| PedD(R97A) | PedDR97Q_For (5'- CTGAGACTCGTGCAGAAGC**AG** GGTGATCTGATGAGTCAGG-3') PedDR97Q _Rev (5'- CCTGACTCATCAGATCACC**CT** GCTTCTGCACGAGTCTCAG-3') | CGT → CAG (Arg → Gln) |

**Table 2.5** Oligonucleotide primers for PsyA ACP3(Δ37,38)

| Mutant | Primers | Mutation |
|---|---|---|
| PsyA ACP3(Δ37,38) | ACP_For(5'-GCGTAGAGGATCGAGATCTCG-3') ACP_Rev (5'-GCATAAGCTTATACCGCTTCGAGCTGC-3') | ΔCys,Val |

PsyA ACP3(Δ37,38)

The PsyA ACP3(Δ37,38) construct was amplified from the plasmid pHis8-ACP3 which contained full-length ACP3 using the primers shown in Table 2.5. ACP_For bound upstream of the ACP3 start codon and incorporated a *BglII* restriction site present in the pHis8-ACP3 vector sequence into the PCR product. ACP_Rev replaced the Cys37 codon with an in-frame stop codon and incorporated a *HindIII* restriction site. The PCR product was digested with *BglII* and *HindIII*, seperated on a TAE-Agarose gel, excised and purified with a Genejet gel extraction kit (Fermentas). The plasmid pHis8-ACP3 was also digested with *BglII* and *HindIII*,

**Table 2.6** Oligonucleotide primers for PsyA ACP1-KS1

| Construct | Primers |
|---|---|
| PsyA ACP1-KS1 | ACPKS1_For (5'-GCGTAGAGGATCGAGATCTCG-3')<br>ACPKS1_Rev (5'-GCATAAGCTTATACCGCTTCGAGCTGC-3') |

separated on a TAE-Agarose gel, and the band corresponding to the vector fragment was excised and purified with a Genejet gel extraction kit (Fermentas). The purified *BglII—HindIII* PCR product was ligated into the purified *BglII—HindIII* pHis8-ACP3 vector fragment using T4 ligase (New England Biolabs), transformed into XL1-blue competent cells (see Sect. 2.2.1.1), and plasmids encoding PsyA ACP3(Δ37,38) was confirmed by sequencing.

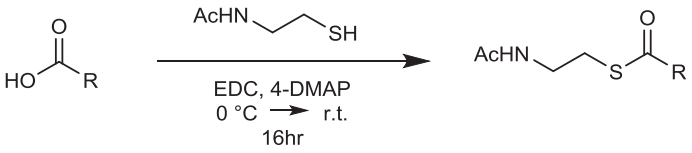
Cloning of PsyA ACP1-KS1 Didomain

The DNA sequence containing ACP and KS1 was amplified from the pPSKF1 fosmid kindly provided by Dr. Anna Vagstad (Piel lab, ETH Zurich), using the primers in Table 2.6, and conducted by José Afonso. These primers were designed in interdomain regions in locations rich in hydrophilic residues and without specific secondary structure, as predicted the PHYRE2 server [1]. The forward primer included an NdeI restriction site while the reverse primer included a Stop codon and a NotI restriction site. The PCR amplification reaction was composed of PCR reaction buffer containing $Mg^{2+}$, diluted from a 10x concentrated stock (Roche), 50 ng template DNA, 200 μM of each dNTP, 0.25 μM of each primer and 5 U of Taq DNA Polymerase (New England Biolabs), in a total volume of 50 μl.

The amplification product was cleaned using the GeneJET PCR Purification kit (Thermo Scientific) and simultaneously digested with NdeI and NotI (New England Biolabs) for 3 h to produce sticky ends. The pET28b plasmid was also double digested with these two enzymes to cut the corresponding restriction sequences within the multiple cloning site. The digested amplification product and the open pET28b vector were then loaded onto a 1 % agarose gel and purified using the Genejet gel extraction Kit (Thermo Scientific). The insert and vector were mixed with final concentrations of 40 and 120 μM, respectively, in T4 DNA ligase buffer (New England Biolabs), diluted from a 10x stock solution. For 20 uL of this reaction mixture, 400 U of T4 DNA ligase (New England Biolabs) was added and the sample was incubated overnight at room temperature. The sample was then used to transform XL1 blue cells.

## *2.2.2 Synthesis of N-Acetylcysteamine Thioesters*

SNAC compounds were obtained from the Piel Lab, ETH, Zurich. The general synthetic procedure for compounds was as follows. A solution of the




**Scheme 2.1** General coupling methodology for the synthesis of acyl-*N*-acetylcysteamine thioesters from *N*-acetylcysteamine and corresponding acid, using DCC and 4-DMAP at room temperature for 16 h

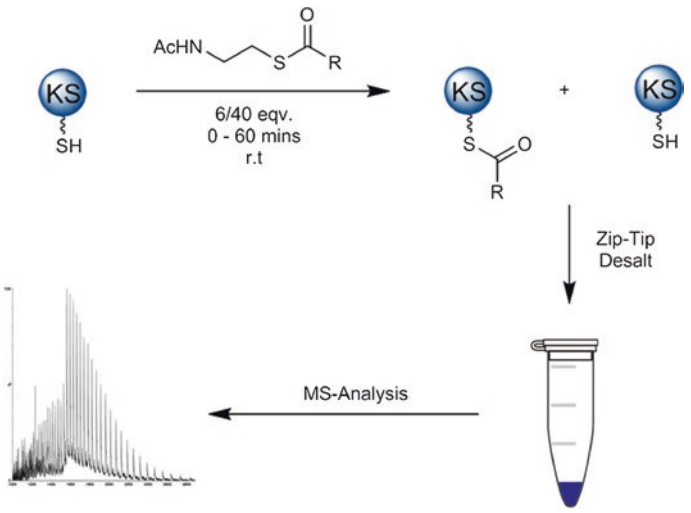
corresponding acid (5 mmol) in dichloromethane (15 mL) was cooled to 0 °C. 4-Dimethylaminopyridine (4-DMAP, 1 mmol), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide (EDC, 6 mmol) and *N*-acetylcysteamine (6 mmol) were added, and stirred overnight at 25 °C. The reaction was then quenched with saturated aqueous ammonium chloride and extracted with dichloromethane, followed by drying with $MgSO_4$ and purification by column chromatography (Scheme 2.1). Characterisation of all compounds can be found in the following publications [2–5]. Compounds were solubilised in DMSO, from which concentrated stock solutions were made, and further diluted when added to incubation reactions (5 % v/v).

### *2.2.3 Ketosynthase Acylation Assay*

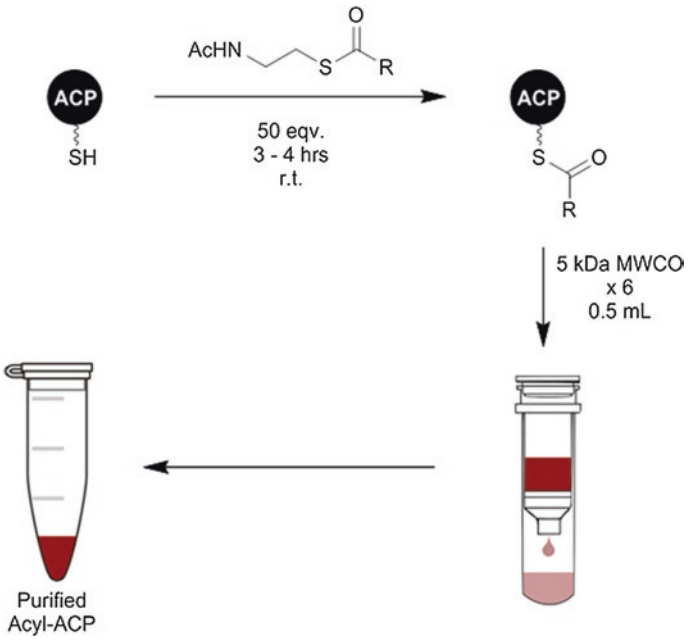
Acylation reactions were conducted in storage buffer (25 mM Tris, 500 mM NaCl, 10 % (v/v) glycerol, pH 7.6). SNAC-thioesters were incubated with KS domains at a final concentrations ranging from 0.5–2 mM for comparative substrate studies. In order to ensure the SNAC-thioesters remained in solution, the concentration of DMSO was adjusted to 5 % (v/v). Acylation reactions were allowed to run for various periods of time between 0–10 min at 25 °C, before quenching with 0.1 % TFA and ZipTip desalting (see Sect. 2.2.15.1) for MS analysis (Scheme 2.2).

### *2.2.4 Synthesis of Acyl–Acyl Carrier Proteins*

Acylation of the ACP PPant thiol was achieved by incubation of 200 μM ACP (25 mM Tris HCl, 500 mM NaCl, pH 7.4) with 10 mM SNAC thioester re-suspended in DMSO, and incubated at a 12.5 % (v/v) ratio to aid SNAC solubility. Reactions were monitored by ESI-MS at regular time intervals for the presence of the acyl-ACP adduct. Generally, acylation reactions were complete after 3–4 h at 25 °C. Removal of excess SNAC was achieved by spin filtration using 5 kDa MWCO columns (Scheme 2.3).




**Scheme 2.2** Standard workflow for performing KS acylation assays followed by MS analysis




**Scheme 2.3** Workflow for the synthesis of acyl-ACPs from SNAC thioesters, and subsequent purification

### 2.2.5 Acyl-ACP Ketosynthase Loading Assay

KS loading assays were conducted in storage buffer (25 mM Tris, 500 mM NaCl 10 % (v/v) glycerol, pH 7.6). Typically, 20 μM KS was incubated with 100 μM acyl-ACP and the reaction was allowed to proceed at 25 °C. Aliquots were removed at various time points between 2–32 h, and subsequent ZipTip clean-up of the samples allowed MS analysis of each time point (see Sect. 2.2.15.1).

### 2.2.6 PedC/PedD(R97Q) Hydrolase Assays

Hydrolase assays were conducted in storage buffer (25 mM Tris, 500 mM NaCl, 10 % v/v glycerol, pH 7.4) in the presence of GroEL. Typically, PedC (5 μM) was incubated with acyl-ACPs (20 μM) in a 30 μL reaction for various periods of time ranging from 0.5–60 min at 25 °C. The reaction was quenched using 0.1 % TFA, followed by ZipTip desalting for MS analysis (see Sect. 2.2.15.1).

### 2.2.7 Acyltransferase Extender Unit Specificity Assays

Stocks of malonyl- methylmalonyl- and acetyl-CoA were made to 20 mM in milliQ $H_2O$, snap frozen and kept at −80 °C. Incubations with AT domains were conducted in storage buffer (25 mM Tris, 500 mM NaCl, 10 % v/v glycerol, pH 7.6) with addition of required amount of extender unit. Reactions were allowed to run for 10 min at 25 °C, before quenching with 0.1 % TFA and ZipTip desalting (see Sect. 2.2.15.1) for MS analysis.

### 2.2.8 PedD Malonyl Loading and Unloading of ACP

PedD catalysed malonyl loading and unloading of the ACP was conducted in storage buffer (25 mM Tris, 500 mM NaCl, 10 % v/v glycerol, pH 7.6). In the case of malonyl loading, 80 μM PedD was incubated with 4 equivalents (320 μM) of malonyl-CoA and allowed to react for 10 min at 25 °C. Excess malonyl-CoA was removed by spin filtration using 30 kDa MWCO columns. The resulting retentate was then incubated with 40 μM *holo*-ACP at a 1:1 molar ratio. The transfer reaction was allowed to proceed for 20 min, with aliquots removed at various time intervals, quenched with 0.1 % TFA and desalted using a $C_4$ ZipTip (see Sect. 2.2.15.1). The unloading reaction was conducted under exactly the same conditions, except PedD was not pre-loaded with malonyl. The ACP was pre-loaded malonyl following the procedure detailed in Sect. 2.2.9.

### 2.2.9 Ketosynthase Elongation Assay

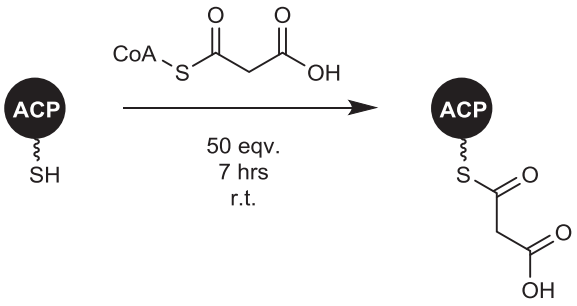
The ketosynthase-catalysed chain elongation assay described in Chap. 6 required multiple preparative synthetic steps before the final assay could be conducted. This section details the required synthetic steps and the final assay procedure.

#### 2.2.9.1 Synthesis of Malonyl-ACP

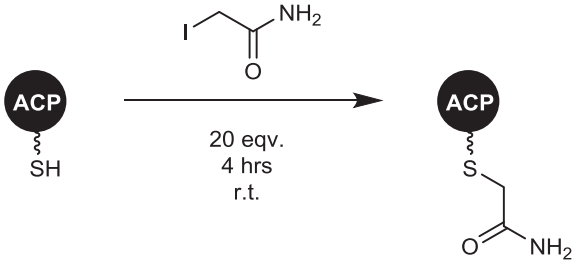
Malonylation of the ACP PPant thiol was achieved by incubation of 200 μM ACP (25 mM Tris HCl, 500 mM NaCl, pH 7.4) with 8 mM malonyl-CoA (Scheme 2.4). Reactions were monitored by ESI-MS at regular time intervals for the presence of the malonyl-ACP adduct. Reactions were complete after 7 h at 25 °C.

#### 2.2.9.2 Synthesis of Alkyl-ACP

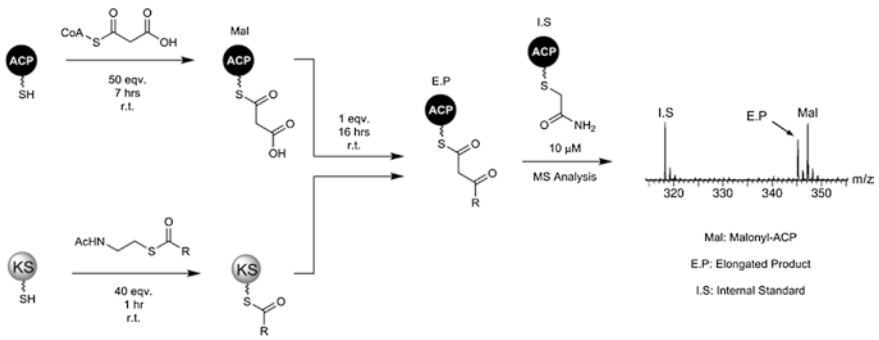
The synthesis of alkyl-ACP, for use as an internal standard, was achieved by incubation of 200 μM ACP (25 mM Tris HCl, 500 mM NaCl, pH 7.4) with 4 mM iodoacetamide, in the absence of light (Scheme 2.5). Reactions were monitored by ESI-MS at regular time intervals for the presence of the alkyl-ACP adduct. Alkylation reactions were complete after 4 h at 25 °C, and excess iodoacetamide was removed by spin filtration using 5 kDa MWCO columns.




**Scheme 2.4** Synthesis of malonyl-ACP from malonyl-CoA and *holo*-ACP




**Scheme 2.5** Synthesis of alkyl-ACP from iodoacetamide and *holo*-ACP, in the absence of light



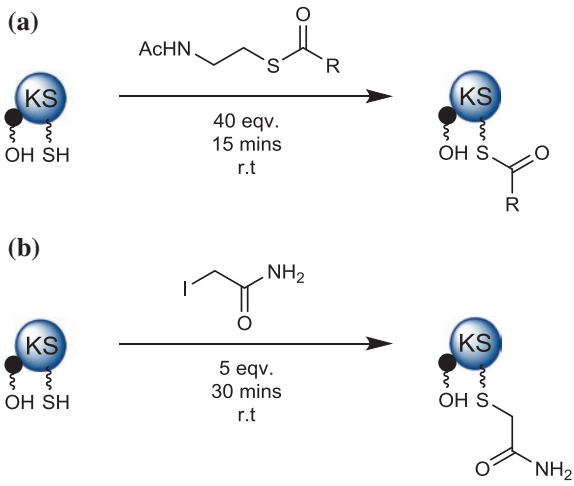

**Scheme 2.6** Overview of the MS-based elongation assay for probing the substrate specificity of KS domains using a range of *N*-acetylcysteamine thioesters and malonyl-ACP. Following incubation of acyl-KS with malonyl-ACP the reaction is quenched and alkylated ACP added as an internal standard. MS/MS analysis reveals the extent of elongation by detecting and quantifying β-ketoacyl-ACP

#### 2.2.9.3 Ketosynthase Elongation Reactions and MS Analysis

KS elongation assays were conducted in storage buffer (25 mM Tris, 500 mM NaCl, 10 % v/v glycerol, pH 7.4). Typically, the KS was pre-acylated by incubation with SNAC thioester (4 mM) for 2 h, followed by concentration to 100 μM using a spin concentrator. Malonyl-ACP was produced by incubation of PsyA ACP3(Δ37,38) (100 μM) with malonyl-CoA (8 mM) for 6 h at 25 °C. The resulting acyl-KS and malonyl-ACP solutions were subsequently mixed in a 1:1 ratio yielding a reaction solution containing 50 μM acyl-KS and 50 μM malonyl-ACP. The reaction was allowed to proceed for 16 h, followed by addition of 10 μM alkyl-ACP and subsequent ZipTip desalting and MS analysis (Scheme 2.6). During MS analysis, the $7^+$ charge state of the ACP was isolated ($m/z = 1579$) with a low quadrupole resolution to encompass all species of this charge state. Activation was achieved by applying collision energy of 20 V to the trap region of the instrument.

### *2.2.10 Monitoring Acyl-Transfer in PsyA ACP1-KS1*

The ability to monitor the location of an acyl chain in the PsyA ACP1-KS1 didomain required several pre-synthesised species. Acyl-ACP1-KS1 was used as a starting point for reverse transfer assays, and alkyl-ACP1-KS1 was used to inactive the active site Cys of the KS, therefore acting as a control for subsequent reactions. In this section, the synthesis of these species is described in addition to the MS-based assay.




**Scheme 2.7** Synthesis of **a** Acyl-PsyA ACP1-KS1 using SNAC thioesters and **b** Alkyl-PsyA ACP1-KS1 under mild conditions using iodoacetamide in the absence of light

#### 2.2.10.1 Acylation and Alkylation of PsyA ACP1-KS1

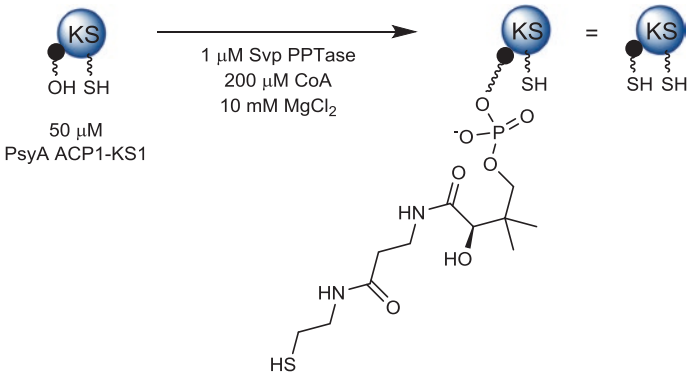
Acylation of PsyA ACP1-KS1 was conducted in a similar manner to that of the ketosynthase acylation assay described in Sect. 2.2.3, followed by removal of excess SNAC using 30 kDa MWCO spin filters (Scheme 2.7a). Alkylation of the KS active-site Cys from the ACP1-KS1 didomain was performed under mild conditions to alkylate the active site Cys residue only, by incubation with 5 equivalents of iodoacetamide in the absence of light (Scheme 2.7b).

#### 2.2.10.2 Phosphopantheinylation of PsyA ACP1-KS1

The PPTase-catalysed phosphopantetheinyl loading of PsyA ACP1-KS1 was carried out in accordance with previously published methods [6]. Minimal amounts of Svp PPTase from *Streptomyces verticillus* were used to reduce the complexity and heterogeneity of the subsequent mass spectra. Optimised conditions allowed 1 μM Svp PPTase to be used for complete loading of 50 μM PsyA ACP1-KS1, in the presence of 10 mM $MgCl_2$ and 200 μM CoA (Scheme 2.8).

#### 2.2.10.3 Monitoring Reverse Acyl-Transfer in PsyA ACP1-KS1

Pre-acylated *apo*-ACP1-KS1 was prepared as described in Sect. 2.2.12. Subsequent addition of Svp PPtase, CoA and $MgCl_2$ generated




**Scheme 2.8** The enzymatic attachment of phosphopantetheine from co-enzyme A to a conserved serine residue on *apo*-ACP1-KS1, catalysed by the Svp PPTase

*holo*-acyl-ACP1-KS1, which was allowed to equilibrate for 10 min. The resulting reaction was then desalted using a $C_4$ ZipTip into 80:20 MeCN: $mQH_2O$, 0.1 % TFA for MS analysis. The location of the acyl unit was monitored using the PPant ejection assay, isolating the $73^+$ charge state of *holo*-acyl-ACP1-KS1 and application of 12 V in the trap region of the MS. Measurement of the ratio between unmodified PPant (*m/z* 261.2) and acyl-PPant species allowed the location of the acyl chain to be interrogated (Scheme 2.9).

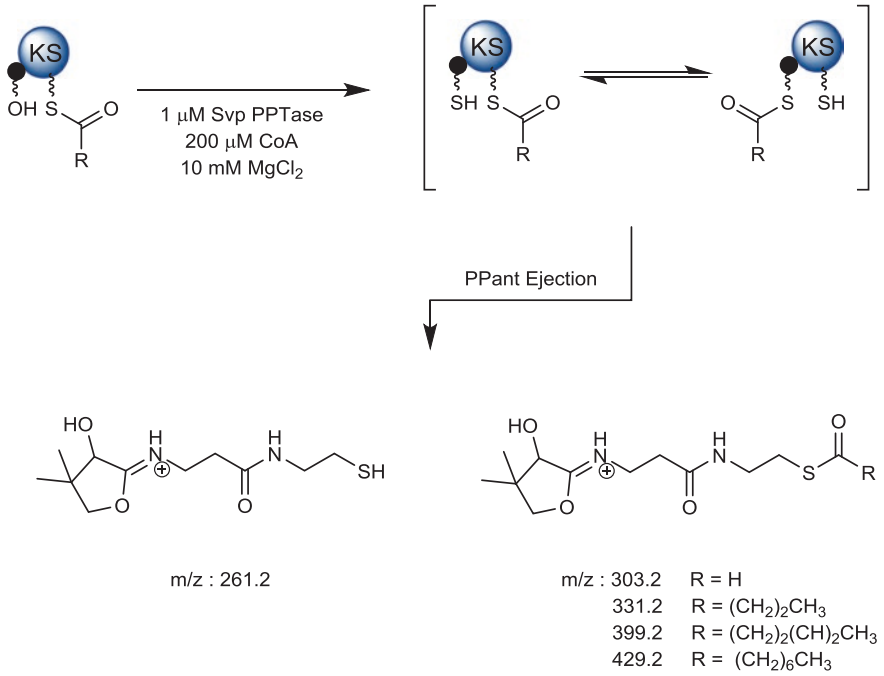
## *2.2.11 Sample Preparation for Mass Spectrometry*

### 2.2.11.1 Zip-Tip Desalting

Both $C_{18}$ and $C_4$ ZipTips™ were used to prepare samples; the $C_{18}$ was primarily used for desalting of the ACP, whereas $C_4$ was used for KS and AT domains. The ZipTip was washed with two 10 μL aspirations of 50 % MeCN, followed by five 10 μL aspirations of $H_2O$/0.1 % TFA solution. The protein sample was then loaded onto the ZipTip column by 15–20x 10 μL aspirations of the sample, each time injecting back into the sample vial. The loaded protein sample was then desalted by x15 10 μL aspirations of 0.1 % TFA, followed by elution of the sample into 8 μL $H_2O$/80 % MeCN/0.1 % TFA.

### 2.2.11.2 Viva-Spin MWCO Desalting

Proteins sprayed from non-denaturing conditions required buffer exchanging into $NH_4OAc$. This was achieved by x5 concentration cycles at 11,500 x*g*, from 0.5 mL → 50 μL using Vivaspin 500 ultrafiltration columns (Sartorius).




**Scheme 2.9** Overview of the MS-based method for monitoring acyl transfer within PsyA ACP1-KS1. Following loading of CoA-SH or acyl-CoA, the PPant ejection assay is used to identify the location of the acyl unit

### *2.2.12 Pulling and Gold-Coating Nanospray Capillaries*

Nano-ESI capillaries were pulled using a Flaming/Brown P-97 micropipette puller (Sutter Instruments). In order to achieve a tip shape shown in Fig. 2.3, the parameters were carefully optimised by to obtain a satisfactory tip shape [7]. Although parameters can change when using different filaments, a typical set of values are detailed in Table 2.7.

Once pulled, the capillaries were coated in gold using a home-built vacuum evaporator, depicted in (Fig. 2.4). Pre-pulled capillaries were placed into a supporting rack and placed in the evaporation chamber, directly underneath the evaporation boat. Gold wire (~20 mg) was then placed into the evaporation boat, and a vacuum (~$1.2 \times 10^{-4}$ mbar) was applied to the chamber. Once a suitable vacuum was achieved, 50 V was applied to the evaporation boat using an autotransformer and evaporation was allowed to proceed. Following evaporation (typically 2 min), the autotransformer was set to 0 V, and the chamber was vented using the vent valve. This procedure was performed twice to coat each side of the capillaries.

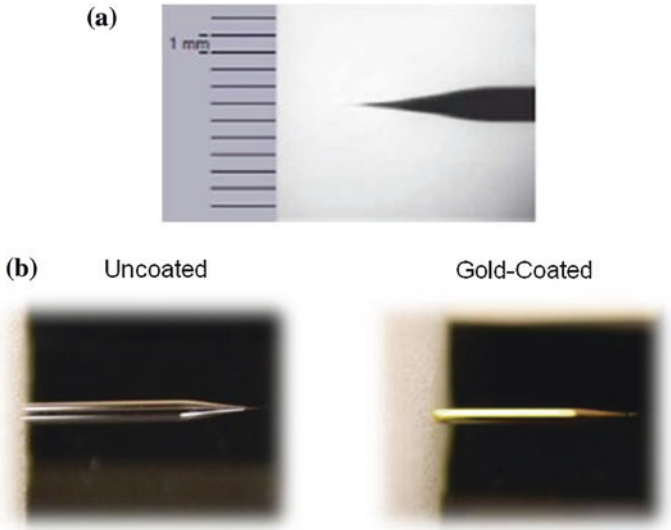



**Fig. 2.3** **a** Mircroscope view of optimal tip shape for nESI capillaries. Taken from [4]. **b** Photographs of nESI capillaries pulled in-house, before and after gold coating

**Table 2.7** Instrument parameters for Flaming/ Brown P-97 micropipette puller

| | | |
|---|---|---|
| Heat | 750 | x3 cycles |
| Time | 80 | |
| Pull velocity | 15 | |

### *2.2.13 Mass Spectrometry Instrument Parameters*

Instrument parameters were optimised for each experiment to provide the highest relative signals. For experiments requiring collision induced activation, the voltage applied in the trap region was optimised such that minimal secondary fragmentation of ejected ions occurred (Table 2.8).

### *2.2.14 Calculation of Acyl-KS Concentrations Using MS*

Denatured KS spectra obtained from acylation experiments were subjected to minimal smoothing and noise reduction. The spectra were deconvoluted using the transform function of MassLynx, taking an average of all charge states. Relevant peak intensities were recorded and converted into the concentration of [KS-SH] and [KS-acyl] respectively. Data was recorded in triplicate, and an average taken for kinetic plots.


**Fig. 2.4** Photograph of home-built vacuum evaporator. Components are labelled as follows: **a** Evaporation chamber. **b** Evaporation boat. **c** Vent valve. **d** Autotransformer. **e** Vacuum gauge controls and read-out

**Table 2.8** Mass spectrometry parameters for analysis of protein domains

| | KS and AT domains | ACP domain |
|---|---|---|
| *Voltages* | | |
| Capillary voltage (kV) | 1.5 | 1.5 |
| Sample cone voltage (V) | 20 | 10 |
| Extraction cone voltage (V) | 5 | 5 |
| Trap collision voltage (V) | 6 | 6–20† |
| Transfer collision voltage (V) | 5 | 5 |
| *Pressures* | | |
| Backing pressure (mbar) | 1.6–1.8 | |
| Trap pressure (mbar) | $1.4 \times 10^{-2}$ | |
| TOF pressure (mbar) | $1.5 \times 10^{-6}$ | |
| *Quadrupole profile* | | |
| m/z | Dwell time (% scan time) | Ramp time (% scan time) |
| 800 | 10 | 30 |
| 1500 | 40 | 10 |
| 3000 | – | – |

†Activation energies varied between experiments. For experiments detailed in Sect. 2.2.11, an activation energy of 20 V was applied to the $7^+$ charge state of PsyA ACP3. For experiments detailed in Sect. 2.2.11, an activation energy of 12 V was applied to the $73^+$ charge state of PsyA ACP1-KS1

### *2.2.15 Structure Prediction*

#### 2.2.15.1 Homology Modelling of Ketosynthase Domains

Homology models of all KS domains in this study were constructed using the CPHmodel server [8], using published KS structures 2QO3 and 2HG4 as templates [9, 10]. Further refinement of the model was achieved by energy minimisation using the YASARA server [11]. Ramachandran plots were produced using the RAMPAGE server to assess the quality of the models (<1 % of residues were found to lie outside the favoured regions in all models, and none of these were close to the active site) [12]. In the case of WT BaeL KS5, the full length biosynthetic intermediate of KS5, including the phosphopantetheine chain, was constructed in the PRODRG server [13]. A similar substrate was generated for BaeL KS5(M237A), which included an additional β-methyl branch. Both substrates were manually docked into the binding site of each domain, and re-submitted to the YASARA energy minimisation server. Homology models with acyl chains attached to the active site Cys were produced using the crystal structure of a FAS KS domain with dodecanoic acid bound as a template [14].

#### 2.2.15.2 Homology Modelling of PedC and PedD

Homology models of PedC and PedD domains were generated using the CPHmodel server, using published AT structures 3RGI and 3BSM as templates [15, 16]. Ramachandran plots were produced using the RAMPAGE server to assess the quality of the models. Addition of a malonyl unit to the active site Ser was achieved by alignment of the respective homology models to the *E.coli* FabD, 2G2Z, structure complexed with malonyl and CoA [17].

## References

1. L.A. Kelley, M.J.E. Sternberg, Protein structure prediction on the web: a case study using the Phyre server. Nat. Protoc. **4**, 363–371 (2009)
2. I.H. Gilbert et al., Synthesis of beta-keto and alpha, beta-unsaturated n-acetylcysteamine thioesters. Bioorgan. Med. Chem. Lett. **5**, 1587–1590 (1995)
3. O. Vergnolle, F. Hahn, A. Baerga-Ortiz, P.F. Leadlay, J.N. Andexer, Stereoselectivity of isolated dehydratase domains of the borrelidin polyketide synthase: implications for *cis* double bond formation. ChemBioChem **12**, 1011–1014 (2011)
4. M. Jenner et al., Substrate specificity in ketosynthase domains from *trans*-AT polyketide synthases. Angew. Chem. Int. Ed. **52**, 1143–1147 (2013)
5. C. Kohlhaas et al., Amino acid-accepting ketosynthase domain from a *trans*-AT polyketide synthase exhibits high selectivity for predicted intermediate. Chem. Sci. **4**, 3212–3217 (2013)
6. C. Sanchez, L.C. Du, D.J. Edwards, M.D. Toney, B. Shen, Cloning and characterisation of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. Chem. Biol. **8**, 725–738 (2001)

7. H. Hernandez, C.V. Robinson, Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. Nat. Protoc. **2**, 715–726 (2007)
8. M. Nielsen, C. Lundegaard, O. Lund, T.N. Petersen, CPHmodels-3.0-remote homology modeling using structure-guided sequence profiles. Nucleic Acids Res. **38**, W576–W581 (2010)
9. Y.Y. Tang, C.Y. Kim, II. Mathews, D.E. Cane, C. Khosla, The 2.7-angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc. Natl. Acad. Sci. U.S.A. **103**, 11124–11129 (2006)
10. Y.Y. Tang, A.Y. Chen, C.Y. Kim, D.E. Cane, C. Khosla, Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. Chem. Biol. **14**, 931–943 (2007)
11. E. Krieger et al., Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. Proteins: Struct. Funct. Bioinform. **77**, 114–122 (2009)
12. S.C. Lovell et al., Structure validation by C alpha geometry: phi, psi and C beta deviation. Proteins: Struct. Funct. Genet. **50**, 437–450 (2003)
13. A.W. Schuttelkopf, D.M.F. van Aalten, PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr. Sect. D-Biol. Crystallogr. **60**, 1355–1363 (2004)
14. J.G. Olsen, A. Kadziola, P. von Wettstein-Knowles, M. Siggaard-Andersen, S. Larsen, Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. Structure **9**, 233–243 (2001)
15. F.T. Wong, X. Jin, I.I. Mathews, D.E. Cane, C. Khosla, Structure and mechanism of the *trans*-acting acyltransferase from the disorazole synthase. Biochemistry **50**, 6539–6548 (2011)
16. A.A. Fedorov et al., 3BSM: crystal structure of D-mannonate dehydratase from *Chromohalobacter salexigens* complexed with Mg, D-mannonate and 2-keto-3-deoxy-D-gluconate. Protein Data Bank (2008)
17. C. Oefner, H. Schulz, A. D'Arcy, G.E. Dale, Mapping the active site of Escherichia coli malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography. Acta Crystallogr. Sect. D-Biol. Crystallogr. **62**, 613–618 (2006)

# Chapter 3
# Substrate Specificity of Ketosynthase Domains Part I: β-Branched Acyl Chains

## 3.1 Introduction

Recent phylogenetic work has shown that KS domains from *trans*-AT PKSs correlate, at the sequence level, with their predicted biosynthetic intermediates [1]. The extent of this evolution-based specificity is believed to reach as far as the β-position for each given biosynthetic intermediate, as detailed in section "KS Specificity-Based Assignment of *trans*-AT PKSs". Compared to textbook colinearity rules, employed for *cis*-AT PKS systems, a KS specificity-based approach works remarkably well as a predictive method for assigning biosynthetic intermediates to their PKS, however it lacks functional testing.

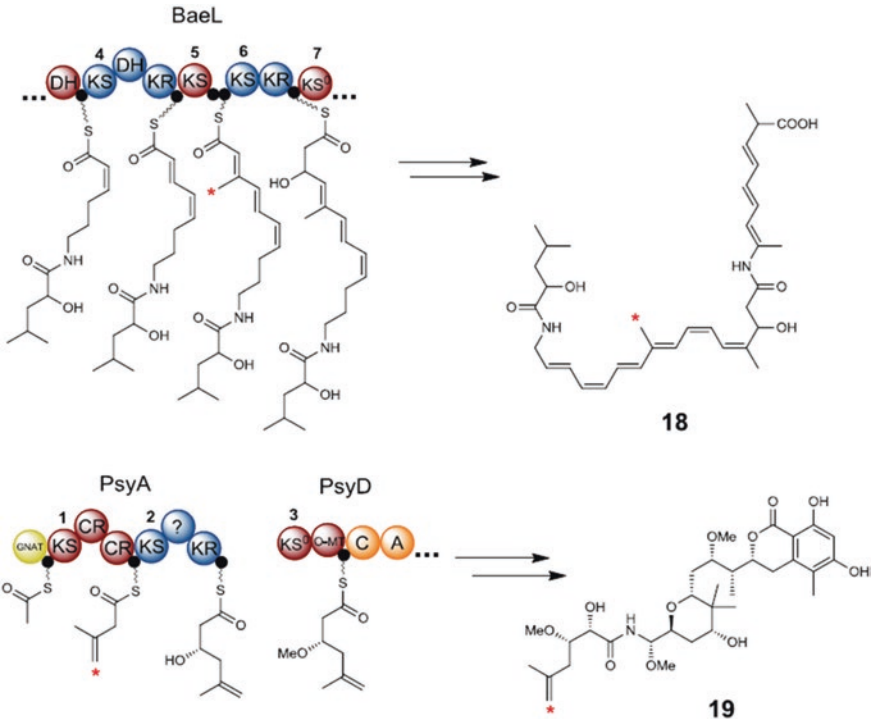
In order to probe the functional substrate specificity of KS domains from *trans*-AT PKSs, a mass spectrometry (MS)-based methodology was developed to test a selection of KS-domains. Previous approaches designed to examine KS specificity from *cis*-AT systems have focussed upon radioactivity assays and MS analysis of trypsin-digested domains [2, 3]. The assay reported in this chapter possesses significant advantages over previous methods, allowing rapid and direct analysis of intact KSs domain and ability to observe exact mass-shifts for the attachment of acyl chains.

Herein, the substrate specificity of BaeL KS5 from the bacillaene (**18**) cluster, and KS1, KS2 and $KS3^0$ from the psymberin (**19**) PKS are reported (Scheme 3.1). Using the data from these assays, in conjunction with homology modelling, the molecular rules dictating the ability of KS domains to accept carbon based β-branched substrates is revealed.

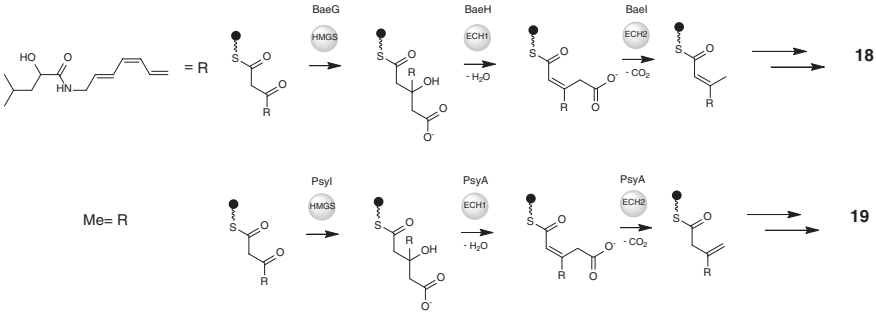
BaeL KS5 and PsyA KS2 are located immediately before and after β-branching enzymes, which install a β-methyl and a β-methylene branch respectively (Scheme 3.2). In each instance, the mechanism of β-branch formation is essentially the same. The synthesis commences with decarboxylation of malonyl-ACP by a non-elongating KS, to yield acetyl-ACP. A 3-hydroxy-3-methylglutaryl-CoA



M. Jenner, *Using Mass Spectrometry for Biochemical Studies on Enzymatic Domains from Polyketide Synthases*, Springer Theses,
DOI 10.1007/978-3-319-32723-5_3



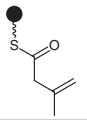
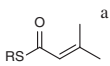
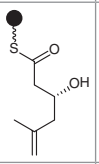
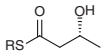
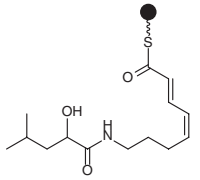
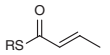

**Scheme 3.1** Partial proposed biosynthetic schemes for bacillaene (**18**) and psymberin (**19**)




**Scheme 3.2** Mechanisms of installation of β-methyl and β-methylene branches during the biosynthesis of **18** and **19**

synthase (HMGS) then catalyses an aldol addition between the acetate unit and the β-keto group of the growing polyketide chain. This is followed by dehydration and decarboxylation by two enoyl-CoA dehydratases (ECH) [4]. During installation of the β-methyl branch in **18**, the HMGS and ECH enzymes all act in *trans* [5],

**Table 3.1** The assigned clades and predicted substrates for KS domains used in this study

| Ketosynthase | Clade | Predicted Intermediate | SNAC Mimic |
|---|---|---|---|
| PsyA KS1 | VI |  |  |
| PsyA KS2 | Iβ |  |  [a] |
| PsyD KS3$^0$ | III |  |  |
| BaeL KS5 | IX |  |  |

R = 

[a]Due to the instability of the β,γ-unsaturated SNAC, the α, β-unsaturated, β-Me SNAC was used as a substrate mimic

whereas during the synthesis of the β-methylene branch in **19** the ECH, or crotonase (CR), domains are integral to the PsyA protein (Scheme 3.2) [6].

*N*-acetyl cysteamine (SNAC) thioester substrate mimics were used to test the specificity of four KS domains with a range of predicted biosynthetic intermediates (Table 3.1). In the case of PsyD KS3$^0$ and BaeL KS5, the synthesised SNAC mimic are shorter in length than the predicted biosynthetic intermediate. However, the phylogenetic KS-based substrate prediction would suggest that the specificity does not extend past the β-position [1], and therefore SNAC substrate mimics synthesised to this length may be sufficient to reveal KS specificity.

## 3.2 Results and Discussion

### *3.2.1 Purification of Ketosynthase Domains*

Psy KS1, KS2, KS3$^0$, BaeL KS5 and KS5(M237A) were used in this chapter, and were expressed and purified as His-Tag fusion proteins as described in Sect. 2.2.1. Post-purification, KS domains were analysed by SDS-PAGE and by mass spectrometry. An SDS-PAGE gel of all the KS domains used in this chapter is shown in Fig. 3.1. The subsequent mass spectra are shown in Fig. 3.2.

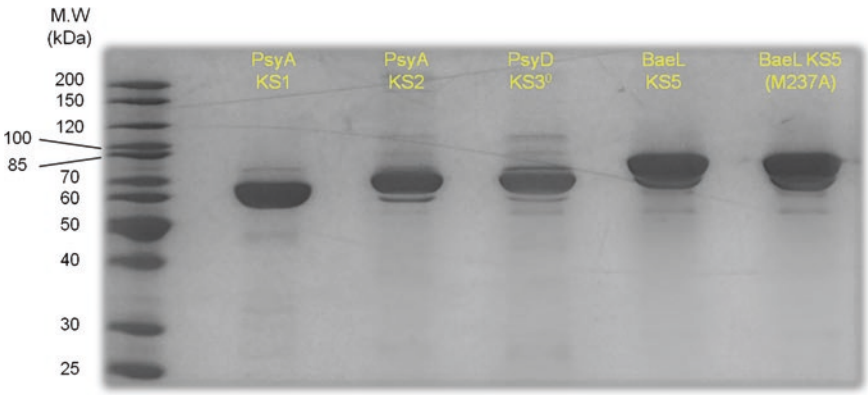



**Fig. 3.1** 12.5 % SDS-PAGE gel of (L → R) Psy KS1, KS2, $KS3^0$ and BaeL KS5, KS5(M237A)

### *3.2.2 Substrate Specificity of BaeL KS5*

Phylogenetic sequence analysis, described in section "KS Specificity-Based Assignment of *trans*-AT PKSs", places BaeL KS5 in clade IX with other KS domains predicted to accept unbranched, α, β-unsaturated intermediates. Furthermore, KS5 is located before a key β-branching step, outlined in Scheme 3.2, and would therefore be expected to be specific for an unbranched substrate. In order to probe the specificity profile of KS5, a range of SNAC-thioesters (**20–26**) were synthesised by our collaborators (Table 3.2). Following incubation of KS5 with each SNAC derivative, reactions were quenched at regular time intervals, and desalted using the ZipTip procedure (see Sects. 2.2.3 and 2.2.11.1). Electrospray ionisation mass spectra were obtained at each time interval with the masses recorded detailed in the Appendix section. Acylation of KS5 was detected by the presence of additional peaks in the mass spectra corresponding to the mass of acyl-KS5, highlighted in Fig. 3.3, showing a time-course acylation of KS5 with butyryl-SNAC (**21**). The relative intensities of the unacylated and acylated peaks were recorded and plotted.

The *E*-2-butenoyl (**22**) and 3-methyl-2-butenoyl-SNAC (**23**) thioesters were synthesised as simplified substrate mimics of the native intermediates of KS5 and KS6, respectively (see Scheme 3.1 and Table 3.2). Using the MS-based acylation assay, KS5 was shown to be highly selective for an unbranched substrate, with no acylation observed with **23** after 40 min. In contrast the unbranched SNAC analogues **20–22** successfully acylated KS5 at similar rates, reaching saturation after ~20 min (Fig. 3.4 and Table 3.2). MS data acquired following incubation with SNAC **22** yielded additional signals corresponding to attachment of the complete SNAC to KS5. This observation can be explained by Michael addition to **22** by nucleophilic sites on the protein surface (Cys or Lys). Conjugate addition resulted in a +187 Da mass shift on the protein which was easily resolvable from that of the attachment of **22** (corresponding to a +69 Da mass shift) Fig. 3.5. The Michael addition did not appear to interfere with the ability of KS5 to be acylated by **22**.

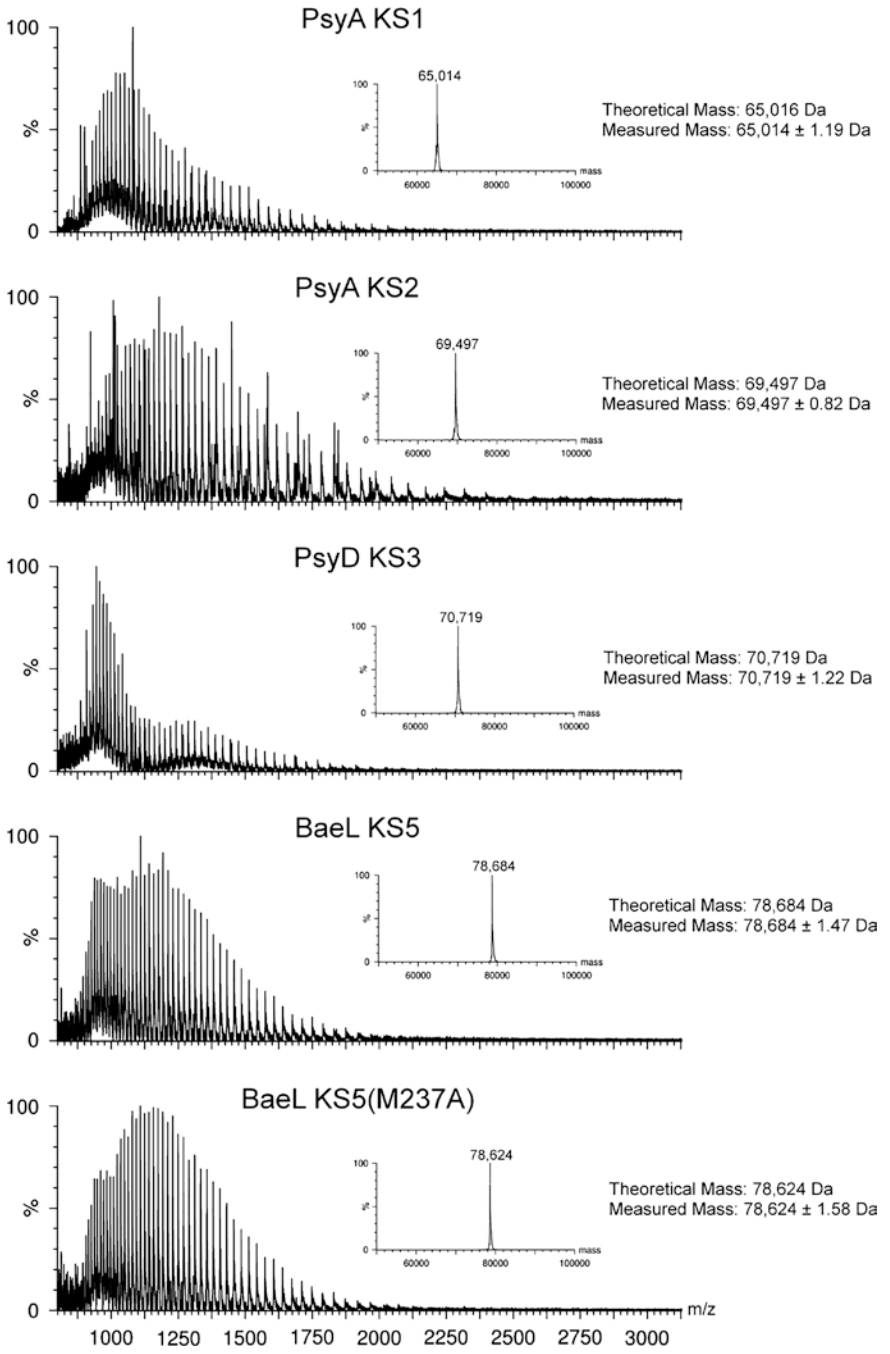



**Fig. 3.2** nESI mass spectra of (T → B) Psy KS1, KS2, $KS3^0$ and Bae KS5, KS5(M237A) sprayed from 80:20 MeCN:$H_2O$ 0.1 % TFA. Theoretical and measured molecular weights are annotated on the spectra

**Table 3.2** Initial acylation rates for BaeL KS5 with SNAC thioesters **20**–**26**

| Rate/ $\times 10^{-6}$ mol dm$^{-3}$ min$^{-1}$ | | |
|---|---|---|
| Substrate | | WT BaeL KS5 |
| **20** |  | 0.07 |
| **21** |  | 0.06 |
| **22** |  | 0.07[a] |
| **23** |  | *N.D* |
| **24** |  | 0.06 |
| **25** |  | 0.01 |
| **26** |  | 0.01 |

R = 

Initial rate was calculated from a ln[KS]/[KS$_0$] versus t plot, given to 2 significant figures

*N.D*—No acylation was detected over the tme scale of the incubation (40 min)

Estimated error in measurements: $\pm 0.005 \times 10^{-6}$ mol dm$^{-3}$ min$^{-1}$

[a]Indicates the SNAC-mimic of the predicted biosynthetic substrate for each KS domain

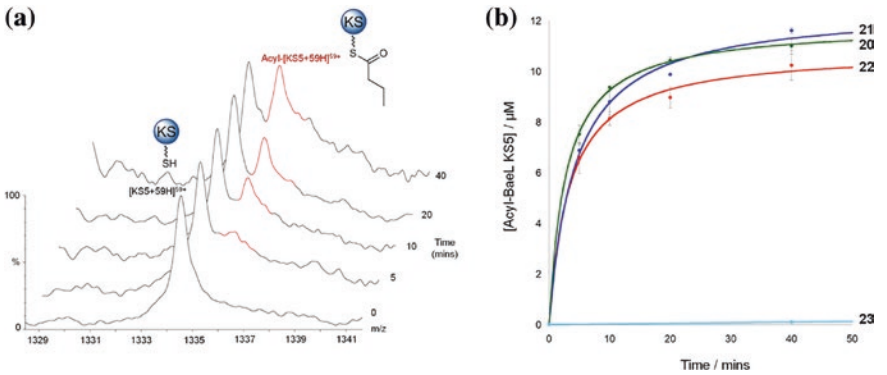



**Fig. 3.3** **a** Stacked nESI mass spectrum of the 59$^+$ charge state of BaeL KS5 showing acylation by butyryl-SNAC **21**. **b** Kinetic plot for the acylation of BaeL KS5 with SNACs **20**–**23**. Each time point was recorded in triplicate, average and error bars applied

The (*R*)/(*S*)-β-hydroxyl SNACs **25** and **26** were found to acylate KS5, but at a far reduced rate to that of the unbranched SNACs. Interestingly, the β-keto SNAC (**24**) acylated KS5 reasonably well, this may be due to the reduced steric bulk of the sp$^2$ β-keto compared to that of the sp$^3$ β-hydroxyl group. Another explanation

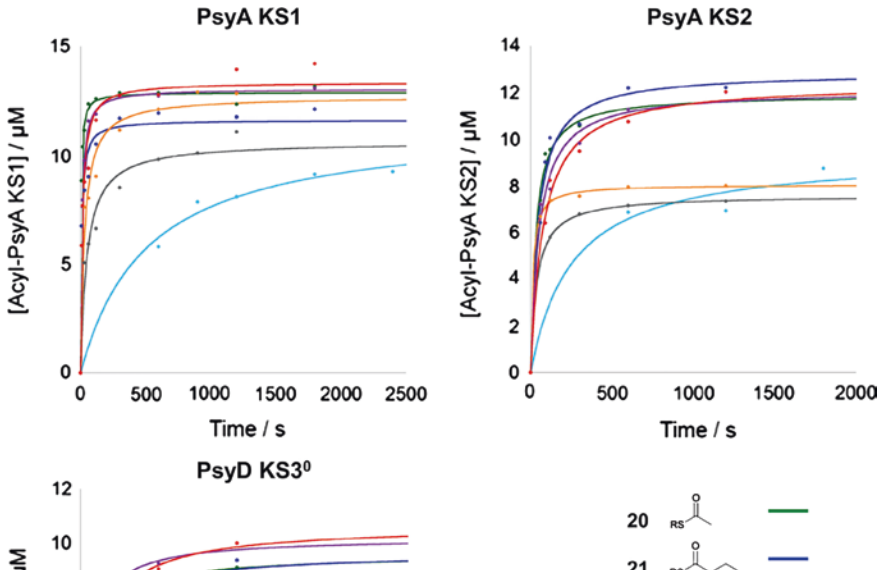



**Fig. 3.4** Deconvoluted ESI mass spectrum of Bae KS5 after incubation with SNAC derivative (**22**). Two attachments of the full SNAC to the KS were observed, highlighted by successive mass shifts of +187 Da

may be that the β-keto moiety is recruited into the binding site in a similar manner to that of the malonyl extender moiety, thus allowing it to acylate the KS domain efficiently.

This initial set of data clearly demonstrates that simple, short-chain SNAC mimics of the acyl-phosphopantetheine chain are able to probe the substrate specificity of BaeL KS5. This KS domain is situated before a key β-branching step in the biosynthesis of **18**, and is shown to from these assays to be highly specific for an unbranched substrate, providing evidence that examining the specificity of KS domains may aid biosynthetic assignment of PKSs.

### *3.2.3 Substrate Specificity Profiles: Psy KS1, KS2, and KS3⁰*

Following initial results with BaeL KS5, the substrate specificity of the first three KS domains from the psymberin PKS were investigated. Each of the KS domains are predicted to accept different substrate intermediates, shown in Scheme 3.1, and include a non-elongating KS domain, PsyD KS3$^0$.

PsyA KS1 is predicted to accept an acetyl unit following the decarboxylative loading of malonyl-CoA from the upstream GNAT domain, and due to the small nature of the predicted substrate, was expected to possess a strict substrate tolerance. However, substrate profiling of KS1 revealed an exceptionally relaxed

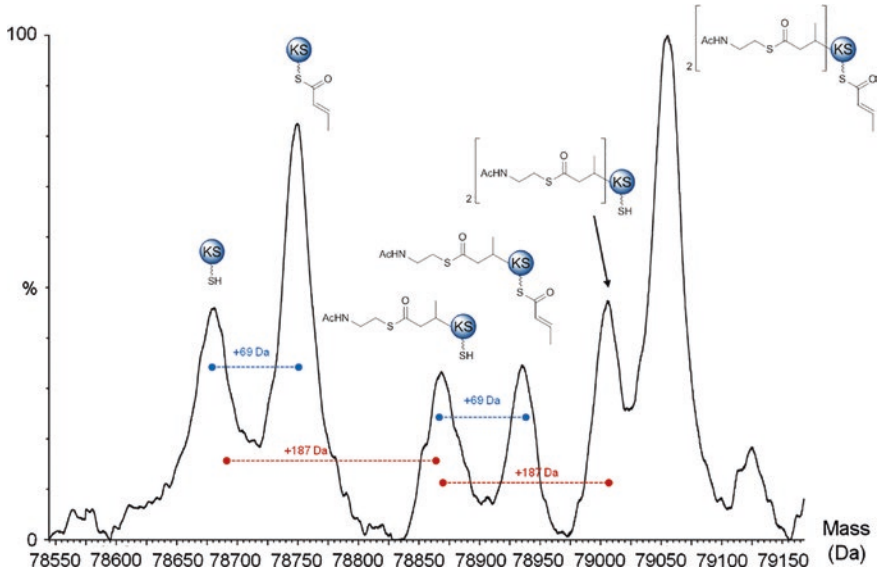



**Fig. 3.5** Kinetic acylation plots of Psy KS1, PsyA KS2 and PsyD KS3$^0$ with SNACs (**20**–**26**)

specificity profile, with all SNAC substrates able to acylate KS1 at varying rates (Table 3.3). Acetyl-SNAC (**20**) was readily incorporated by KS1, as anticipated. Butyryl- and *E*-2-butenoyl-SNACs (**21**, **22**) acylated at a comparable rate to that of **20**, suggesting that chain length in not a dictating factor. Surprisingly, the β-methyl branched SNAC (**23**) was observed to acylate KS1, albeit at a 50-fold slower rate than that of the unbranched substrates. This observation suggested key differences between the active sites of PsyA KS1 and BaeL KS5, as the latter was unable to accept the β-methyl branched SNAC. Additionally, the β-keto SNAC (**6**) acylated at the same rate to that of **21** and **22**; an observation which is conserved across all KS domains tested. Both enatiomers of the β-hydroxyl SNACs acylated, with a slight preference shown towards the (*S*) configuration.

Despite the initial hypothesis that PsyA KS1 would be a highly specific KS domain, substrate screening has shown that, functionally, KS1 accepts a wide range of substrates. It is plausible that there may be little or no evolutionary pressure upon KS1 to acquire specificity, as the preceding GNAT domain harbours AT-like specificity towards the loading unit [7]. Although a surprising result, a promiscuous substrate profile coupled with relatively high turnover could make KS1 an attractive choice for biosynthetic engineering.

PsyA KS2 is particularly interesting, as the KS is believed to accept a β-methylene branched intermediate resulting from crotonase (CR) activity in the previous module (Scheme 3.1). Interestingly, KS2 produced a similar substrate specificity profile to that of PsyA KS1, and importantly, the β-methyl branched

**Table 3.3** Initial acylation rates for PsyA KS1, PsyA KS2 and PsyD KS3$^0$ with SNAC thioesters **20–26**

| Rate/ $\times 10^{-6}$ mol dm$^{-3}$ min$^{-1}$ | | | | |
|---|---|---|---|---|
| | Substrate | PsyA KS1 | PsyA KS2 | PsyD KS3$^0$ |
| **2** |  | 3[a] | 0.3 | 0.3 |
| **3** |  | 2 | 0.3 | 0.1 |
| **4** |  | 2 | 0.3 | 0.2 |
| **5** |  | 0.04 | 0.03 | *N.D* |
| **6** |  | 2 | 0.3 | 0.2 |
| **7** |  | 0.7 | 0.3 | 0.03[a] |
| **8** |  | 0.4 | 0.02 | 0.02 |

R = 

Initial rate was calculated from a ln[KS]/[KS$_0$] versus t plot, given to 2 significant figures

*N.D*—No acylation was detected over the tme scale of the incubation (40 min)

Estimated error in measurements: $\pm 0.005 \times 10^{-6}$ mol dm$^{-3}$ min$^{-1}$

[a]Indicates the SNAC-mimic of the predicted biosynthetic substrate for each KS domain

SNAC (**23**) was observed to be a viable substrate (Table 3.3). Due to the similarities in substrate profile, it was postulated that PsyA KS1 and KS2 may have similar characteristics in the binding channel, despite being located in different phylogentic clades.

Psy KS3$^0$ is denoted as a non-elongating KS due to the presence of a glutamine in place of the critical elongating histidine residue in the active site. The predicted intermediate of KS3$^0$ is an (*S*)-β-hydroxyl acyl chain, and although not expected to elongate this substrate, transthioesterification to the active site cysteine is still believed to occur. The corresponding (*S*)-β-hydroxyl-SNAC (**25**) was found to be a viable substrate for KS3$^0$, however the β-methyl branched SNAC (**23**) was not observed to acylate. Thus, with respect to the substrate tolerance of (**23**), BaeL KS5 and PsyD KS3$^0$ are somewhat similar.

### *3.2.4 Homology Modelling of KS Domains*

In an effort to rationalise the observed specificity of the KS domains, homology models of each KS were constructed using the CPHmodel server as described in Sect. 2.2.15. Models used a previously published structure of the DEBS KS3-AT3 didomain (PDB:2QO3) as a template, and homology models generated

were overlaid upon 2QO3 to assess positioning of key residues in the active site (Fig. 3.6), in addition to Ramachandran plots (see Appendix).

Initially, models of BaeL KS5 and the downstream, β-branch accepting, BaeL KS6 were constructed. Upon examination of the respective binding sites a potentially important difference was identified. The nature of the residue immediately N-terminal to the active site cysteine (X-Cys) appeared to affect the size of the binding pocket. In the case of BaeL KS5, X was identified as Met, compared to the far less bulky Ala of BaeL KS6. To investigate the effect of this position upon substrate binding, the unbranched biosynthetic intermediate of KS5 was manually docked into the binding cleft of KS5 (Fig. 3.7). Upon docking, the X residue was found to be positioned directly adjacent to the β-position of the substrate, which in this instance was a hydrogen atom from the $sp^2$ β-carbon. It was therefore hypothesised that the nature and positioning of this residue had the potential to dictate the ability of the KS domain to accept a β-branched substrate.

The same homology modelling approach was applied to the three psymberin KS domains. PsyA KS1 and KS2 both harboured Ala residues at the X positions, yielding a binding pocket similar to that of BaeL KS6 (Fig. 3.8). Interestingly,

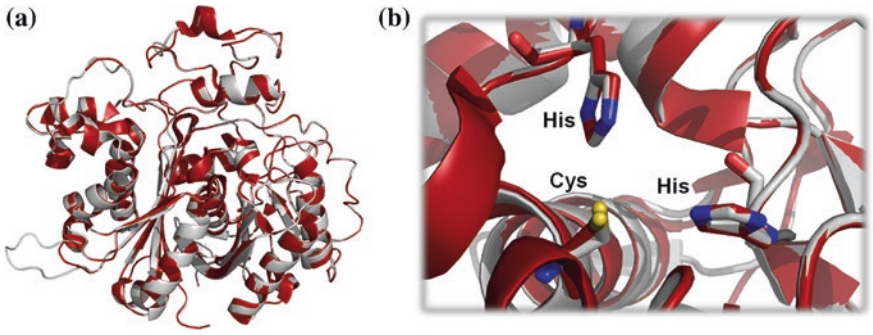



**Fig. 3.6** **a** Homology model of BaeL KS5 (*Grey*) overlaid onto DEBS KS3 X-Ray crystal structure (PDB:2QO3) showing high level of structural similarity. **b** Active site residues from BaeL KS5 and DEBS KS3 overlay. Spatial positioning of the residues is extremely similar for the model and the crystal structure

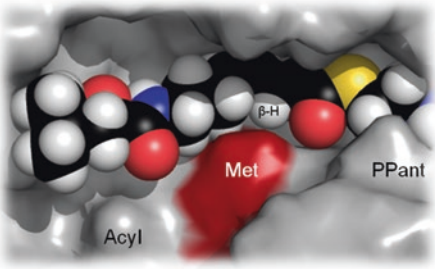



**Fig. 3.7** WT BaeL KS5 homology model with predicted unbranched biosynthetic intermediate docked in the active site. The Met residue highlighted in *red* is postulated to provide steric bulk in the binding pocket to prevent acylation by β- branched substrates

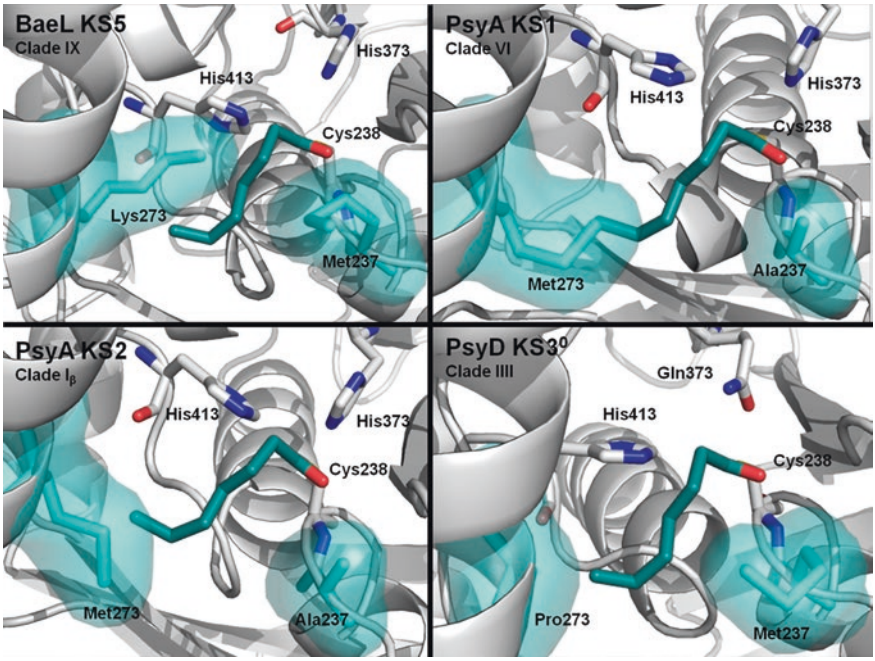



**Fig. 3.8** Comparison of the active sites for the KS domains investigated. A hexanoyl chain is modelled into the binding pocket in each case, and the steric effect of the X-Cys residue (Met/Ala237) is shown

experimentally PsyA KS1 and KS2 were also shown to tolerate the β-methyl SNAC (**23**) during the SNAC acylation assays. In contrast, PsyD $KS3^0$ and BaeL KS5, which did not accept **23**, both have a Met residue at the X-position. These observations added further evidence to the premise that a Met residue at the X-Cys position excludes bulky β-carbon branched substrates, whereas an Ala residue removes the steric restriction in the binding site.

Comparison of the KS binding sites from homology models tallies well with the experimental data for these KS domains. In order to test the hypothesis that a bulky residue at the X-Cys position precludes acylation by β-carbon branched substrates, a M237A mutant was proposed. Simple alteration to the homology model suggested that a Met to Ala mutation would indeed provide additional space in the binding pocket for a β-branched substrate (Fig. 3.9).

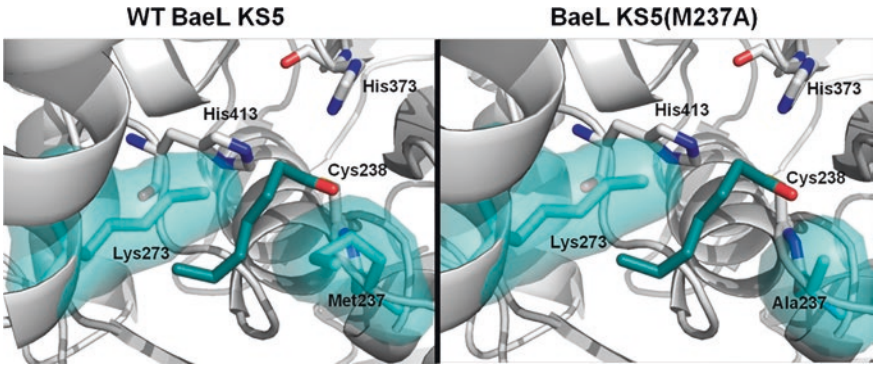



**Fig. 3.9** Comparison of WT BaeL KS5 and BaeL KS5(M237A) binding pockets. The removal of the bulky Met residue affords additional space in the binding pocket for bulkier substrates

### *3.2.5 Substrate Specificity of BaeL KS5(M237A)*

As discussed previously, the Met to Ala mutation was predicted to provide additional space in the binding pocket to accommodate a β-carbon branched substrate. The KS5(M237A) mutant was successfully constructed using standard molecular cloning techniques (detailed in Sect.\,2.2.1.4), followed by expression and purification (Figs. 3.1 and 3.2).

Incubation of KS5(M237A) with SNACs **20**-**22** resulted in acylation rates similar to those of WT KS5. This was expected, as increasing the space within the binding pocket was unlikely to affect the ability of straight chain SNACs to acylate. Reassuringly, it also indicated that the binding site mutation had not rendered the KS inactive. However, there was a notable change in the tolerance of the β-methyl branched substrate (**23**), which was found to acylate BaeL KS5(M237A) at a reasonable rate (Fig. 3.10 and Table 3.4). These observations indicate that KS domains from *trans*-AT PKSs have the ability to discriminate between substrates based upon the extent of β-branching, and furthermore, they can be engineered to accept non-natural acyl units.

### *3.2.6 Analysis of the X-Cys Position*

The results presented provide strong evidence that the nature of the X-Cys residue has the ability to dictate KS specificity with regard to β-branching. In order to investigate the potential role of this residue in KS domains from other clades, a sequence analysis was conducted using the 138 KS domains from the initial phylogenetic study, in addition to 12 subsequently identified KSs from the psymberin PKS [1, 6]. The sequence analysis focussed upon the nature of the X-Cys residue, and the most highly represented residues were reported (Table 3.5). The study

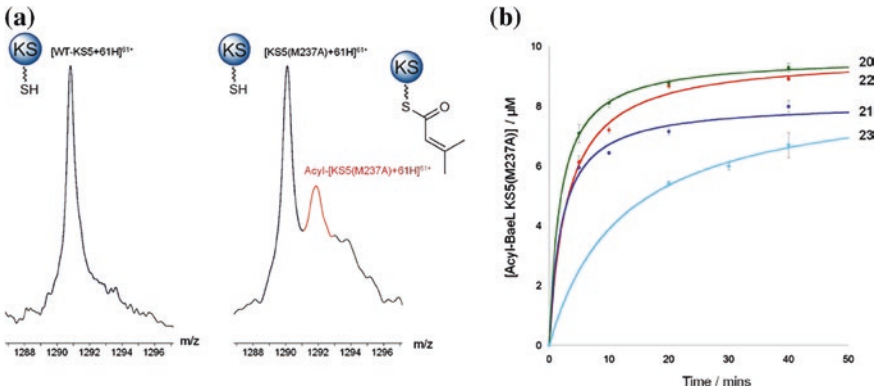



**Fig. 3.10** **a** nESI mass spectra of the $61^+$ charge state of WT BaeL KS5 and BaeL KS5(M237A) following incubation with SNAC **23** for 40 min, showing acylation with KS5(M237A) only. **b** Kinetic plot for the acylation of BaeL KS5(M237A) with SNACs **20–23**

**Table 3.4** Initial acylation rates for BaeL KS5 and Bael KS5(M237A) with SNAC thioesters **20–26**

| Rate/ $\times 10^{-6}$ mol dm$^{-3}$ min$^{-1}$ | | | |
|---|---|---|---|
| | Substrate | WT BaeL KS5 | BaeL KS5(M237A) |
| **20** |  | 0.07 | 0.06 |
| **21** |  | 0.06 | 0.05 |
| **22** |  | 0.07[a] | 0.05 |
| **23** |  | *N.D* | 0.01 |
| **24** |  | 0.06 | 0.06 |
| **25** |  | 0.01 | 0.01 |
| **26** |  | 0.01 | 0.01 |

R = 

Initial rate was calculated from a ln[KS]/[$KS_0$] versus t plot, given to 2 significant figures

*N.D*—No acylation was detected over the tme scale of the incubation (40 min)

Estimated error in measurements: $\pm 0.005 \times 10^{-6}$ mol dm$^{-3}$ min$^{-1}$

aIndicates the SNAC-mimic of the predicted biosynthetic substrate for each KS domain

**Table 3.5** Most highly represented residues located at the X-Cys position for each KS clade

| Clade | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII | XIII | XIV | XV | XVI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X-Cys | A | A | A/Q/M | A | A | A | A | M | M | A | V | A | A | A | A | N |

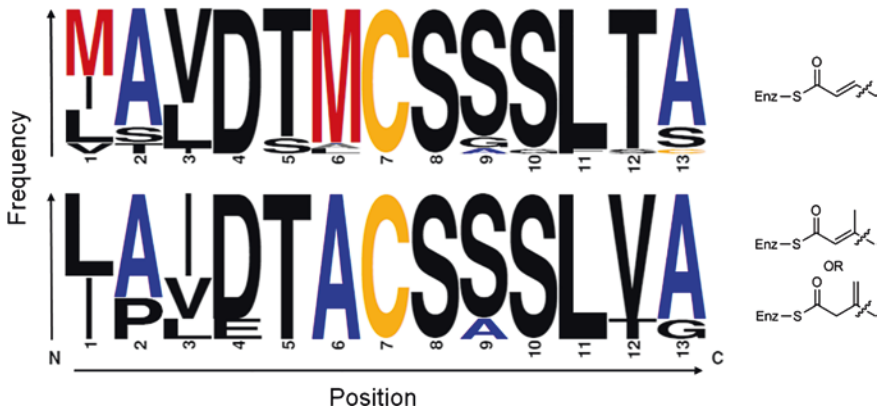



**Fig. 3.11** Sequence logo of a sequence alignment conducted upon KS domains from clades IX (*top*) and I (*bottom*), with predicted substrates from the respective clades shown. The overall height of a given amino acid indicates the sequence conservation at that position. This alignment reveals that the X-Cys position is predominantly a Met residue for KSs in clade IX, but exclusively an Ala residue for clade I. Constructed using WebLogo

revealed that for a large proportion of KS clades, X = Ala. However, some clades harboured a markedly different amino acid at this position. For example, in the clade containing BaeL KS5 (IX), 94 % of the KS domains possess a Met residue at the X position. This observation added further evidence to the hypothesis that X = Met is required to confer specificity towards unbranched substrates.

In contrast, the clade containing BaeL KS6 (I), harbours exclusively X = Ala and is likely to provide the necessary space in the binding pocket to accommodate a β-methyl branch (Fig. 3.11). Clade XVI contains all KS domains positioned immediately downstream of an NRPS module, and therefore accept amino acid-containing acyl chains. In this instance, X = Asn and the role of this residue is investigated further in Chap. 4.

As discussed in section "KS Specificity-Based Assignment of *trans*-AT PKSs", *cis*-AT and *trans*-AT PKSs are believed to have evolved via different routes. The *trans*-AT PKSs are thought to have arisen from horizontal gene transfer events, whereby substrate-specific KS domains are assembled in a mosaic fashion. In comparison, the evolutionary origin of *cis*-AT PKSs is a result of complete gene duplication events of entire modules, in which the KS domains are integrated [8, 9]. The relatively low substrate specificity observed in KS domains from *cis*-AT clusters is indicative of such gene-duplication events, as there is minimal evolutionary pressure upon the KS to confer substrate specificity. Comparing the X-Cys motif in KS domains from *cis*- and *trans*-AT clusters yields an extremely clear and interesting difference. For *trans*-AT PKSs, the nature of the X-Cys position is variable with regard to the nature of the KS substrate. For example, in order to accommodate an α- or β-methyl branch, the X-position is always a less sterically demanding substrate (Ala or Gly), or a Met residue to promote unbranched

substrates. In contrast, *cis*-AT PKSs would not be expected to have evolved such variety at the X-Cys position as a mechanism of specificity. Indeed, examination of the available sequence data for fully assigned *cis*-AT PKSs reveals that X = Ala in all instances. This observation provides additional evidence for different modes of evolution and suggests that KS domains from *trans*-AT PKSs may be more amenable to bioengineering than *cis*-AT PKSs.

## 3.3 Conclusions

The work presented in this chapter describes a novel MS-based assay to probe the substrate specificity of the acylation step catalysed by KS domains using short *N*-acetylcysteamine (SNAC) substrate mimics. The MS-methodology affords rapid, intact analysis of the KS domain, with attachment of acyl substrates to the active site Cys observable by mass shifts in the signal. The advantages to this approach over standard radioactivity-based assays include: absolute certainty regarding acyl attachment (afforded by mass measurement), ability to detect Michael addition conjugates when using α, β-unsaturated substrates, and a short analysis time reducing the propensity for hydrolysis of the acyl chain.

Using this novel approach, four KS domains from *trans*-AT PKSs have been probed for their substrate specificity at the acylation stage using a range of SNAC thioesters. BaeL KS5 was shown to be highly specific for unbranched intermediates, as predicted by the phylogenetic clades. PsyA KS1, predicted to accept an acetyl unit following decarboxylative loading of malonyl-CoA from the upstream GNAT domain, was surprisingly promiscuous in substrate specificity with the ablility to accept the β-Me SNAC (**23**). The substrate profile of the β-methylene accepting PsyA KS2 was similar to that of KS1, albeit ~10 times slower in rate. The non-elongating PsyD $KS3^0$ on the other hand had a specificity profile similar to KS5, not accepting the β-Me SNAC (**23**) over the time scale of the experiment. This suggested that BaeL KS5 and PsyD $KS3^0$ may have similar features in the active site excluding the β-Me SNAC (**23**) from the binding pocket.

Using this data and subsequent homology modelling, the molecular basis dictating KS β-branch specificity was revealed. A single residue in the binding pocket of BaeL KS5 was identified as being responsible for directing the tolerance for β-branched substrates. Upon examination of the homology models, BaeL KS5 and PsyD $KS3^0$ the residue at the X-position is a bulky Met residue, precluding acylation by bulky β-Me. A single Met → Ala point mutation in KS5 successfully removed the steric bulk in the binding site and allowed acylation with the β-Me SNAC (**23**).

In summary, using a novel MS-based assay the molecular basis of carbon β-branch specificity in *trans*-AT PKSs KS domains has been identified. A key residue has been characterised in the active site of BaeL KS5 which sterically dictates the tolerance for β-branched substrates. Examination of this residue in the sequences of 150 KS domains from *trans*-AT PKSs suggests that the presence of

Gly or Ala residues are required for acceptance of a carbon β-branched substrate. This general rule provides insight for potential *trans*-AT PKSs engineering, and may add another discriminatory factor to the functional assignment of natural product biosynthesis. These results also indicate the exciting potential to tune the specificity of KS domains towards desired biosynthetic products.

## References

1. T. Nguyen et al., Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nat. Biotechnol. **26**, 225–233 (2008)
2. J.Q. Wu, K. Kinoshita, C. Khosla, D.E. Cane, 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**, 16301–16310 (2004)
3. N.A. Schnarr, A.Y. Chen, D.E. Cane, C. Khosla, Analysis of covalently bound polyketide intermediates on 6-deoxyerythronolide B synthase by tandem proteolysis-mass spectrometry. Biochemistry **44**, 11836–11842 (2005)
4. T.W. Geders et al., Crystal structure of the ECH2 catalytic domain of CurF from Lyngbya majuscula - insights into a decarboxylase involved in polyketide chain beta-branching. J. Biol. Chem. **282**, 35954–35963 (2007)
5. C.T. Calderone, W.E. Kowtoniuk, N.L. Kelleher, C.T. Walsh, P.C. Dorrestein, Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis. Proc. Natl. Acad. Sci. U.S.A. **103**, 8977–8982 (2006)
6. K.M. Fisch et al., Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. Nat. Chem. Biol. **5**, 494–501 (2009)
7. L. Gu et al., GNAT-like strategy for polyketide chain initiation. Science **318**, 970–974 (2007)
8. T. Hochmuth, J. Piel, Polyketide synthases of bacterial symbionts in sponges - evolution-based applications in natural products research. Phytochemistry **70**, 1841–1849 (2009)
9. H. Jenke-Kodama, A. Sandmann, R. Muller, E. Dittmann, Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evol. **22**, 2027–2039 (2005)

# Chapter 4
# Substrate Specificity of Ketosynthase Domains Part II: Amino Acid-Containing Acyl Chains

## 4.1 Introduction

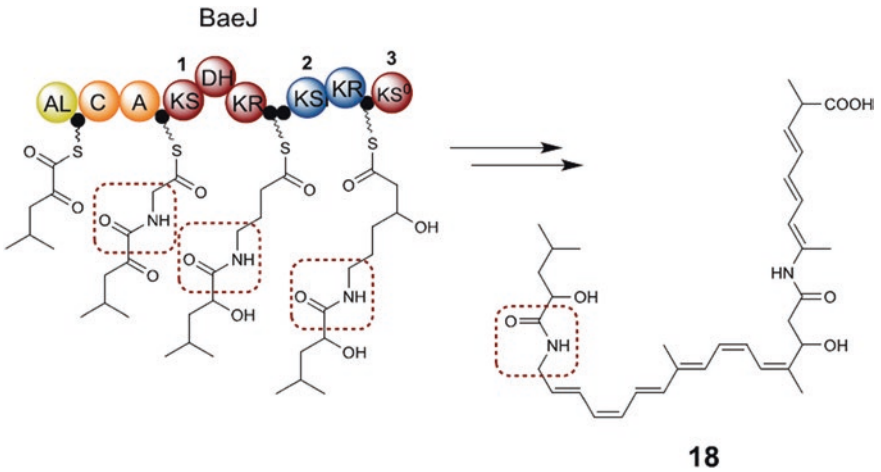
The *trans*-AT PKSs are notorious for incorporation of non-canonical modules, which often contain novel enzymatic domains [1]. Although this provides a wealth of exciting enzymology to be explored, this factor also makes assignment of biosynthetic clusters somewhat difficult. Similar to *cis*-AT PKSs, the *trans*-AT systems also integrate NRPS modules into the biosynthetic assembly line (PKS-NRPS hybrids), which allows the introduction of amino acid units into the growing polyketide chain (see Sect. 1.2.2.5) [2, 3]. Therefore, this feature requires KS domains from downstream modules to accept NRPS-derived intermediates. Although in *cis*-AT systems some insights exist into the association between NRPS and PKS modules, very little is known for either type with regard to substrate requirements for a successful assembly line switch [4, 5].

In this chapter, the first study of a KS domain immediately downstream of a NRPS module is reported. Using full-length acyl precursors the substrate specificity of BaeJ KS1 from the bacillaene *trans*-AT PKS was examined. BaeJ KS1 is the first PKS module in the biosynthesis of bacillaene (**18**), and is believed to accept a glycine-derived intermediate incorporated by the previous NRPS module (Scheme 4.1). KS1 is positioned in a phylogenetic clade with other amino-acid accepting KS domains, of which the majority accept a glycine intermediate.

Using the MS-based methodology from Chap. 3, BaeJ KS1 is shown to possess high specificity towards 2-amidoacetyl intermediates, which are derived from incorporation of alpha amino acids into the polyketide chain. Using a range of full-length *N*-acetylcysteamine (SNAC) analogues as substrates the specificity of the domain is successfully probed. The natural glycine-derived acyl-SNAC was found to acylate KS1 with highest efficiency, with an alanine variant also incorporated, but its valine equivalent was not, which indicated limited tolerance of substitution at the α-position.



M. Jenner, *Using Mass Spectrometry for Biochemical Studies on Enzymatic Domains from Polyketide Synthases*, Springer Theses,
DOI 10.1007/978-3-319-32723-5_4




**Scheme 4.1** Partial proposed biosynthetic scheme for bacillaene (**18**). The amide region of the relevant intermediates and product is highlighted in *red*

Substrate analogues without an amine or amide nitrogen substituted on the 2-position were not accepted by KS1 at the standard assay concentration of 0.5 mM. Moreover, removal of Asn-206 from the active site of KS1 by site-directed mutagenesis reduced $k_{cat}/K_m$ by a factor of ~2. This residue is conserved in most known 2-amidoacetyl-accepting KS domains from *trans*-AT PKSs and we postulate an important interaction between Asn-206 and the amide nitrogen of the substrate.

## 4.2 Results and Discussion

### *4.2.1 Purification of Ketosynthase Domains*

BaeJ KS1 and respective mutants used in this chapter were expressed and purified as His-Tag fusion proteins, as described in Sect. 2.2.1. Post-purification, KS domains were analysed by SDS-PAGE and by mass spectrometry. An SDS-PAGE gel of all the KS domains used in this chapter is shown in Fig. 4.1. The subsequent mass spectra are shown in Fig. 4.2.

### *4.2.2 Substrate Specificity of BaeJ KS1*

The initial aims of this study were to test the full-length glycine-containing substrate with BaeJ KS1 and examine the tolerance towards other amino acids. The glycine, alanine and valine derived α-keto acid intermediates **27–29** were

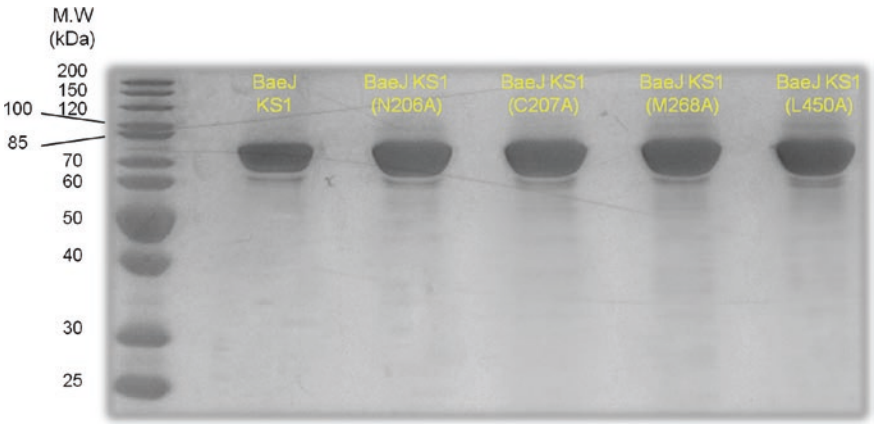



**Fig. 4.1** 12 % SDS-PAGE gel of WT BaeJ KS1 and KS1(N206A), KS1(M268A) and KS1(L450A) post His-tag purification

incubated with BaeJ KS1 according to the protocol in Sect. 2.2.3, at a final concentration of 0.5 mM. Electrospray mass spectra were recorded to monitor the acylation reaction as a function of time. The unacylated mass of WT BaeJ KS1 was determined to be 73,507 Da, in good agreement with the theoretical value of 73,509 Da (Fig. 4.2). Successful acylation was monitored by the appearance of additional peaks corresponding to acylated KS1 (Fig. 4.3).

Kinetic analysis of KS1 with substrates **27–29** revealed that the glycine-derived SNAC **27** was a slightly better substrate than the alanine variant **28** (Fig. 4.3 and Table 4.1). This suggests that the presence of an α-methyl branch may reduce the rate of acylation on steric grounds. No acylation was detected for the valine-derived SNAC **29**, adding further evidence to the theory that excessive steric bulk around the α-position may preclude entry into the binding pocket, or in some cases, prevent nucleophilic attack at the carbonyl.

In order to further interrogate the structural features that are important for substrate selectivity and recognition by KS1, an additional set of SNAC-thioesters were synthesised by our collaborators, each designed to possess partial functionality of the natural acyl substrate (**27**). SNACs **30** and **31** were used to dissect the individual effect of the nitrogen and the carbonyl of the amide. SNAC **32** was used to probe the effect of moving the amide to the 4-position (corresponding to a γ-amino acid derived chain), and SNAC **33** was used to examine the effect of any potential long-range interactions associated with the carbonyl group in the 2-keto-4-methylpentanoyl moiety of **27**.

Incubation of KS1 with the 2-amino SNAC (**30**) and the 4-keto SNAC (**31**) provided extremely valuable information. Interestingly, SNAC **30** was able to acylate KS1 efficiently at a rate faster than the native substrate (**27**). In contrast, upon incubation of the 4-keto SNAC (**31**) with KS1, no acylation was observed over the time-frame of the experiment. These observations indicate that the nitrogen of the amide functionality in **27** is a crucial component for substrate recognition.

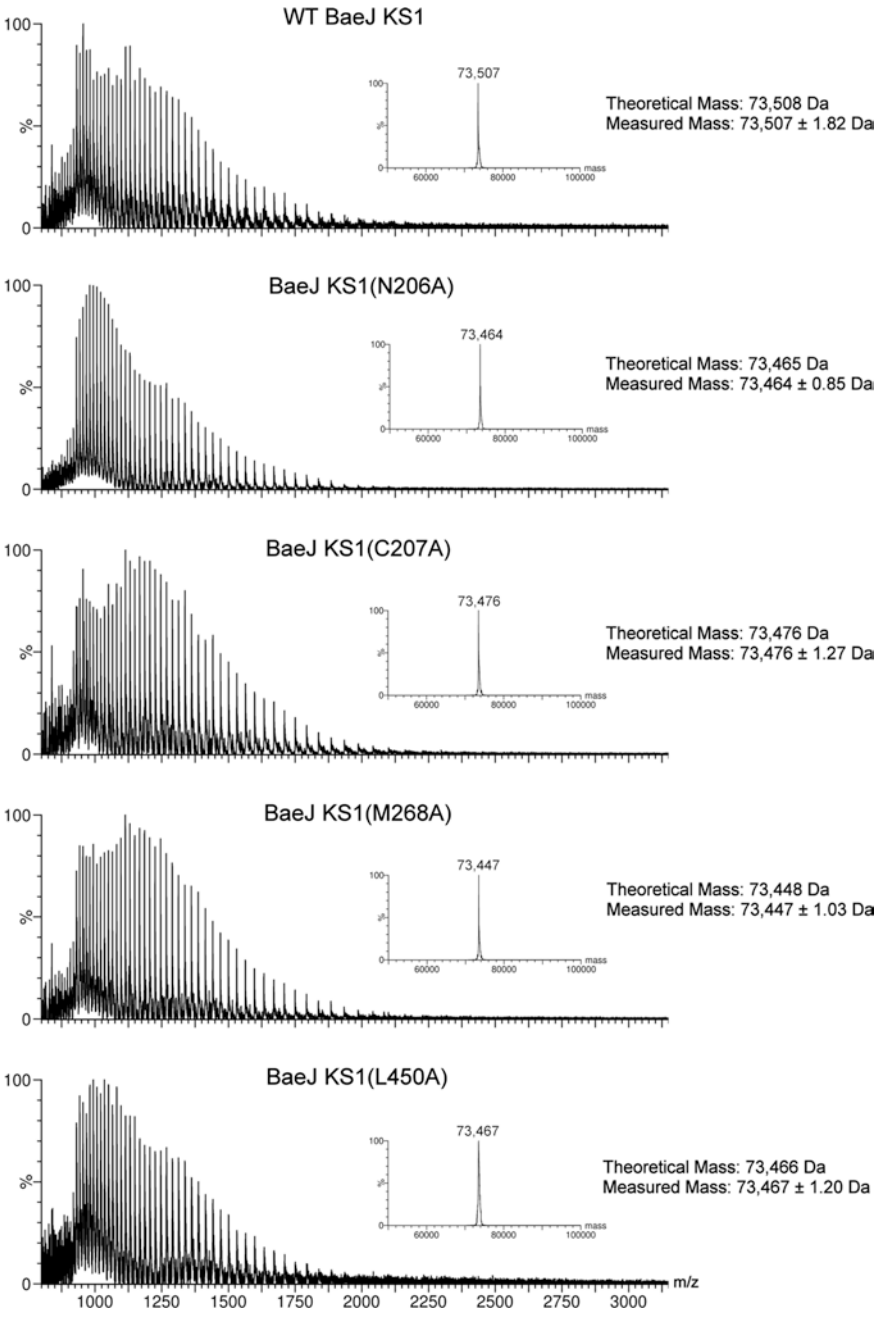



**Fig. 4.2** nESI mass spectra of BaeJ KS1 and mutants sprayed from 80:20 MeCN:$H_2O$ 0.1 % TFA. Deconvoluted spectra are displayed as an *inset*. Measured and theoretical masses are noted on each spectrum

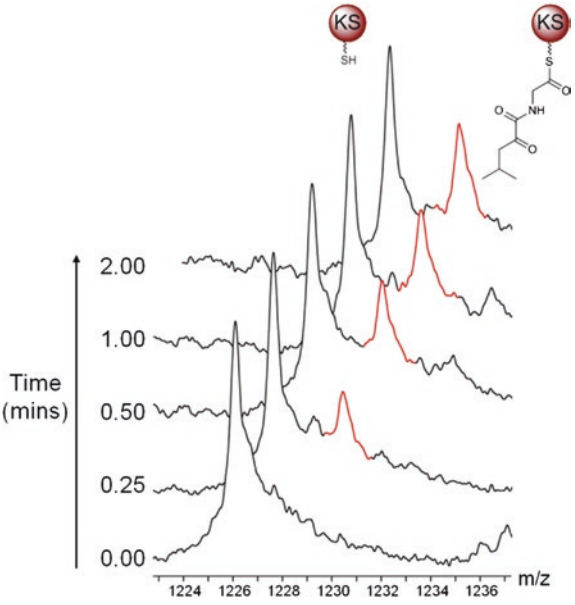


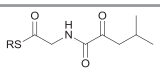
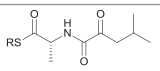
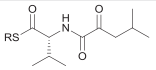
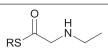
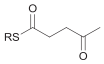
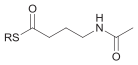
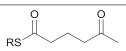
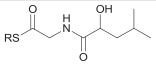
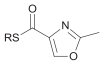
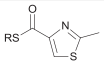

**Fig. 4.3** Stacked ESI mass spectra of the $60^{+}$ charge state of WT BaeJ KS1 during kinetic acylation. Peaks coloured in *red* correspond to +170 Da mass shift on the KS1 domain, indicating acylation by (**27**)

It is worth considering that the amido nitrogen of **27** and the amino nitrogen of **30** will possess considerably different $pK_a$ values, however, both substrates would be expected to competent hydrogen bond donors.

Incubation of SNAC **32** yielded no observable acylation of KS1, suggesting that the amide substrate should be derived from an α-amino acid. Equally, no acylation was detected with SNAC **33** indicating that the distal carbonyl is not a factor in substrate viability. Previous work by Claderone et al. revealed that the reduction of the 2-keto-4-methylpentanoyl intermediate to the 2-hydroxy-analogue is performed by the KR domain of module 3 (Scheme 4.1) [6]. This unorthodox reduction step is believed to occur post-KS1 elongation, therefore suggesting that the oxidation level of SNAC **27** is correct, and therefore a substrate mimic for KS1. To probe this issue further, a 2-hydroxy-4-methyl-pentanoyl SNAC (**34**) was synthesised. Upon incubation of SNAC **34** with KS1, a slight reduction in acylation rate was observed (~20 %) compared to SNAC **27**. The reduction in acylation rate may be due to unfavourable loss of planarity in the substrate, or decreased intrinsic reactivity of the thioester (Fig. 4.4).

These observations suggest that the substrate selectivity of BaeJ KS1 may not extend past the first 3–4 carbons of the substrate, which at least in this instance, is in good agreement with the prediction of *trans*-AT KS specificity. It is noteworthy that, SNAC **27** was capable of acylating BaeL KS5, despite its presence in a different phylogenetic clade to BaeJ KS1. BaeL KS5 is situated in clade IX, predicted to accept unbranched and unsaturated substrates. The previous work from Chap. 3 revealed that KS5 does not accept β-branched acyl chains, but unbranched chains acylate the domain efficiently. Therefore, on this basis SNAC

**Table 4.1** Estimated initial acylation rates for WT KS1 and KS1(N206A) with SNACs **27**–**36**

| Rate/ × $10^{-6}$ mol $dm^{-3}$ $min^{-1}$ | | | |
|---|---|---|---|
| | Substrate | WT BaeJ KS1 | BaeJ KS1 (N206A) |
| **27** |  | 0.73[a] | 0.26 |
| **28** |  | 0.31 | 0.14 |
| **29** |  | *N.D* | *N.D* |
| **30** |  | 0.90 | 0.39 |
| **31** |  | *N.D* | *N.D* |
| **32** |  | *N.D* | *N.D* |
| **33** |  | *N.D* | *N.D* |
| **34** |  | 0.60 | 0.14 |
| **35** |  | 0.06 | 0.06 |
| **36** |  | 0.06 | 0.04 |

R = 

Initial rate was calculated from a ln[KS]/[$KS_0$] versus t plot, given to 2 significant figures

*N.D*—No acylation was detected over the tme scale of the incubation (10 min)

Estimated error in measurements: ±0.005 × $10^{-6}$ mol $dm^{-3}$ $min^{-1}$

[a]Indicates the SNAC-mimic of the predicted biosynthetic substrate for each KS domain

**27** might be expected to be a reasonable substrate for KS5 due to its unbranched nature at the β-position. Although no favourable interactions were observed past the β-position with KS1, this result does not rule out the potential for other interactions in the binding pocket which help to stabilise the intermediate once bound.

NRPS modules that incorporate amino acid building blocks can also contain heterocyclisation (HC) and oxidation (OX) domains, which together are capable of catalysing formation of thiazole and oxazole moieties from *N*-acetylcysteine and *N*-acetylserine respectively [7, 8]. In an effort to test whether these cyclic, non-amide intermediates are viable substrates of KS1, 2-methyloxazole (**35**) and 2-methylthiazole (**36**) SNACs were synthesised. Despite the additional bulk of a five-membered ring; a nitrogen atom, albeit tertiary in nature, is present at the β-position in these substrates. The $sp^2$ nitrogen of **35** and **36** will be hydrogen

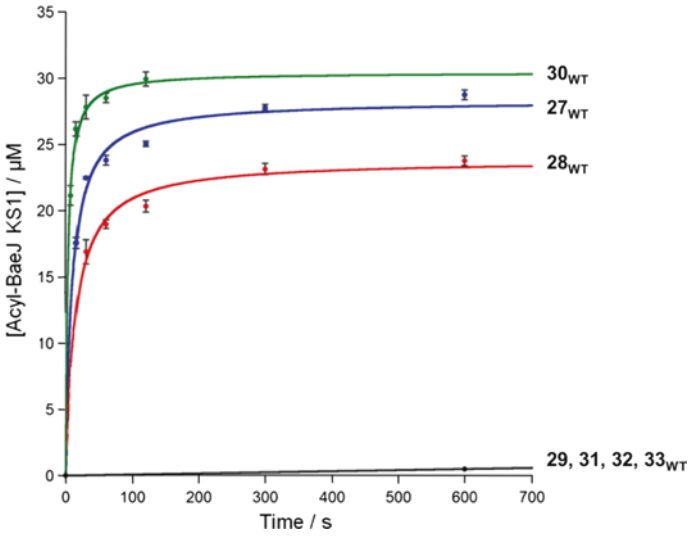



**Fig. 4.4** Kinetic plot of WT KS1 with SNAC thioester substrate mimics (**27–33**)

bond acceptors, whereas SNACs **27**–**30** possess hydrogen bond donors. Evidence to this point would suggest that a hydrogen bond donating nitrogen is favourable in the substrate. Incubation of SNACs **35** and **36** with KS1 resulted in low levels of acylation at a far reduced rate to that of straight chain amido/amino-SNACs (see Table 4.1). This is likely due to the increased steric effect incurred by the ring, and the lack of favourable interactions formed with the tertiary amine.

### *4.2.3 Analysis of X-Cys Position*

Previous work from Chap. 3 highlighted the importance of the amino acid residue 'X' immediately preceding the active site Cys with regard to substrate specificity. It therefore seemed logical to investigate the nature of this position in amino acid accepting KS domains. A sequence alignment of all 17 known KS domains immediately downstream of a NRPS module was performed, and revealed that in 11 instances (including BaeJ KS1) an Asn residue occupied the X position (Fig. 4.5). This suggested that the Asn residue may well be of biochemical significance, and could be involved in substrate recognition due to its apparent sequence conservation. KS domains predicted to accept oxazole and thiazole intermediates were included in the alignment; however none of these KSs harboured an Asn in the X-position. Instead, smaller Gly or Ser residues were favoured, potentially providing additional space in the binding pocket for the heterocycle.

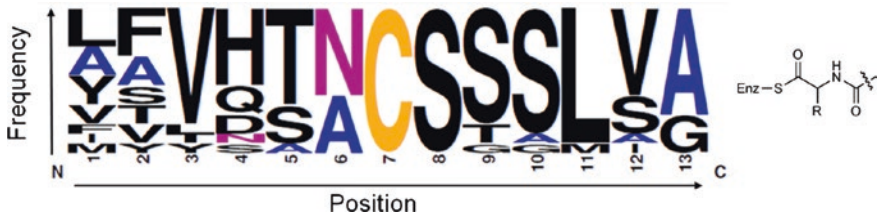



**Fig. 4.5** Sequence logo of a sequence alignment conducted upon amino acid-accepting KS domains. The overall height of a given amino acid indicates the sequence conservation at that position. This alignment reveals that the X-Cys position is occupied by either an Asn or Ala residue. Constructed using WebLogo

A phylogenetic tree was constructed from the sequence alignment, which grouped BaeJ KS1 with other KS domains with X = Asn. In addition, a separate clade was formed for amino acid-accepting KSs where X = Ala, suggesting that the Asn residue is not absolutely required for a functional enzyme (Fig. 4.6). Interestingly, no other known KS domain from *cis* or *trans*-AT PKSs possesses an Asn residue at the X position, other than the KSs in clade XVI. It is therefore reasonable to postulate that Asn may form important interactions with 2-amidoacetyl substrates, potentially assisting the acylation reaction.

### *4.2.4 Homology Modelling of BaeJ KS1*

In an effort to rationalise the substrate selectivity exhibited by BaeJ KS1 and examine the effect of an Asn residue at the X position, a homology model of KS1 was constructed using the same methodology as in Chap. 3. Due to the shorter nature of the predicted intermediate, it was possible to model the native acyl chain of **27** into the binding pocket. Using a previously published structure of a FA:FAS complex (PDB:1EK4) [9] as a guide, the acyl chain of **27** was successfully modelled into the active site of KS1.

Using this approach, two models of acyl-KS1 were produced to visualise the role of the Asn residue in the binding pocket. The first of which shows a potential hydrogen bond between the amine of the Asn and the δ-carbonyl of the acyl chain (Fig. 4.7a). However, the SNAC acylation assays revealed no experimental evidence to suggest that the δ-carbonyl of **27** plays a role in substrate recognition. In the second structure however, the carboxamide region of Asn206 is rotated ~90° anticlockwise which allows the formation of a hydrogen bond between amide group of the acyl chain and the carboxyl of Asn206 (Fig. 4.7b). This conformation of the Asn is in good agreement with the experimental data obtained using the acyl SNACs.

A further finding from the homology modelling was the putative existence of a pocket in the KS binding cleft which was able to accommodate the terminus of

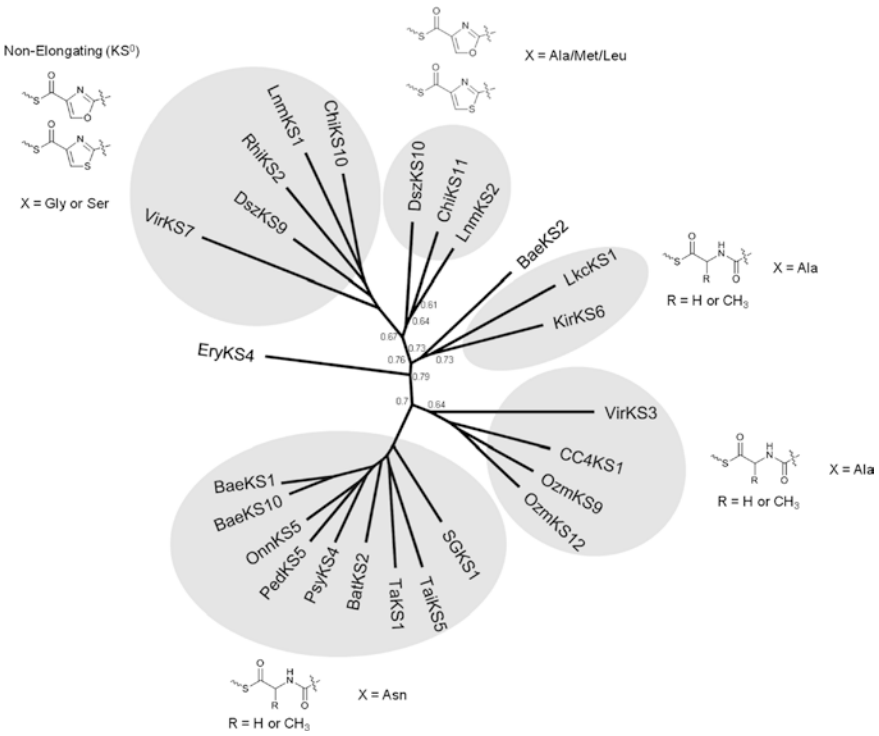



**Fig. 4.6** Phylogram of KS domains predicted to accept amino-acid derived substrates. Tip labels consist of KS number and cluster, with an outgroup of the erythromycin KS4 domain. X indicates the identity of the amino acid preceeding the active site Cys residue. The phylogram shows a distinct clade for KS domains which harbor an Asn residue at the X-Cys position. However, amino acid-accepting KSs where X = Ala fall into separate clades. N.B. BaeKS1 and BaeKS10 from B. Amyloliquefaciens FZB42 are orthologous to PksKS1 and PksKS11, respectively, from the bacillaene PKS of Bacillus subtilis 168. Bootstrap values ≥0.60 are highlighted. *Abbreviations* Bat, batumin; Bae, bacillaene; CC4, uncharacterized PKSs from Clostridium cellulolyticum; Chi, chivosazol; Dsz, disorazol; Ery, erythromycin; Kir, kirromycin; Lkc, lankacidin; Lnm, leinamycin; Onn, onnamide; Ozm, oxazolomycin; Ped, pederin; Psy, psymberin; Rhi, rhizoxin; SG, uncharacterized PKSs from Streptomyces griseus; Ta, myxovirescin; Tai, thailandamide; Vir, virginiamycin

the acyl chain (Fig. 4.7c). Although no evidence for longer range interactions was revealed in this study, this model may indicate the benefit of using longer chain substrate mimics in KS enzyme assays.

The homology modelling approach also enabled identification of residues potentially blocking the KS active site from acylation by α-branched SNACs, such as **29**. Met268 and Leu450 appeared to be providing steric bulk around the α-position of the substrate (Fig. 4.8). A general hypothesis was that removing the steric bulk from Leu450 might allow a bulkier L-configured amino acid-derived acyl chain to enter the active site, whereas removal of Met268 may favour bulky D-configured amino acid.

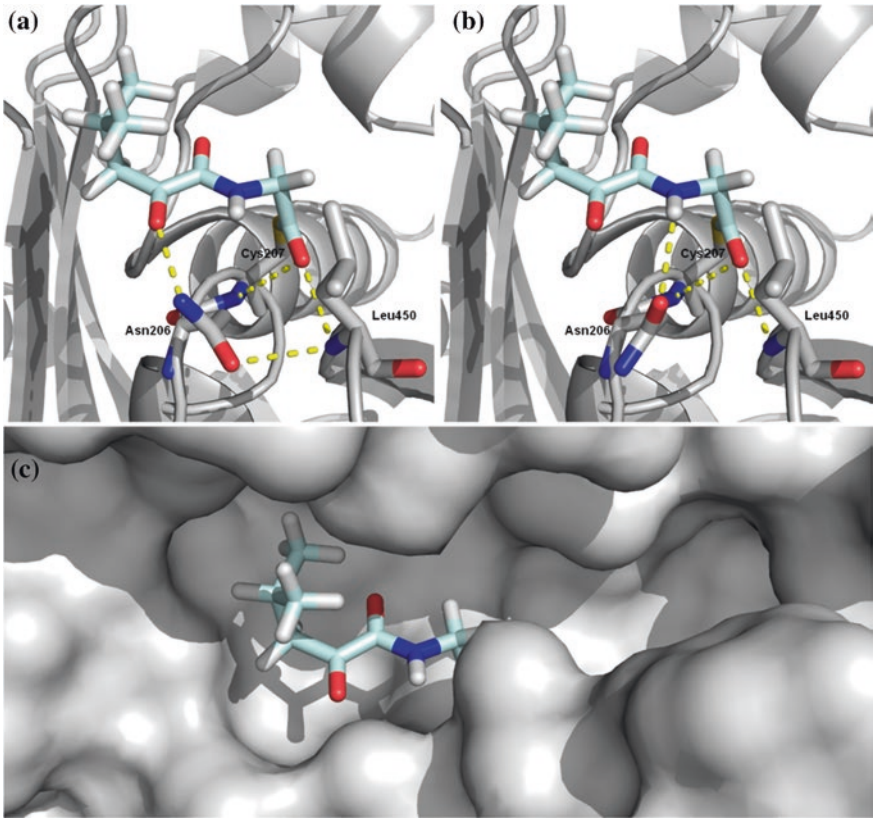



**Fig. 4.7** Models of SNAC thioester **27** bound to BaeJ KS1. **a** BaeJ KS1 with Asn206 able to form a hydrogen bond to the δ-carbonyl of SNAC **2** and to the backbone amide nitrogen of Leu450. **b** In contrast, BaeJ KS1 with the carboxamide of the Asn206 residue rotated ~90° anti-clockwise, which is now able to form a hydrogen bond with the β-amide nitrogen of SNAC **2**. In both cases the thioester carboxyl is stabilized by an oxyanion hole formed by the backbone amide of Leu450, in addition to a potential stabilizing interaction with the backbone amide of Cys207. *Dashed yellow lines* designate distances between 2.4–3.3 Å. **c** Surface representation of BaeJ KS1 binding pocket with SNAC **2**. The distal region of the binding cleft provides space to accommodate the rest of the acyl chain

### *4.2.5 Substrate Specificity of BaeJ KS1(N206A)*

To test the influence of the Asn residue upon KS1 acylation rate, a BaeJ KS1(N206A) mutant was constructed. Upon initial incubation of KS1(N206A) with SNAC **27**, markedly reduced levels of acylation were observed compared to WT KS1, with the initial rate reduced by ~65 % (Table 4.1 and Fig. 4.9). Approximately the same reduction in acylation rate was observed for the other nitrogen-containing SNACs **28**, **30** and **34**, indicating the mutation may well have removed a favourable interaction between the Asn side-chain and the nitrogen of

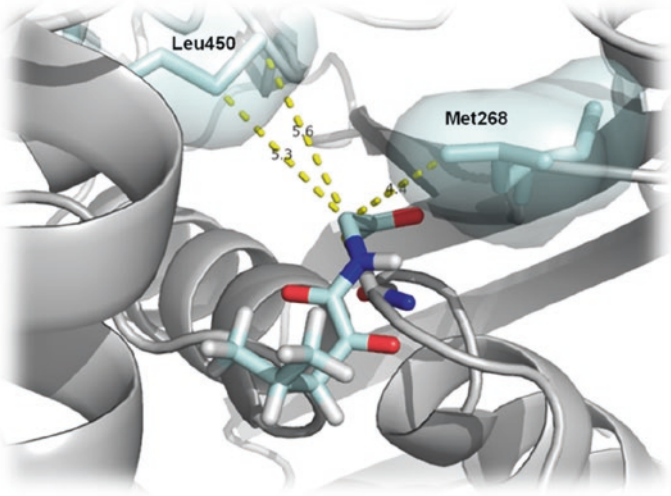



**Fig. 4.8** Model of SNAC **27** covalently bound to active site Cys of WT BaeJ KS1. Residues Met268 and Leu450 are shown in teal, and highlight the steric hindrance they apply to the α-position of the substrate

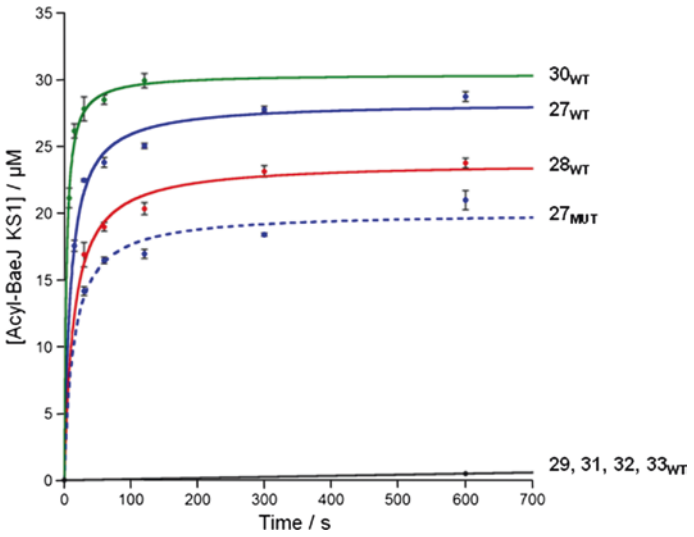



**Fig. 4.9** Kinetic plot of WT KS1 with SNAC thioester substrate mimics (**27–33**) and KS1(N206A) with SNAC **27** for comparison

the substrate, as previously hypothesised. Similar to WT KS1, no acylation was observed for the non-nitrogenous SNACs **31**, **32** and **33** indicating that the Ala residue is not promoting acylation of these substrates.

The replacement of Asn with an Ala residue potentially provided additional space in the binding pocket for larger substrates, such as **29**, **35** and **36**. However, in all cases, no dramatic change from the WT acylation rates was observed for the mutant. These findings, combined with the previous substrate-based analysis suggest that the presence of an Asn residue at the X position promotes the incorporation of 2-amidoacetyl substrates into this KS domain. Although it is tempting to speculate that such an effect may exist for all members of clade XVI, however the existence of KSs in this clade with X = Ala suggests that an Asn residue is desirable but not essential for function.

### *4.2.6 Michaelis–Menten Treatment of WT KS1 and KS1(N206A)*

In order to gain more insight into the interaction between the active site Asn206 and SNAC **27**, a Michaelis–Menten treatment was performed on WT KS1 and KS1(N206A). This analysis required incubation of KS domains with a range of SNAC **27** concentrations (0.250–2 mM), with the measured acylation velocities used to construct Lineweaver–Burk plots (Fig. 4.10).

When using higher concentrations of SNAC (1–2 mM), a small amount of secondary, non-specific, acylation was observed. A KS1 mutant lacking the active site Cys (C206A) was therefore constructed to monitor the degree of non-specific acylation occurring during these measurements. At a SNAC concentration of 1 mM, the non-specific acylation was measured to be 5, and 8.5 % at 2 mM. These values were subtracted from the percentage acylation values measured for WT and N206A mutant of KS1. It is likely that the non-specific acylation observed is occurring at one of the 3 additional Cys residues in the KS1 sequence, presumably positioned such that it is surface exposed.

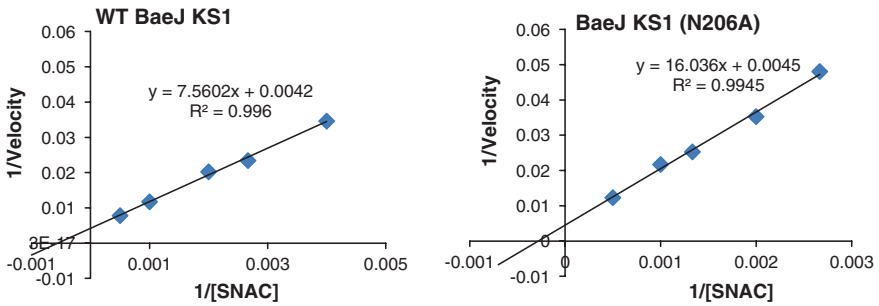



**Fig. 4.10** Michaelis–Menten analysis of WT KS1 and KS1(N206A). Lineweaver–Burk plots for BaeJ KS1 WT and N206A are shown. Kinetic parameters were calculated from the plot using the equations shown, and errors were determined using a σ confidence interval

**Table 4.2** Calculated Michaelis–Menten kinetic parameters for WT KS1 and KS1(N206A) with SNAC **27**

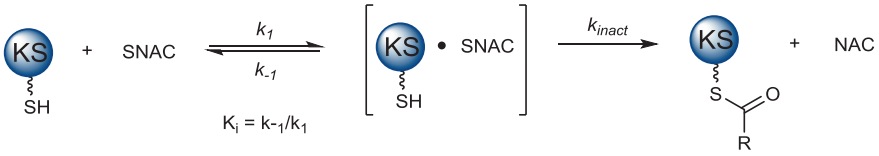
| | $K_m$ (mM) | $k_{cat}$ ($min^{-1}$) | $V_{max}$ (mM $min^{-1}$) |
|---|---|---|---|
| WT KS1 | 1.8 ± 0.4 | 3.0 ± 0.6 | 0.24 ± 0.05 |
| KS1(N206A) | 3.7 ± 1.3 | 2.8 ± 0.9 | 0.22 ± 0.07 |

The Michaelis–Menten analysis yielded a $K_m$ value of 1.8 mM ± 0.4 for WT KS1. In comparison, the $K_m$ value for KS1(N206A) was 3.7 mM ± 1.3, suggesting a decreased affinity towards SNAC **27** upon removal of the active site Asn residue. Additionally, the $k_{cat}/K_m$ value was reduced by a factor of 2, implying that the catalytic efficiency of KS1(N206A) has also been reduced by removal of the Asn residue. It is noteworthy that this reduction in $k_{cat}/K_m$ is largely due to the difference in $K_m$ values, as $k_{cat}$ values for WT and mutant are extremely similar (Table 4.2).

One of the key assumptions in Michaelis–Menten kinetics is that the concentration of the enzyme-substrate complex remains constant in the steady-state, even if the concentration of substrate and product are changing. However, the irreversible KS-SNAC acylation reaction monitored in Chaps. 3 and 4 results in dramatic depletion of the active enzyme concentration during the course of the reaction. Therefore, the concentration of the enzyme-substrate complex can no longer be assumed as constant throughout the reaction. Although numerous examples in the literature apply Michaelis–Menten kinetics to analogous systems, and are in good agreement with the values reported in this chapter [10, 11], its application to such systems is debateable. Therefore, fitting the KS-SNAC system to an irreversible kinetic equation is likely to be a more accurate approximation (Scheme 4.2 and Eq. 4.1).

$$k_{obs} = k_{inact}\frac{[SNAC]_0}{[SNAC]_0 + K_i} \tag{4.1}$$

where $k_{obs}$ is the initial rate of reaction for a given concentration of SNAC and $k_{inact}$ and $K_i$ are defined in Scheme 4.2. Fitting the $k_{obs}$ data to Eq. 4.1 for both WT BaeJ KS1 and BaeJ KS1(N206A) yielded plots shown in Fig. 4.11, and generated the kinetic values of $k_{inact}$ and $K_i$ shown in Table 4.3.




**Scheme 4.2** Kinetic profile of acylation of a KS domain with a SNAC thioester. The rate constants for each step are highlighted and defined

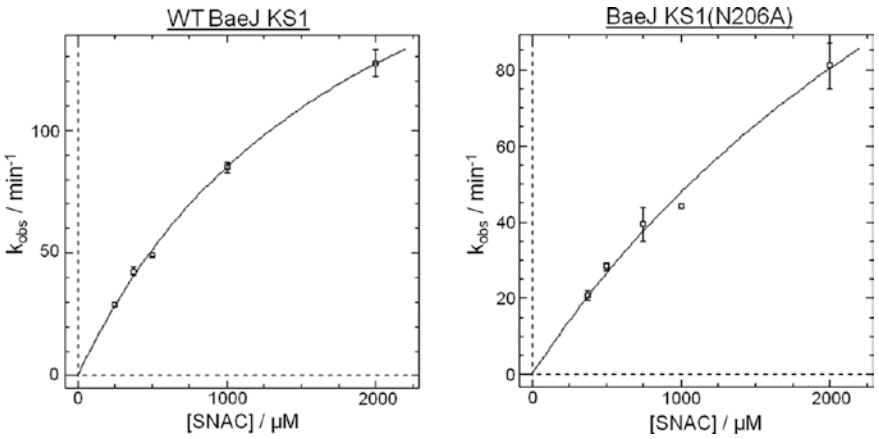



**Fig. 4.11** Non-linear regression analysis of BaeJ KS1 WT and N206A using initial rate data ($k_{obs}$). Kinetic parameters $k_{inact}$ and $K_i$ were calculated from the plot using the Eq. 4.1, and errors were determined using a σ confidence interval

**Table 4.3** Calculated irreversible inhibition kinetic parameters for WT KS1 and KS1(N206A) with SNAC **27**

| | $K_i$ (mM) | $k_{inact}$ ($min^{-1}$) | $k_{inact}/K_i$ ($mM\ min^{-1}$) |
|---|---|---|---|
| WT KS1 | 2.0 ± 0.2 | 251 ± 11 | 125 ± 18 |
| KS1(N206A) | 4.2 ± 1.2 | 248 ± 55 | 59 ± 29 |

Applying the irreversible inhibition analysis to both KS domains yielded a $K_i$ value of 2.0 mM ± 0.2 for WT KS1 and 4.2 mM ± 1.2 for KS1(N206A), again suggesting a decreased affinity towards SNAC **27** upon removal of the active site Asn residue. Furthermore, the rate constant for the formation of the acyl-KS thioester ($k_{inact}$) is extremely similar for both WT and mutant, which is to be expected. Comparison of the $k_{inact}/K_i$ values reveals a decrease in efficiency of a factor ~2, similar to the Michaelis–Menten analysis. It is worth noting, that fitting the data shown in Fig. 4.11 to a linear model (which assumes a single-step, as opposed to the formation of a [KS•SNAC] complex) yields a decrease in the $k_{inact}/K_i$ values of ~1.7.

These data strongly suggests that the presence of an Asn at the X-Cys position enhances the preference of KS1 towards 2-amidoacetyl substrates. The occurrence of such intermediates is a result of incorporation of NRPS modules into the PKS assembly line, with upstream KS domains having to accept amino acid-derived intermediates. Although the Asn residue appears to aid substrate recognition, the phylogenetic analysis and experimental work on the KS1(N206A) mutant suggest that it is not essential for function. This is borne out by these KSs which do not possess an Asn residue at this position.

### 4.2.7 Substrate Specificity of BaeJ KS1(M268A) and (L450A)

As proposed in Sect. 4.2.4, the homology modelling identified two residues in the active site that may be sterically restricting α-substituted substrates from entering the KS active site. The homology models suggested that Met268 and Leu450 provided steric bulk either side of the α-position of the substrate. These residues were individually mutated to an Ala residue and initially incubated with SNAC **29**. However, in both cases no acylation was observed, but similar acylation levels to WT KS1 were obtained for the less sterically demanding SNACs. Due to the inability to observe acylation with SNAC **29**, it was hypothesised that the bulky isopropyl moiety of **29** may be preventing nucleophilic attack at the carbonyl of the thioester.

### 4.2.8 Effect of Salt on Dimerisation and Acylation

Published crystal structures of KS domains suggest that it is dimeric in structure [12, 13]. Additionally, a recent study of a KS from a *trans*-AT PKS, suggests that salt concentration can alter the oligomeric state of the KS [14]. The dimer interface is composed of ionic interactions; therefore high salt concentrations favour the monomeric form and vice versa. This effect was investigated for the range of KS domains available using a combination of analytical gel filtration and native MS. KS domains were analysed in 25 mM Tris and either 5 mM or 500 mM NaCl, to represent low and high salt conditions. In all cases, KS domains were observed to be monomeric at 500 mM salt, as expected. However, in 5 mM NaCl/$NH_4OAc$ most KS domains were either monomeric or insoluble at such low concentrations of NaCl/$NH_4OAc$.

In contrast to most KS domains, BaeJ KS1 was soluble in 5 mM NaCl/$NH_4OAc$, allowing the oligomeric state of the KS to be monitored using a combination of analytical gel filtration and native MS. Analytical gel filtration revealed exclusively monomer at 500 mM NaCl, however at 5 mM NaCl, the dimeric KS1 was observed at a lower elution volume. The dimer peak appeared to leech into the monomer peak suggesting the existence of an equilibrium between the two species (Fig. 4.12).

This observation was further confirmed using native MS. A range (5–250 mM) of $NH_4OAc$ concentrations were investigated for their influence upon the oligomeric state of KS1. A dimer-monomer equilibrium was observed up to 100 mM $NH_4OAc$, with 250 mM $NH_4OAc$ completely depleting the dimeric KS1 population (Fig. 4.13).

With the ability to control the oligomeric state of the KS, it seemed logical to test the capacity of the native SNAC **27** to acylate KS1 at different salt concentrations. Under the standard assay conditions used previously, KS1 in either 5 mM or

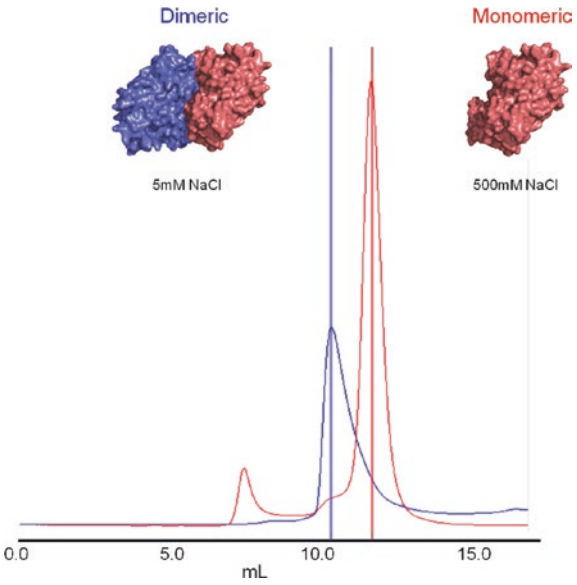



**Fig. 4.12** $UV_{280}$ traces from the analytical gel filtration, showing KS1 in 500 mM NaCl (*red trace*), and in 5 mM NaCl (*blue trace*). The shift in retention time suggests that dimeric KS1 is formed under low salt conditions

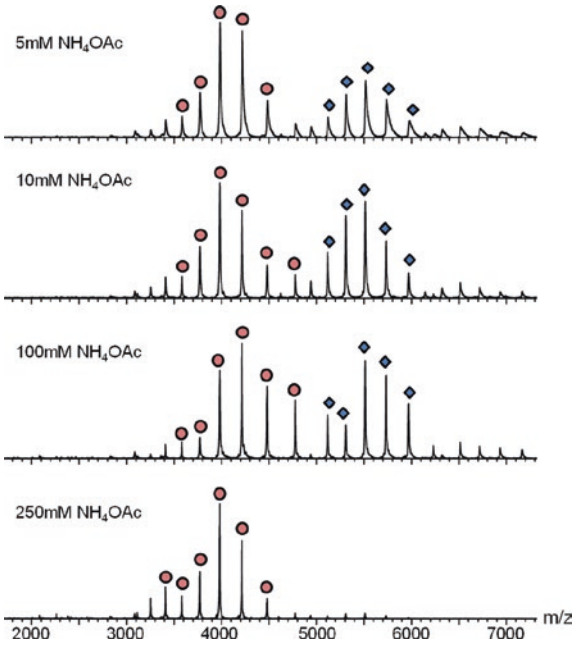



**Fig. 4.13** Native ESI-MS spectra of BaeJ KS1 sprayed from increasing concentrations of $NH_4OAc$. *Red circles* indicate monomeric KS1, and *blue diamonds* represent dimeric KS1

500 mM NaCl was incubated with SNAC **27** for 5 min, before quenching and MS analysis. Interestingly, approximately 3 times more acylation was observed with KS1 in the monomeric form (Fig. 4.14). This result was somewhat unexpected,

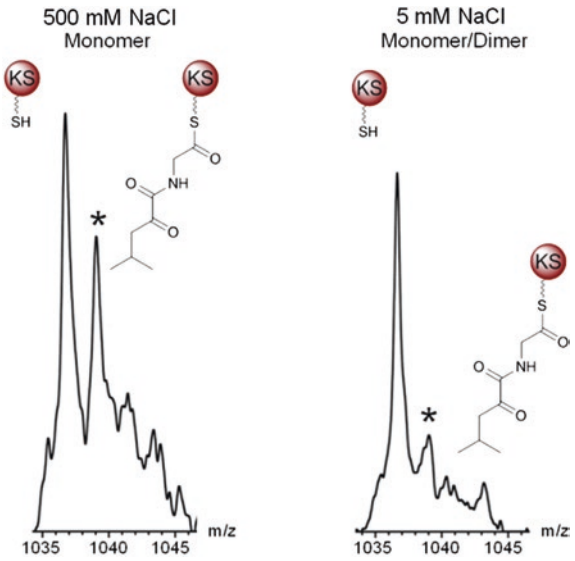



**Fig. 4.14** ESI mass spectra of the $71^{+}$ charge state of BaeJ KS1 following acylation with (**27**) under high (500 mM) and low (5 mM) $NH_4OAc$ conditions

as in the 'native' arrangement the KS is a dimer, and therefore a slow rate of acylation seems illogical for polyketide biosynthesis. To date, published crystal structures of KS domains show that the dimeric form of the KS closes off a large amount of the acyl-binding region of the pocket [14]. Additionally, relying on diffusion of small SNAC substrates into the KS active site is a somewhat artificial way to measure substrate specificity, when in the 'native' system; the substrate is tethered to an ACP and delivered to the active site. With these considerations in mind, it might not be surprising that the dimeric form of KS1 has a decreased rate of acylation, as the active site is far more closed than that of the monomeric form with proportionately less substrate able to access the active site.

This study of the KS oligomeric state provides valuable information regarding the use of SNACs as substrate probes. Although structurally SNACs effectively mimic the acyl-PPant chain, the fact that they rely on diffusion to access the active site is very different to the native situation of a tethered acyl-ACP delivering the substrate to the active site. The tethered ACP is likely to allow correct orientation for delivering the acyl-PPant moiety to the active site, regardless of the oligomeric state of the KS.

## 4.3 Conclusions

The work in this chapter has successfully demonstrated the first example of substrate specificity and selectivity exhibited by a KS domain immediately downstream of an NRPS module. Using full length substrate mimics of biosynthetic intermediates, BaeJ KS1 is shown to readily accept the both the natural

glycine-derived substrate (**27**) and its alanine analogue (**28**). It is noteworthy that PKSs do exist which naturally incorporate alanine from NRPS modules, including KS10 from the bacillaene cluster and KS5 from the thailandamide PKS. Experiments using a valine-derived substrate (**29**) resulted in no observable acylation; this can probably be attributed to steric effects restricting incorporation of substrates with a bulky group at the α-position. This constraint might be a reason why only Gly, Ala and Ser are found incorporated by NRPS modules preceding KS domains. Building on this point, from all known KS domains from *trans*-AT PKSs, none are predicted to accept an α-substituent larger than a dimethyl branch (a result of two consecutive methylations by a MT domain) [15, 16]. This would suggest that although KS domains downstream of NRPS modules are tolerant of the incorporation of small amino acid R-groups, the potential to engineer KS domains downstream of NRPS module introducing bulky amino acids may require more work.

Under the assay conditions, conducted at 0.5 mM substrate, SNAC thioesters lacking an amide or amine nitrogen at the 2-position were unable to acylate BaeJ KS1. The importance of the nitrogen at the 2-position was ascertained using SNACs with functional groups sequentially removed to examine their role upon acylation. Following on from work in Chap. 3, which identified a crucial specificity-dictating residue at the X-Cys position, structural and sequence analyses were carried out to identify the role of this residue in amino acid-accepting KS domains. A semi-conserved Asn residue was identified at the X-position, which was present in BaeJ KS1. Homology modelling suggested the potential for a hydrogen bond between the carbonyl of the Asn206 side chain and the nitrogen of the amide. Mutation of Asn206 to Ala reduced the affinity of KS1 towards 2-amidoacetyl substrates ($k_{cat}/K_m$ lowered by a factor of ~2).

In addition to the substrate specificity profiling of KS1, the effect of salt concentration upon the oligomeric state of the KS was examined. BaeJ KS1 was found to dimerise under low salt conditions (5 mM), albeit in an equilibrium with the monomer, and in the presence of high salt ($\geq$250 mM) was found to be exclusively monomeric. The acylation of KS1 was found to be 3 times more effective when in the monomeric state. With knowledge from crystal structures that the dimer interface blocks part of the KS binding pocket, it is conceivable that relying on diffusion to promote substrates into the KS active site is relatively artificial compared to the native arrangement of a tethered acyl-ACP.

Understanding the factors governing specificity and selectivity of KS domains can aid both biosynthetic assignment of PKS clusters and guide potential PKS engineering approaches. Using the data presented in this chapter is may be possible to augment the selectivity of other KS domains towards 2-amidoacetyl substrates, and further probe the possibility of incorporation of non-natural amino acid units into PKS assembly lines.

## References


1. T. Gulder, M. Freeman, J. Piel, The catalytic diversity of multimodular polyketide synthases: natural product biosynthesis beyond textbook assembly rules. Top. Curr. Chem. 1–53 (2011)
2. L.H. Du, C. Sanchez, B. Shen, Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. Metab. Eng. **3**, 78–95 (2001)
3. S.C. Wenzel, R. Muller, Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. Curr. Opin. Chem. Biol. **9**, 447–458 (2005)
4. K.J. Weissman, R. Muller, Protein–protein interactions in multienzyme megasynthetases. ChemBiochem **9**, 826–848 (2008)
5. S.J. Admiraal, C. Khosla, C.T. Walsh, A switch for the transfer of substrate between nonribosomal peptide and polyketide modules of the rifamycin synthetase assembly line. J. Am. Chem. Soc. **125**, 13664–13665 (2003)
6. C.T. Calderone, S.B. Bumpus, N.L. Kelleher, C.T. Walsh, N.A. Magarvey, A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both alpha- and beta-ketone reduction during chain growth. Proc. Natl. Acad. Sci. U.S.A. **105**, 12809–12814 (2008)
7. H.W. Chen, S. O'Connor, D.E. Cane, C.T. Walsh, Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. Chem. Biol. **8**, 899–912 (2001)
8. T.L. Schneider, B. Shen, C.T. Walsh, Oxidase domains in epothilone and bleomycin biosynthesis: Thiazoline to thiazole oxidation during chain elongation. Biochemistry **42**, 9722–9730 (2003)
9. J.G. Olsen, A. Kadziola, P. von Wettstein-Knowles, M. Siggaard-Andersen, S. Larsen, Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. Structure **9**, 233–243 (2001)
10. K. Watanabe, C.C.C. Wang, C.N. Boddy, D.E. Cane, C. Khosla, Understanding substrate specificity of polyketide synthase modules by generating hybrid multimodular synthases. J. Biol. Chem. **278**, 42020–42026 (2003)
11. J.Q. Wu, K. Kinoshita, C. Khosla, D.E. Cane, 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**, 16301–16310 (2004)
12. Y.Y. Tang, C.Y. Kim, G. Mathews II, D.E. Cane, C. Khosla, The 2.7-angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc. Natl. Acad. Sci. U.S.A. **103**, 11124–11129 (2006)
13. G. Pappenberger et al., Structure of the human fatty acid synthase KS-MAT didomain as a framework for inhibitor design. J. Mol. Biol. **397**, 508–519 (2010)
14. D.C. Gay et al., A close look at a ketosynthase from a *trans*-acyltransferase modular polyketide synthase. Structure **22**, 444–451 (2014)
15. J. Piel et al., Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. Proc. Natl. Acad. Sci. U.S.A. **101**, 16222–16227 (2004)
16. N. Pulsawat, S. Kitani, T. Nihira, Characterisation of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic. Streptomyces virginiae. Gene **393**, 31–42 (2007)

# Chapter 5
# Synthesis of Acyl-Acyl Carrier Proteins and Their Use in Studying Polyketide Synthase Enzymology

## 5.1 Introduction

As previously discussed, PKSs can be grouped into three distinct classes based on their enzymatic architecture, types I, II and III. Of these the type I and II PKSs both utilise an ACP to which the growing polyketide chain is tethered via a PPant prosthetic moiety attached to a conserved serine residue [1, 2]. Additional modification to the tethered acyl chain may also be carried out by other enzymatic domains, a feature particularly apparent in *trans*-AT PKSs. These systems are known to incorporate unusual enzymatic activities into biosynthetic pathways, including non-elongating KSs ($KS^0$), pyran synthases (PS), heterocyclase domains (HC) and in some cases, complete cassettes of *trans*-acting enzymes responsible for introducing carbon β-branches into the biosynthesis [3–6]. Therefore, cogent in vitro testing of such enzymatic activities ideally requires a substrate that accurately reflects the true intermediate in the biosynthesis, namely an acyl-ACP.

Currently, there are several published methodologies for the in vitro synthesis of acyl-ACPs. The most commonly employed method is the enzymatic transfer of the acyl-PPant group of an acyl-CoA on to an *apo*-ACP using a phosphopantetheinyl transferase (PPTase) enzyme [7]. This approach suffers from several disadvantages. Firstly, it requires production of *apo*-ACP from a strain deficient in the 4'-phosphopantetheinyl transferase *sfp* gene, and the subsequent synthesis of acyl-ACP is often slow and requires either production or purchase of the PPTase enzyme. Furthermore, attachment of more elaborate acyl groups requires access to synthetically challenging and expensive acyl-CoAs. The use of *N*-acyl imidazolides has also been reported for the synthesis of acyl-ACPs, and although effective at acylating holo-ACPs, these compounds are often difficult to synthesise [8].

The specific case of β-oxoacyl transfer on to holo-ACPs from either –CoA or *N*-acetylcysteamine (SNAC) was first reported by Cox and Simpson [9, 10]. Using a type II ACP, 10 equivalents of malonyl-CoA were found to be sufficient to produce



M. Jenner, *Using Mass Spectrometry for Biochemical Studies on Enzymatic Domains from Polyketide Synthases*, Springer Theses,
DOI 10.1007/978-3-319-32723-5_5

fully malonylated ACP. Crucially, however, simple acetyl and butyryl-CoAs were unable acylate the ACP, suggesting that the β-oxo functionality was important. The authors suggest a possible role for a conserved arginine side chain, present in type II ACPs, in stabilising the enolate tautomer of the acetoacetyl moiety [9]. This work was further developed using a range of β-oxoacyl SNACs to probe the nature of the transacylation reaction. Here, 50 equivalents of β-oxoacyl-SNAC thioester were used to acylate the ACP, with yields at ~60 % [10]. This promising work has allowed rapid access to a range of β-oxoacyl-ACPs, which can be used as surrogate intermediates of pathways in type II PKS systems. To date, this methodology has been limited to acyl groups containing β-keto functionality, and has not been explored using ACPs from type I modular PKSs, which often produce highly reduced intermediates.

Given the aforementioned restrictions, acyl *N*-acetylcysteamine (SNAC) thioesters are often used as small molecule mimics of ACP-bound acyl units to test reactivities of PKS components [11, 12]. Although these simple substrate analogues are relatively straightforward to synthesise, their ability to act as a realistic surrogate of the acyl-ACP is open to debate. Generally, high concentrations (0.5–2 mM) of SNAC are required to drive the reaction [13, 14]. In addition, the small and simple nature of the SNAC group may allow binding with higher degrees of freedom, potentially masking true substrate specificity of domains. Therefore, for a more realistic representation of a tethered substrate, utilising the full extent of the PPant arm, an acyl-ACP is required.

In this chapter, a simple and general method for the production of acyl-ACPs from a range of SNAC-thioesters is detailed, and their valuable usage as complete substrate mimics for studying the activity of a KS (PsyA KS1) domain, AT (PedD) domain and an AH (PedC) is highlighted.

### *5.1.1 The Phosphopantetheine Ejection Assay*

Studies conducted by Kelleher and co-workers in 2006 characterised an efficient gas-phase tandem MS elimination reaction to identify phosphopantetheine-bound substrates on ACPs [15]. During collisional activation of *holo*-ACP domains, it was found that the phosphopantetheinyl moeity is ejected from the ACP, yielding a characteristic protonated imine ion at m/z 261. The mechanism of this elimination is believed to leave a phosphate anion attached to the serine of the ACP (*apo*-ACP + 80 Da) minus one charge, as the carbonyl of the amide displaces the pantetheinyl unit, generating a five-membered tetrahydrofuranol ring with a protonated imine (Fig. 5.1).

This methodology has allowed identification of biosynthetic intermediates during enzymatic synthesis, by effectively reducing large >100 kDa proteins to small molecule fragment ions. In this chapter, the phosphopantetheine ejection assay will simply be used to characterise the synthesis acyl-ACPs, however, in later chapters the assay will be employed for identification of other intermediates.

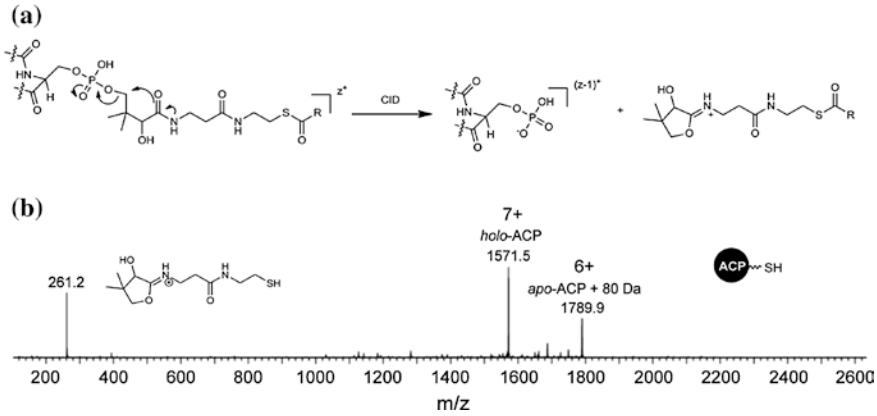



**Fig. 5.1** **a** Proposed mechanism of phosphopantetheine elimination reaction. **b** Example tandem MS spectrum of an *holo*-ACP domain. In this instance, the 7 + *holo*-ACP ion is isolated and activated, yielding the ejection ion at m/z 261 and the 6 + *apo*-ACP + 80 Da

## *5.1.2 PedC: Acyltransferase-like Proofreading Domain*

The pederin PKS cluster, which is biosynthetically similar to that of psymberin, harbours two AT homologs PedC and PedD. Although both domains are similar at sequence level, a phylogenetic analysis places PedD into the AT1 clade with other known malonyl transferases, whereas PedC is located in the largely unstudied AT2 clade [16] (Fig. 5.2). An in vitro functional analysis of these domains has shown that PedD serves as a malonyl-acyltransferase enzyme as predicted, whereas PedC catalyses the hydrolysis of acyl thioesters derived from both SNACs and acyl-ACPs, thus likely acting as a proof-reading component that removes stalled polyketide intermediates from the PKS [16].

A selection of short chain SNACs have been used previously to examine the specificity of PedC, with the 3-hydroxybutyryl, α,β-unsaturated and 4-methylpentanoic SNACs among the better substrates. In addition, PedC showed hydrolytic activity against both acetyl- and hexanoyl-ACP, but crucially, not malonyl-ACP which represents an essential precursor for polyketide biosynthesis. The two acyl-ACPs used in the aforementioned study were produced using the time-consuming PPTase coupling methodology, hindering the profiling of a full range of acyl chains. Kinetic parameters for PedC were determined to be $K_m = 88.48$ mM and $k_{cat}/K_m = 14.41\ M^{-1}\ s^{-1}$ using the 3-hydroxybutyryl (**25/26**) SNAC, although these values are advised to be preliminary by the authors due to the simplistic nature of the substrate [16].

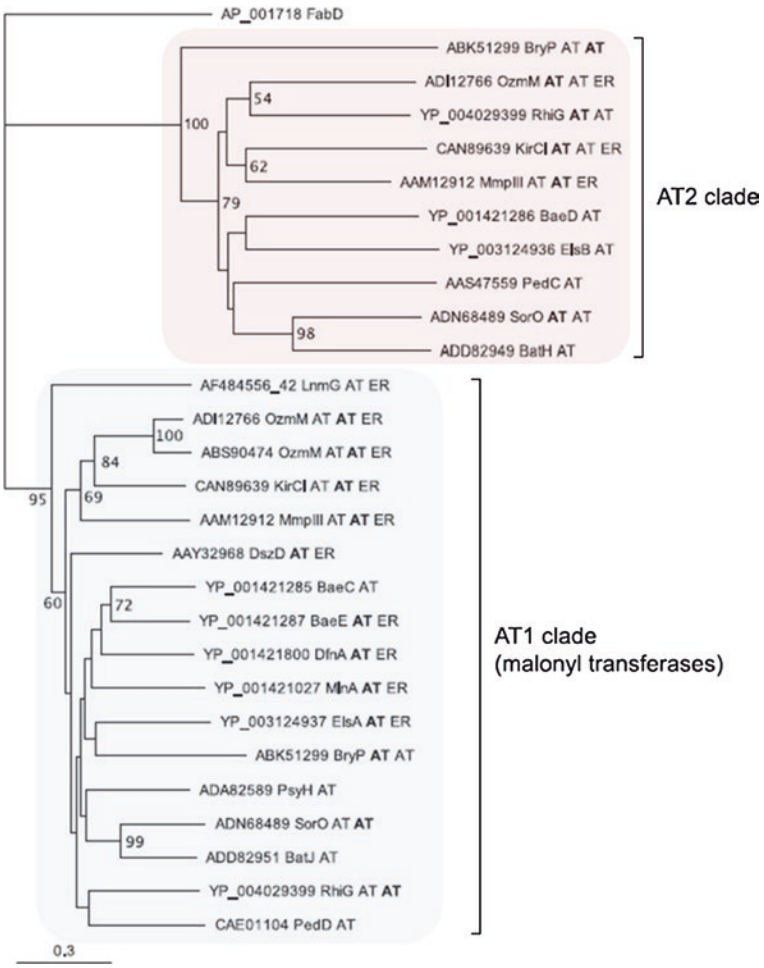



**Fig. 5.2** Phylogram of AT-like domains from *trans*-AT PKS clusters. Tip labels indicate protein name and the enzymatic architecture they are contained within. For multi-domain proteins, the hydrolase domain is shown in *bold*. The two seperate clades are shown, the malonyl transferases in *blue* and the AT2 clade (potential AH domains) in *red*

## 5.2 Results

### *5.2.1 Purification of PsyA ACP3(Δ37,38)*

The ACP used in this study originated from the psymberin (*psy*) *trans*-AT PKS. Psymberin is an extremely rare and highly potent cytotoxin isolated from the marine sponge *Psammocinia* aff. *Bulbosa* [17], but produced by an as-yet unculturable symbiotic bacterium [18]. Since gene inactivation experiments for pathway

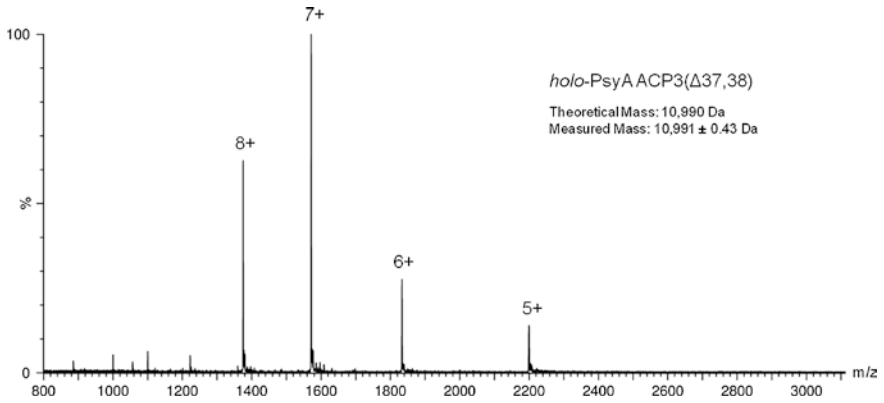



**Fig. 5.3** nESI mass spectrum of PsyA ACP3(Δ37,38) sprayed from 80:20 MeCN:$H_2O$ 0.1 % TFA, resulting in a measured mass of 10,991 ± 0.43 Da

characterization cannot be conducted for such bacteria, research is often focussed on utilising alternative methods, often relying on heterologously expressed components of these complex PKS machineries [5, 13].

To investigate MS-based strategies using ACP-bound intermediates, we focused on PsyA ACP3 as a model, which is situated at the C-terminus of the PsyA protein. The *holo*-ACP was prepared as described in the Sect. 2.2.1. ESI-MS measurement of the ACP revealed a mass of 10,991 ± 0.43 Da, in good agreement with the theoretical mass of 10,990 Da (Fig. 5.3). Collision induced activation of the ACP yielded the ejected PPant moiety at *m/z* 261, demonstrating an unacylated *holo*-ACP [15].

### *5.2.2 Synthesis of Acyl-ACPs*

Acylation of the ACP PPant thiol was achieved by incubation of this *holo*-ACP with 50 equivalents of SNAC thioester as detailed in Sect. 2.2.4 and outlined in Scheme 2.3. Reactions were monitored by ESI-MS for the presence of the acyl-ACP product (Fig. 5.4). Generally, acylation reactions were complete after 3–4 h at 25°C. Removal of excess SNAC was achieved by spin filtration using 5 kDa MWCO columns. Initial acylation experiments conducted with WT PsyA ACP3 revealed that, under the forcing conditions employed, Cys37, near the C-terminus of the ACP was acylated in addition to the thiol of the PPant chain. To avoid resulting complications in quantifying the amount of acylation at the PPant arm, a PsyA ACP3(Δ37,38) mutant was constructed to remove the C-terminal Cys and Val residues. This mutant was used in subsequent experiments, and it is recommended that a similar procedure be adopted for other ACPs. The transthioesterification reaction between acyl-SNAC and *holo*-ACP reached yields of 85–95 % after 3–4 h for acyl units **2–7**, **11** and **12**. The three β-hydroxyl species (**8**, **9** and **10**) gave

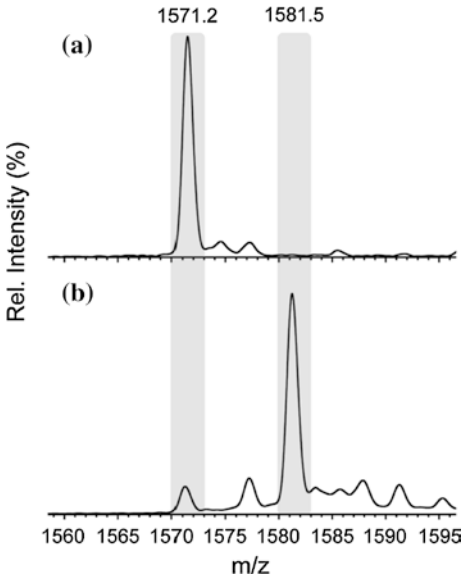



**Fig. 5.4** nESI mass spectra (ESI-MS) of the $7^+$ charge state for **a** *holo*-ACP and butyryl-ACP, **b** following incubation with butyryl-SNAC thioester

somewhat reduced yields of ~75 %, whilst no acylation was observed with a valine derived SNAC (**13**), which is likely to be due to the steric hindrance incurred at the α-position. Several α,β-unsaturated SNACs were also tested for loading, however a competing Michael addition inhibited acylation of the PPant chain.

Using a high molar excess of acyl-SNAC was found to force the transacylation of PsyA ACP3 for many substrates, even in the absence of β-oxo functionality in the acyl donor. Having devised a simple method for the general and facile production of a range of acyl-ACPs, we sought to use these as tools to probe the specificity of enzymatic domains used in polyketide biosynthesis (Fig. 5.5).

### *5.2.3 Loading of KS Domains Using Acyl-ACPs*

The psymberin polyketide core is the putative biosynthetic product of two genes *psyA* and *psyD* [17]. The first three PKS modules are encoded by *psyA*, which initiates the biosynthesis via an N-terminal GCN5-related *N*-acetyltransferase (GNAT) domain. GNAT domains catalyse the decarboxylation of malonyl-CoA to acetyl-CoA and subsequent loading to the downstream ACP [19]. Therefore, the native substrate of PsyA KS1 is predicted to be acetyl-ACP.

The substrate selectivity of PsyA KS1 has been previously examined using SNAC-thioesters, with acetyl-SNAC being shown as the preferred substrate, despite a rather promiscuous substrate profile (see Sect. 3.2.3). It was postulated that the use of the corresponding acyl-ACPs would represent a more realistic approximation of the native system. Moreover, de-acylation of the small ACP domain could be monitored by mass spectrometry. The assay consisted of incubation of PsyA

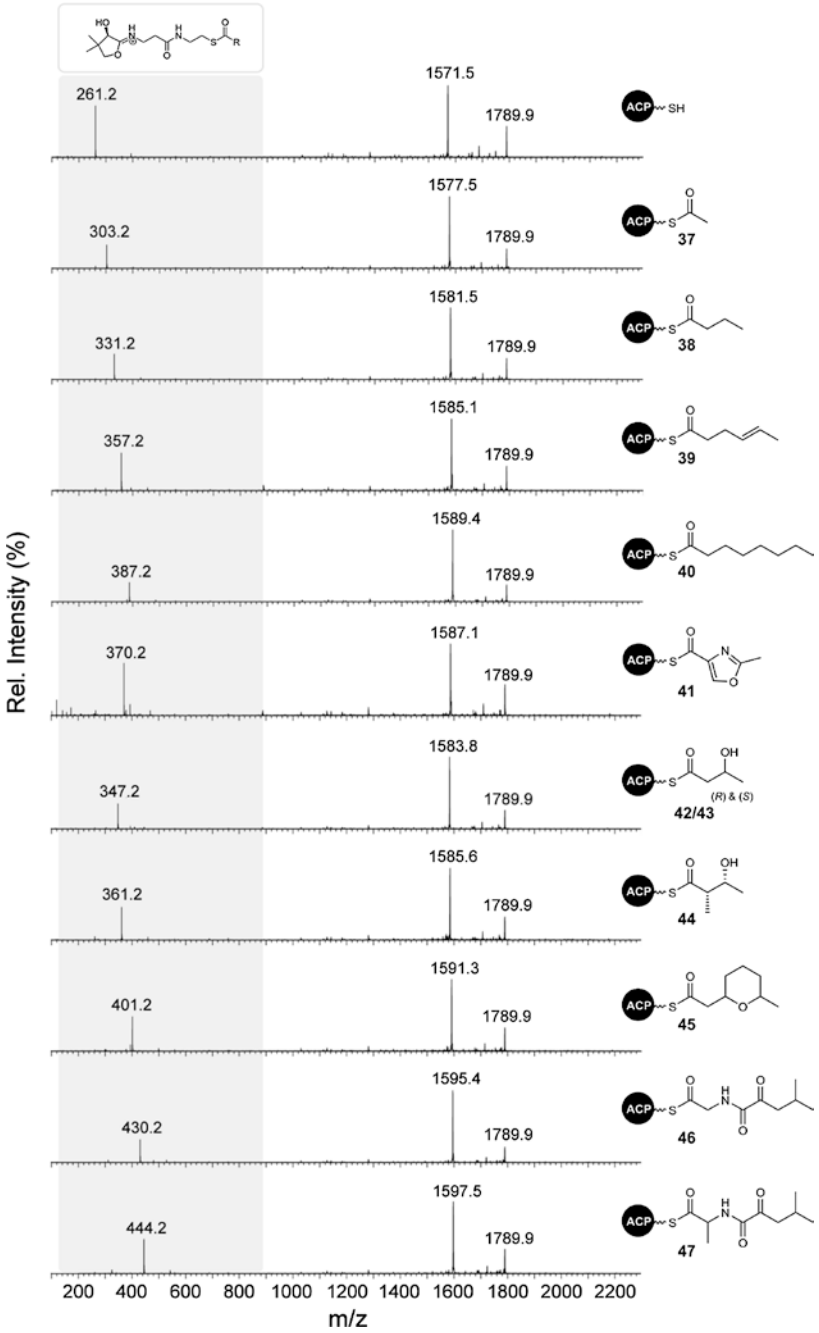



**Fig. 5.5** $MS^2$ spectra of the isolated $7^+$ charge state of acyl-ACPs. The *holo*-ACP $MS^2$ spectrum (*top*) shows the detection of the PPant 261.2 species. Acyl chains attached to the PPant arm with corresponding masses shown

KS1 with 100 μM acetyl-ACP (**37**). Acyl-transfer was monitored over a 30 h time period using mass spectrometry to measure the emergence of the *holo*-ACP signal and disappearance of the acetyl-ACP (see Sect. 2.2.5). Given the indirect nature of monitoring KS acylation in this assay, control measurements were recorded in the absence of KS to account for any hydrolytic loss of the acyl chain. In addition, a further control point was measured with BSA in place of the KS in order to account for any non-specific transfer or hydrolysis catalysed by the presence of protein. Using these aforementioned controls it was therefore possible to deduce that any increase in *holo*-ACP signal would be solely due to the function of a KS. Finally, the assay was conducted at three KS concentrations, 20, 40 and 80 μM, to establish that the rate of ACP deacylation correlated with KS activity.

Acetyl-ACP3 (**37**) was synthesised as a mimic of PsyA ACP1, which is believed to deliver an acetyl unit from a decarboxylative loading catalysed by an upstream GNAT domain. Acetyl-ACP (**37**) was incubated with increasing concentrations of PsyA KS1 over a 30 h time period, with MS measurements taken at regular intervals. The rate of acyl transfer to PsyA KS1 was found to increase with KS concentration, indicating that the KS is solely responsible for catalysing removal of the acyl group by transthioesterification (Fig. 5.6). Initial rates of PsyA KS1 acylation were determined to be: 0.95 ± 0.08 μM $h^{-1}$at 20 μM, 1.59 ± 0.04 μM $h^{-1}$ at 40 μM and 2.83 ± 0.12 μM $h^{-1}$ at 80 μM. The pseudo-linear regions of the curves shown in Fig. 5.6 were used to deduce the initial

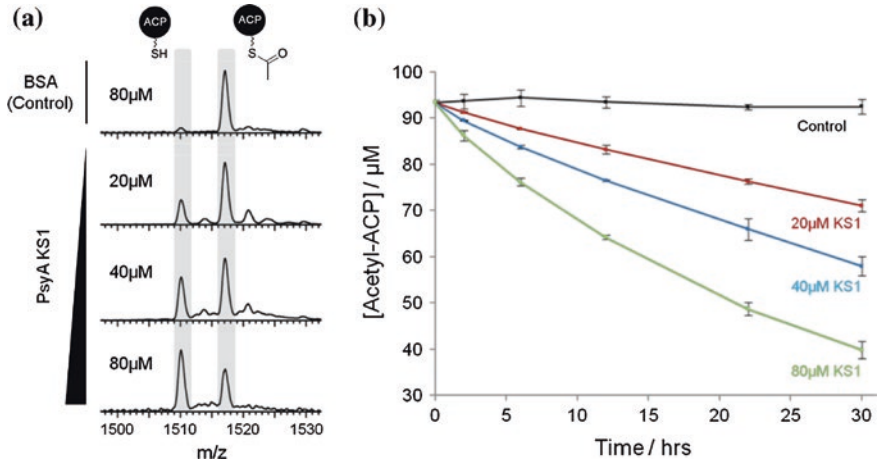



**Fig. 5.6** **a** Stacked ESI-MS spectra of the $7^+$ charge state of acetyl-ACP and deacetylated ACP following 30 h incubations with increasing concentrations of PsyA KS1. The 30 h BSA control was used to monitor the presence of non-specific transfer/hydrolysis of the acetyl moiety resulting from the presence of protein in the sample. **b** Kinetic plot showing acyl transfer from acetyl-ACP to PsyA KS1 at increasing concentrations of KS1. Error bars in the mean of 3 repeats are given at a ± 2σ (95 %) confidence interval. Control measurements were taken in the absence of KS domain. In addition, a 30 h control point was taken in the presence of BSA, which was indistinguishable from the 30 h control point, indicating no non-specific acyl-transfer. Data supplied by J. Afonso

velocity in each case. The 30 h BSA control point was observed to be indistinguishable from that of the 30 h control point, indicating that the transthioesterification reaction results from KS activity. These data indicate that acyl-ACPs can be used as substrate probes of KS domains using this methodology. Despite its indirect nature, the assay possesses a number of advantages over direct monitoring of KS acylation by SNACs. In addition to the more realistic substrate mimetic, monitoring de-acylation of the smaller ACP protein improves the resolution at which the acylated and unacylated forms can be observed by MS.

### *5.2.4 Purification of Acyl Hydrolase PedC*

PedC was co-expressed with GroEL to aid solubility using a previously described methodology [16], therefore calculation of PedC concentration was achieved using a band intensity approach calibrated with known amounts of BSA. Band intensities of PedC and BSA were measured using ImageJ software, and following construction of a calibration curve the concentration of PedC stocks was calculated to be 25 μM, and in a 3:1 ratio with GroEL (Fig. 5.7a). ESI-MS analysis of the GroEL-PedC mixture revealed a mass of 39.1 kDa for the PedC domain and 57.2 kDa for GroEL (Fig. 5.7b).

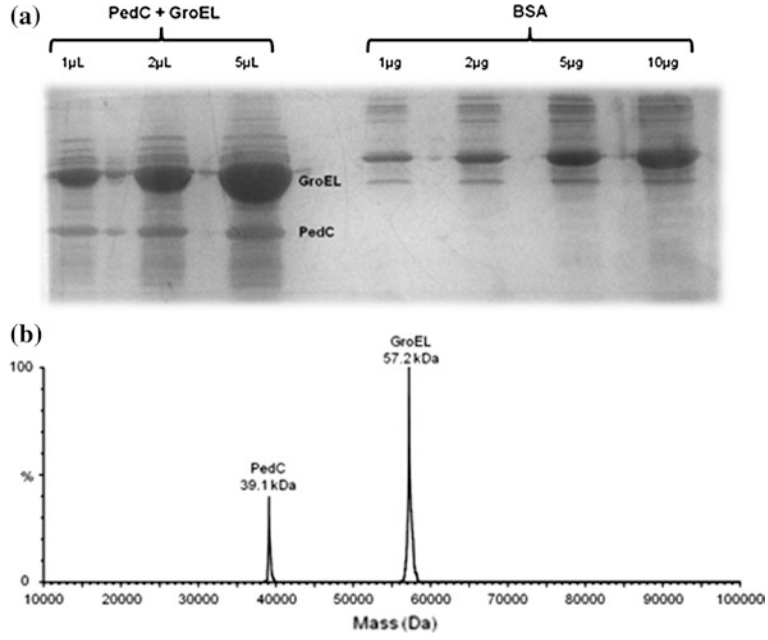



**Fig. 5.7** Analysis of PedC + GroEL co-expression. **a** 12 % SDS-PAGE of purified PedC and GroEL, showing expression in a ~1:4 ratio. The BSA band intensities used for the calibration are shown. **b** Transformed nESI-MS spectrum of PedC + GroEL, with PedC reporting a measured mass of 39.1 kDa

### 5.2.5 *Hydrolysis of Acyl-ACPs By PedC*

Utilising our facile synthesis, acyl-ACPs (**37–47**) were produced to mimic the full range of functionality seen in the intermediates of pederin biosynthesis. The first aspect of PedC specificity probed was the acyl-chain length. Here straight chain acetyl, butyryl, hex-4-enyl and octanoyl acyl-ACPs (**37–40**) were employed. Upon incubation of the acyl-ACPs with PedC and subsequent MS analysis, the appearance of the *holo*-ACP signal indicated hydrolysis of the acyl-ACP (Fig. 5.8). Under the conditions employed (see Sect. 2.2.6), PedC successfully hydrolysed acetyl-ACP at a rate of 10.1 ± 0.75 μM $min^{-1}$. Increasing the chain length to butyryl caused the rate to half, 4.8 ± 0.67 μM $min^{-1}$. The hexenyl- and octanyl-ACPs were observed to decrease the hydrolysis rates still further (Fig. 5.9). The gradual decrease in hydrolytic rate with increased substrate chain length may suggest that PedC is actually optimized for short chain intermediates.

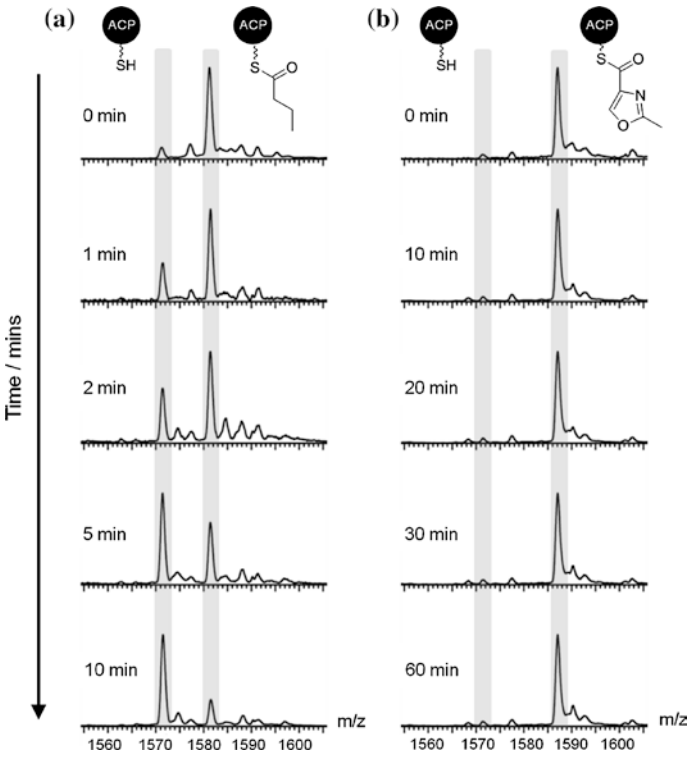



**Fig. 5.8** Stacked nESI-MS spectra of the $7^+$ charge state of **a** butyryl-ACP and **b** oxazole-ACP, following incubation with PedC. The PedC catalysed hydrolysis of butyryl-ACP is shown over 0–10 min by an increase in the *holo*-ACP signal, however no hydrolysis was observed with oxazole-ACP after 60 min

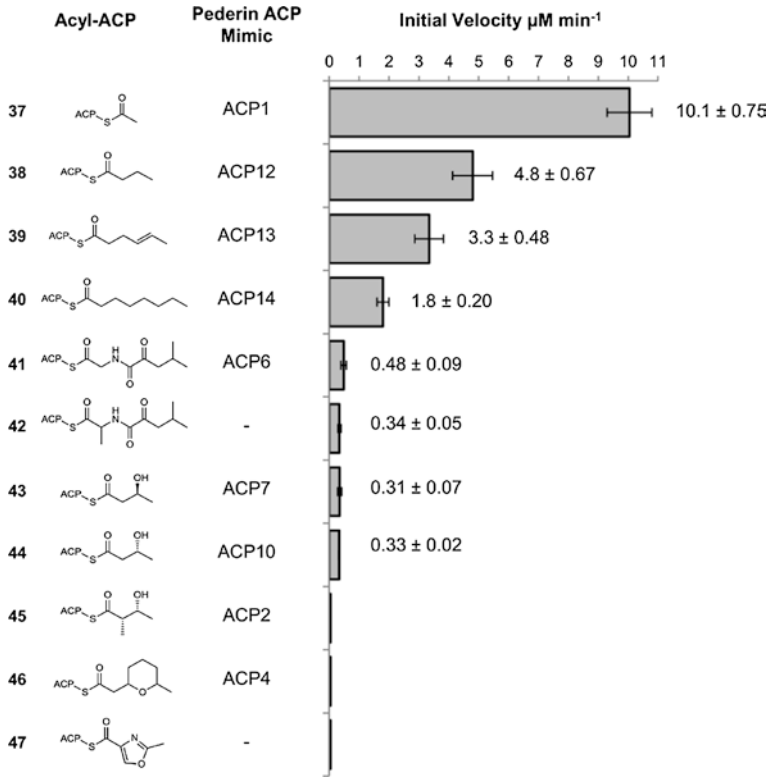



**Fig. 5.9** PedC hydrolysis rates observed for acyl-ACPs **37–47**. Initial rate was calculated from pseudo-linear region of PedC hydrolysis plots. Errors in the mean of 3 repeats are given at a ± 2σ (95 %) confidence interval. For acyl-ACPs **10–12** no hydrolysis greater than ~0.008 μM $min^{-1}$ was observed over the time scale of the experiment (60 min) compared to a control measurement

Although technically a mimic of the bacillaene biosynthetic intermediate at ACP2, the glycine-derived acyl-ACP (**41**) was used to approximate the pederin biosynthetic intermediate at ACP6; owing to a common β-amido functionality (Fig. 1.10). Limited hydrolysis was observed with (**6**) over the course of the experiment, with an initial rate of 0.48 ± 0.09 μM $min^{-1}$, which is ten-fold slower than the butyryl-ACP. An alanine derivative (**42**) was also tested with PedC to examine the effect of an α-methyl branch on the rate of hydrolysis. A small decrease in activity was observed for ACP (**42**), when compared to its non-α-methyl analogue (**41**).The (*R*) and (*S*) configured β-hydroxyl acyl-ACPs (**43** and **44**) were used to test whether PedC harbours any stereo-selectivity towards these acyl-chains. Both β-hydroxyl configurations are proposed to exist during pederin biosynthesis occurring at ACP7 and ACP10 (Fig. 1.10). The latter is believed to be a site of PedG oxygenase cleavage, causing decarboxylation of the (*S*) configured β-hydroxyl

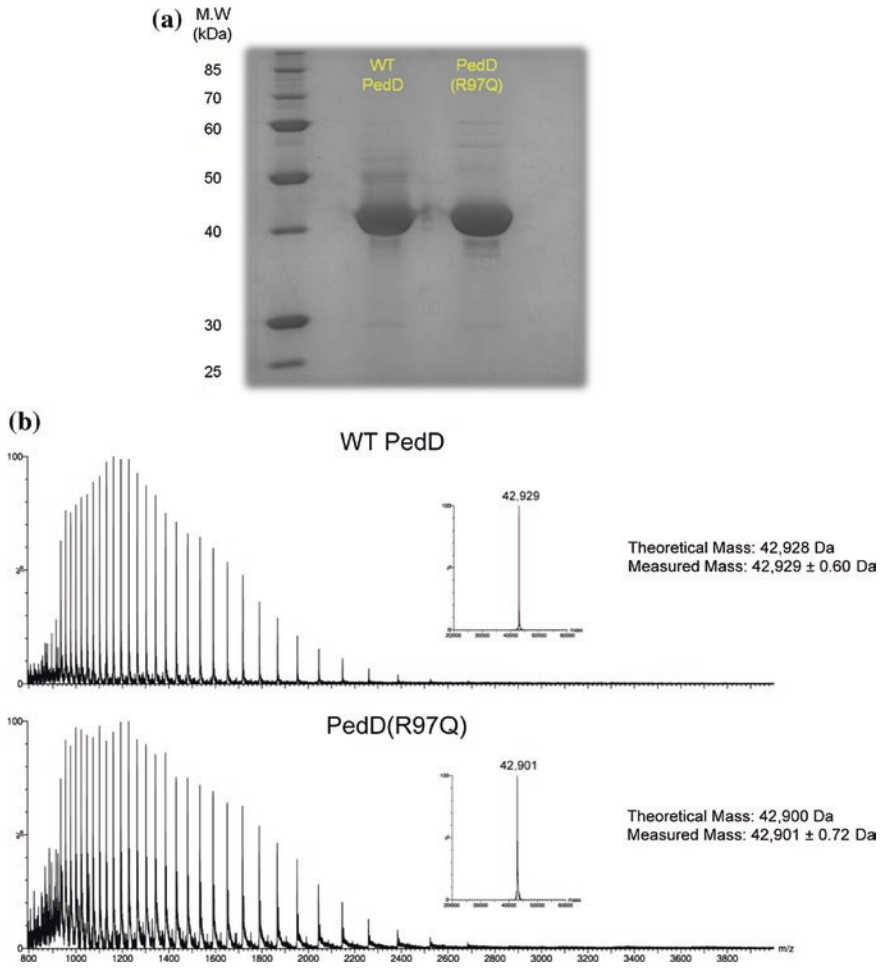



**Fig. 5.10** **a** 12 % SDS-PAGE gel of WT PedD and PedD(R97Q). **b** nESI mass spectra of WT PedD and PedD(R97Q) sprayed from 80:20 MeCN:$H_2O$ 0.1 % TFA. Theoretical and measured molecular weights are annotated on the spectra

chain, resulting in the release of a pederin precursor. Upon incubation with PedC, both (*R*) and (*S*) stereoisomers produced similar rates of hydrolysis: 0.31 ± 0.07 and 0.33 ± 0.02 μM $min^{-1}$ respectively. These data show that no stereoselectivity is exhibited by PedC for these simple β-hydroxyl chains.

The α-methyl-β-hydroxyl acyl-ACP (**45**) is in the correct stereoconfiguration to act as a substrate mimic of the intermediate at ACP2 in pederin biosynthesis. Incubation of (**10**) with PedC yielded no measurable hydrolysis, implying that an α,β-branched substrate provides sufficient steric bulk to prevent hydrolysis under the conditions employed. The 6-methyltetrahydro-2*H*-pyran-ACP (**46**) was synthesised to represent the ACP4 intermediate in pederin biosynthesis. PedC was unable

to hydrolyse the 6-methyltetrahydro-2*H*-pyran-ACP after 1 h incubation. This demonstrates that, in this assay, the pyran ring is too sterically demanding for the active site of PedC.

It has been proposed that the proofreading ability of PedC could be harnessed to elucidate the mechanistic steps in the biosynthesis of other biosynthetic pathways [20]. This requires PedC to be sufficiently promiscuous with regard to its specificity to hydrolyse unnatural intermediates from other clusters [16, 20]. Therefore, in addition to the acyl-ACP mimics of the pederin intermediates, we also tested oxazole-ACP (**47**), present as an intermediate of both rhizoxin and chivosazol biosynthesis. No hydrolytic activity was observed towards the oxazole-ACP, suggesting that the five-membered heterocycle is not tolerated by the PedC active site in this assay.

The acyl-ACP assay shows that PedC has a strong preference for short, unbranched substrates. This result suggests that the proofreading function of PedC may be most effective at the ACP1 stage or in cases of unproductive decarboxylation of ACP-bound malonyl units, as previously suggested for type II TEs occurring in *cis*-AT PKS pathways [21, 22]. It has previously been observed that inactivation of the terminal thioesterase domain in the bacillaene and rhizoxin pathway results in hydrolysis of short- and long-chain intermediates off the PKS [3, 23, 24]. This apparent discrepancy might be due to an additional proof-reading activity present in these pathways or to a different activity profile of the cognate AHs. Alternatively, PedC and its homologues could hydrolyse off the acyl chain with reasonable efficiency under conditions at which polyketide biosynthesis has completely stalled. As a further point of investigation, the predicted malonyl-specific AT for the pederin cluster, PedD, was expressed and purified to explore the biochemical factors dictating transferase activity instead of hydrolase activity.

### *5.2.6 Purification of WT PedD and PedD(R97Q)*

Both WT-PedD and PedD(R97Q) were overexpressed and purified as His-Tag fusion proteins, as described in Sect. 2.2.1. ESI-MS measurement of WT-PedD and PedD(R97Q) revealed masses of 42,929 ± 0.60 Da and 42,901 ± 0.72 Da respectively, in good agreement with their theoretical masses (Fig. 5.10).

### *5.2.7 Extender Unit Specificity of PedD*

The extender unit specificity of an AT can be predicted using the 'fingerprint residues' within the two catalytic motifs. Presence of the GH**S**(I/V)G and HAF**H** motifs signifies a malonyl-specific AT, whereas the GH**S**QG and YAS**H** motifs direct for a methylmalonyl-specific AT (catalytic residues highlighted in bold). The binding site residues in a malonyl-specific AT are believed to exlcude α-substituted

extender units via steric clashes with the phenylalanine residue of the HAF**H** motif [25, 26]. However, the presence of a comparatively smaller serine residue in methylmalonyl-accepting ATs (YA**S**H) means that these AT domains are unable to discriminate against malonyl-CoA. Despite this apparent contradiction in specificity, the methylmalonyl-specific ATs from the pikromycin PKS have been shown to hydrolyse the incorrect acyl-enzyme at a faster rate than the correct acyl-enzyme intermediate. Therefore, the hypothesis is that methylmalonyl-specific ATs act as transferases when presented with the correct extender unit, and a hydrolase when the incorrect extender unit is loaded [27]. The mechanism by which these ATs are able to preferentially 'shield' methylmalonyl units from hydrolysis remains unclear.

PedD is denoted as a malonyl-specific AT as it contains GH**S**LG and GAF**H** motifs [16, 28]. Although these motifs differ slightly from the aforementioned conensus sequences, these rules have been generated using predominantly ATs from *cis*-AT PKS or FA systems, whereas PedD is a *trans*-acting AT. The presence of the sterically demanding phenylalanine residue in the GAF**H** motif suggests PedD is likely to be specific for malonyl-CoA. Incubation of PedD with 4 equivalents of malonyl-CoA yielded ~70 % malonylated PedD (Fig. 5.11). To further investigate the substrate tolerance of PedD, acetyl- and methylmalonyl-CoA were incubated with PedD (see Sect. 2.2.7). Even with an 8-fold excess of each CoA, no acylation was observed in each case, indicating that PedD is indeed malonyl-specific as the motifs hypothesise (Fig. 5.11).

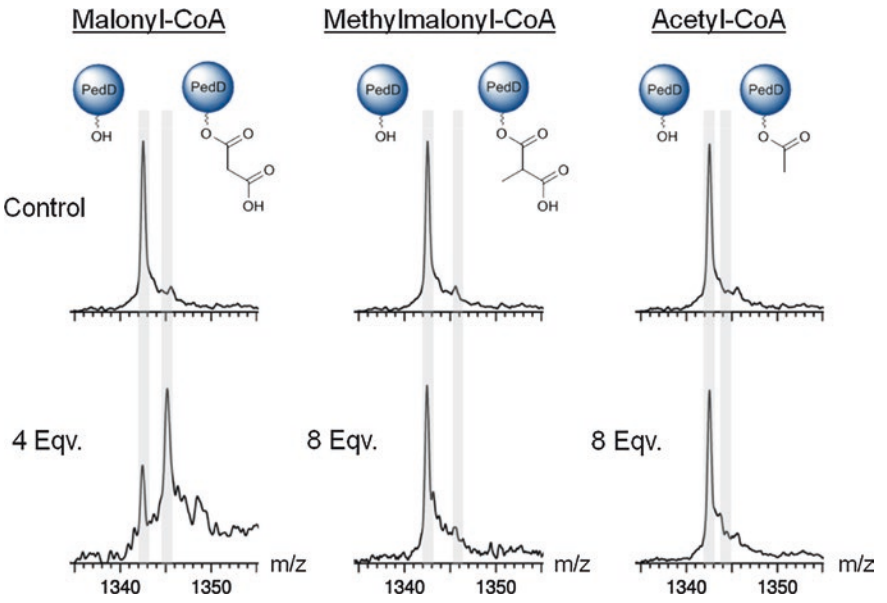



**Fig. 5.11** nESI mass spectra of WT PedD following incubation with malonyl- (4 eqv.), methylmalonyl- (8 eqv.) and acetyl-CoA (8 eqv.). Approximately 70 % malonyl-ACP was generated from 4 eqv. of malonyl-CoA, whereas no methylmalonylation/acetylation was observed used 8 eqv. of methylmalonyl-/acetyl-CoA

### 5.2.8 PedD-Catalysed Malonly-Loading of ACP

Following loading of the AT domain with the appropriate extender unit, the second half of the AT catalytic cycle is the transfer of this unit onto the phosphopantetheine moeity of the *holo*-ACP. In order to investigate the ability of PedD to transfer malonyl, fully malonated PedD was incubated in a 1:1 ratio with *holo*-PsyA ACP3. Aliquots were removed at various time intervals, desalted and analysed by MS (see Sect. 2.2.8). Interestingly, the transfer of the malonyl unit was found to be almost immediate, with ~70 % of the ACP malonyled after 1 min. However, following further incubation, the amount of unmalonated, *holo*-ACP began to increase slightly, plateauing at ~45 % after 20 min (Fig. 5.12). This suggested that either the malonyl moeity was hydrolysing from the ACP, or that back-transfer onto the AT was occuring as part of an equilibrium.

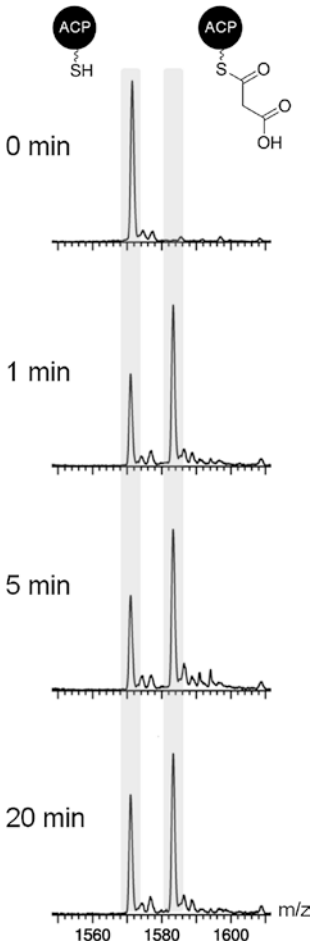



**Fig. 5.12** Stacked nESI-MS spectra of the $7^+$ charge state of *holo*-ACP and malonyl-ACP following a 20 min incubation with malonyl-PedD. The transfer of the malonyl unit from PedD to the ACP can be clearly monitored by the emergence of the signal at m/z 1583

Due to the nature of the analysis it was possible to monitor the malonylation state of PedD at the same time as the ACP. Upon examination of the PedD peaks, it was found to be ~50 % malonylated after 20 min. This observation suggested that an equilibrium was reached during the transfer assay, with back-transfer of the malonyl moeity onto the AT allowing the equilibrium to approach a 50:50 ratio. These results provide the intriguing possiblity that back-transfer of extender units (such as malonyl) may indeed happen in the native PKS system.

To investigate this phenomenon, the reverse reaction was carried out by incubating malonyl-ACP with unmalonylated-AT and monitoring any back-transfer of the malonyl unit onto the AT. Interestingly, the back reaction was found to proceed to approximately the same extent as the forward reaction, with ~55 % of the PedD population observed to be malonyled (Fig. 5.13).

This result suggests that the AT-catalysed loading of malonyl-units onto the ACP is, indeed, a reversible reaction setting up an equilibrium (Scheme 5.1). This provokes the logical question of how, in the native PKS system, does the malonyl loading reaction become essentially irreversible to allow the subsequent Claisen condensation reaction to occur. A likely scenario is that excess extender unit drives the equilibrium towards the malonylation of ACP, and disfavouring the reverse reaction by Le Chatelier's principle (Scheme 5.2).

This hypothesis can be easily probed using the current MS-methodology by providing an excess of malonyl-CoA to the reaction. The increased concentration (50 equivalents) of malonyl-CoA effectively mimicks the limitless –CoA substrate pool in the cell, theoretically pushing the equilibrium at position 2 to the right (Scheme 5.2), producing exclusively malonyl-ACP. This reaction was conducted and analysed after 5 min, resulting in a yield of approximately 95 %

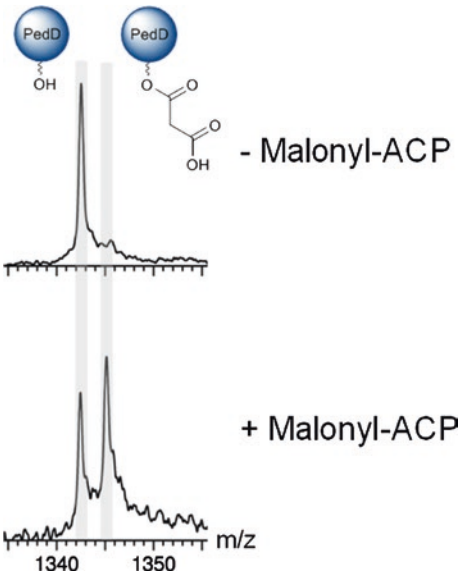
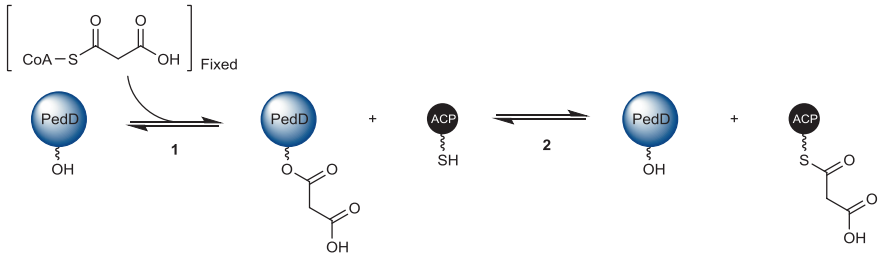
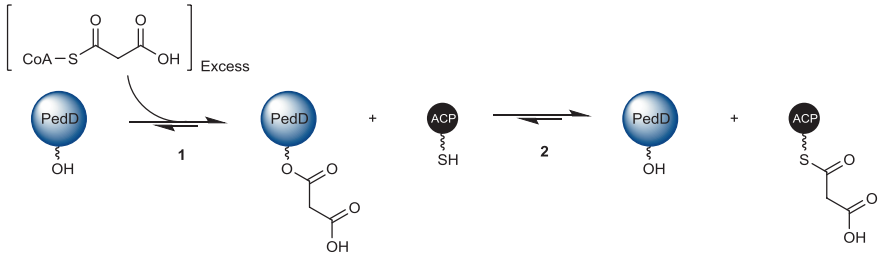



**Fig. 5.13** Stacked nESI-MS spectra of the $32^{+}$ charge state of WT PedD before (*top*) and after (*bottom*) addition of 1 equivalent of malonyl-ACP. The latter clearly shows the attachment of a malonyl unit to the PedD active site at ~55 %, indicative of the back reaction occurring




**Scheme 5.1** The 2-stage equilibrium set up between malonyl-CoA, malonyl-PedD and *holo*-ACP at a fixed, 1:1:1, ratio of malonyl-CoA:PedD:*holo*-ACP. In this instance, the equilibrium at position 2 lies at 50:50, due to the fixed concentration of malnoyl-CoA




**Scheme 5.2** The 2-stage equilibrium set up between malonyl-CoA, malonyl-PedD and *holo*-ACP at a 50:1:1, ratio of malonyl-CoA:PedD:*holo*-ACP. In this instance, the equilibrium at position 2 lies at 5:95, due to the excess concentration of malonyl-CoA, driving the equilibrium to the right

malonyl-ACP suggesting an equiblibrium indicative of Scheme 5.2 has been generated (Fig. 5.14). This scenario is likely to be a more accurate representation of the native system, with the continous supply of –CoA derived extender units favouring the formation of malony-ACP.

### *5.2.9 Modelling and Sequence Analysis of PedC and PedD*

The data presented in Sects. 5.3.5–5.3.8 indicate that PedC is a hydrolytic enzyme with the ability to cleave acyl intermediates from ACPs, and PedD is a malonyl-specific acyltransferase domain. Despite such different catalytic roles, at the sequence level the domains are extremely similar. In an attempt to rationalise the difference in catalytic activities of these domains, homology models of PedC and PedD were constructed as described in Sect. 2.2.15.2.

The homology models were subsequently aligned with the crystal structure of FabD, the AT domain from the *E.coli* FAS, which was successfully crystallised with a malonate unit attached to the active site serine [29]. This step allowed the same

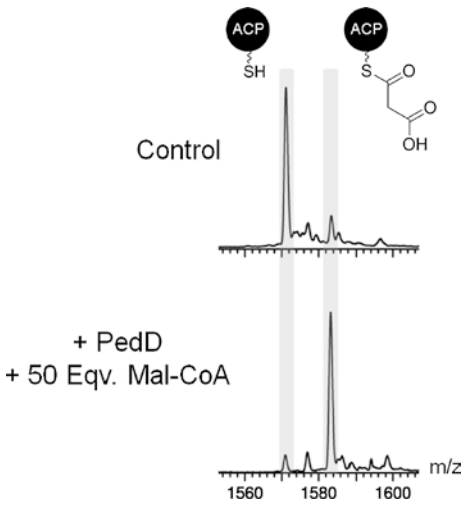



**Fig. 5.14** Stacked nESI-MS spectra of the $7^+$ charge state of *holo*-ACP before (*top*) and after (*bottom*) incubation with 1 eqv. PedD and 50 eqv. malonyl-CoA, yielding ~95 % malonyl-ACP from the reaction

malonate unit to be artifically inserted into the active site of the homology models. Analysis of the malonyl-complexed homology models for PedC and PedD revealed a strikingly obvious difference in the active site. The PedD and FabD structures both harbour an Arg residue at the back of the binding channel, which has been predicted to form a stabilising bidentate salt bridge with the carboxyl region of the malonyl unit [29–32]. In contrast, the same position in PedC is occupied by a Gln residue, which would be unable to participate in such an interaction (Fig. 5.15).

In order to investigate whether this is a common feature of AH domains, a sequence alignment of all the AT and AH domains from Fig. 5.2 was performed. All domains predicted to be ATs (AT1 clade) harbored an Arg residue at this position. However, all but one of the domains in the AT2 clade possessed a Gln residue in the equivalent position (Fig. 5.16).

The homology modelling and sequence analysis performed has highlighted a key difference in the active sites of these enzymes. Using this information a PedD(R97Q) mutant was constructed (as detailed in sect. "*BaeJ KS1 (C207A) and PedD(R97Q)*") to test the importance of the arginine residue in malonyl-binding, and test the exciting possibility that removal of this residue may switch the activity of PedD from and AT to a AH.

## *5.2.10 Examining the Activity of PedD(R97Q)*

### 5.2.10.1 Activity Towards Malonyl-CoA

PedD(R97Q) was incubated with increasing molar equivalents of malonyl-CoA to acertain the ability of the mutant to be malonylated. However, even with 8 eqvuivalents of malonyl-CoA, no malonyl-PedD(R97Q) was observed (Fig. 5.17).

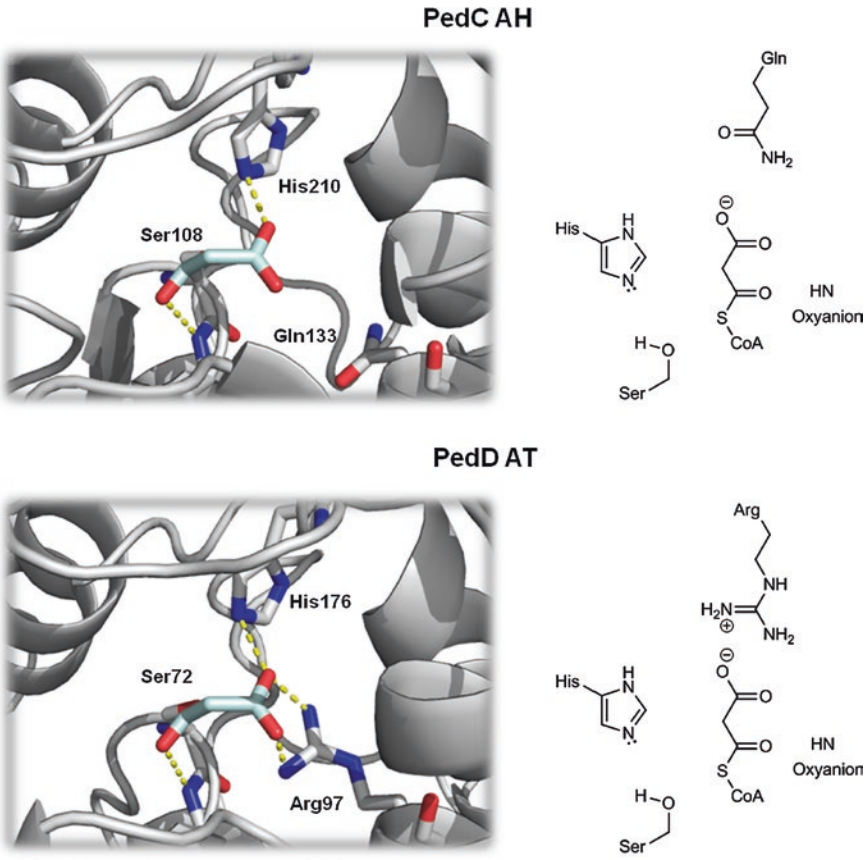



**Fig. 5.15** Homology models of PedC and PedD showing the arrangement of the active site residues around the malonate unit. The active site schematics to the right simplify the interactions identified between the protein and the malonate. The major difference between the two domains is the presence of an arginine residue in PedD at the rear of the pocket, which is hypothesised to participate in a bidentate salt bridge with the carboxyl region of the malonate. In contrast, PedC has a glutamine at this position which is unable to provide the cationic species for the salt bridge

This suggested that removal of the stabilising Arg residue (Fig. 5.15) has effectively made the AT domain catalytically inactive as a malonyl-transferase.

#### 5.2.10.2 Activity Towards Acyl-ACPs

The lack of activity towards malonyl-CoA, similar to that of PedC, suggested that PedD(R97Q) may also be able to catalyse hydrolysis of acyl-ACPs. Initially, PedD(R97Q) was incubated with acetyl-ACP at a 1:1 molar ratio to determine any hydrolytic activity of the mutant. After 5 min of incubation, ~45 % of the

AT1 Clade: Malonyl Transferases

```
ElsAAT1        YLDRLDTMP-KPDFVLGHSLGEYVALFAAGAFSFETGLKLVKQRAEIMGRVKN-GGMAAL 123
PedD           YRQHLQQGGGLPDFVAGHSLGEYNALESAGVFSFEDGLRLVQKRGDLMSQAPR-GAMAAI 112
BryPAT1        FLEKIELDSHKPSYVAGHSLGEYNALFAAGAFDFLTGLKLVKKRGLFMEEAPK-GAMAAI 124
BaeC           YLKKIQDNDIKPDYVAGHSLGEYNALFAAGAFDFITGLQLVRKRGELMSMATD-GKMAAV 127
BaeEAT1        YLKKTEETGLKPDFTAGHSLGEYNALYASGAFDFEEGLQLVKKRGELMSRAKG-GGMAAV 122
DfnAAT1        YLKKMKESGREPDYAAGHSLGEYNALFAAGCFDFETGLQLVKKRGELMSKAAP-GGMAAV 125
MlnAAT1        YLKGLKEGHTRPDAVAGHSLGEYTALFAAGVFNFETGLRLVAKRGELMSEASE-GGMAAV 124
SorOAT2        YYQLLAEGA-RPDYIAGHSLGEYNALHAAGAFDLLTGLRLVKRRGELMSQAKK-GGMAAV 126
BatJ           YYKAVEQGC-YPDFVAGHSLGEYNALLAAGVFDFLTGLRLVQQRGELMAQAKS-GSMGAV 129
DszDAT1        YLKRREEE-APPDFLAGHSLGEFSALFAAGVFDFETGLALVKKRGELMGDARG-GGMAAV 123
PsyH           YLQRMKETGGRPGYALGHSLGEYSALFAAEVFDFETGLRLVQKRGALMAEAKG-GGMAAV 229
RhiGAT2        FLNKQQQDRRQPDYLAGHSLGEYCALFAAGAFSFETGLKLVKKRGELMARATG-GSMAAV 125
OzmMAT2        WLRKLRDDPTPPDYLVGHSLGEYVALFAAGAFDFETGLRLVKRRGELMSGADG-GRMAAV 125
KirCIAT2       WLATVQEGGRLPDYLLGHSLGEFAALFAAGVYDFETGLRLVAERGRLMGQVTG-GTMAAV 123
LnmG           YLAR-DPGLPQPTLLAGHSLGEYGALFAAGCFDFATGVRLVRERGALMGRAQG-GGMLAV 122
```

AT2 Clade: Acyl Hydrolases

```
SorOAT1        ARALLEAGV-EPGMVLGASLGTFTAATIAGCLDVEDGLKAAIQHAKALEST---CEPGAM 131
BatH           AQALIRAGV-VPDITLGASAGSFAAAAVAGFIDVEDALAAVVMQAGIFEKH---CEPGGM 138
BaeD           AYALEQRGI-RPDYVIGASLGEYAAAAVSGVLSAEDALDCVLEQARIVTET---CRNGSM 137
PedC           AKSLLARGLPAPDFLIGASLGEFIAISLAGDTHVENILFNLIKQARLFDEY---CNAGAM 146
BryPAT2        TQVLIDKGI-KPDAFLGHSLGEYIAATVAGIIPLEDGLNLILTQAQLLEKH---CNTGKI 130
ElsB           ARTVSELLPAAPAYVLGSSLGETVAAAVSGAIPAEDMLEIVIRQAALVERS---CAEGGM 141
OzmMAT1        ADTLAGYGI-EPDLLLGASLGEFTACVLAGVVDRDTCLRLLIRQARAARG----APPGGM 131
KirCIAT1       ARMLRAHGF-EPELVLGASLGEVVAAAVAGIFDPEECLRSLLEQVALFEAE---CPRGGM 130
RhiGAT1        GKTLIEQQI-QPDYLLGTSLGELAAAALAETLPLSDAIRFVTRQGQLFHRRQSPATEGTM 135
```

**Fig. 5.16** Sequence alignment of all AT-like domains from the phylogenetic tree shown in Fig. 1.13. The Arg/Gln residue identified from the homology modelling is highlighted in each clade. The alignment clearly shows that domains denoted as acyltransferases harbor the Arg residue, whereas acylhydrolases replace this residue with a Gln. The alignment was performed using the ClustalW2 server

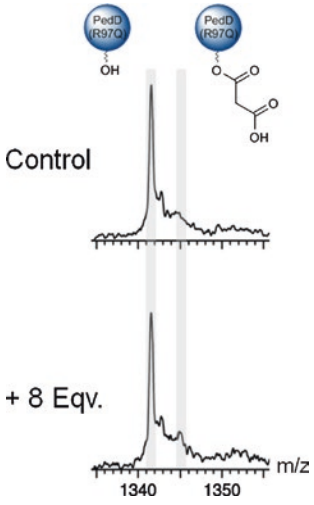



**Fig. 5.17** Stacked nESI-MS spectra of the $32^{+}$ charge state for and PedD(R97Q) following incubation with 8 eqv. of malonyl-CoA. No attachement of a malonyl unit was observed despite the high concentration of malonyl-CoA

acetyl-ACP was removed, with WT PedD exhibiting no activity (Fig. 5.18a). This result indicated that the point mutation may have yielded hydrolase activity; however, leaving the reaction for longer (20–30 min) resulted in no further deacylation. This suggested that PedD(R97Q) either has a population of inactive enzyme or, more likely, that following acyl transfer to PedD(R97Q) no hydrolysis of the acyl-enzyme intermediate occurred. Examination of the PedD(R97Q) MS signals indicated that the latter is correct, with ~45 % of acetyl-PedD(R97Q) observed, the same ratio observed on the ACP (Fig. 5.18b).

Although the Arg → Gln point mutation did not generate an enzyme with true acyl hydrolase activity, the ability of the domain to allow transfer of an acyl chain to the active site serine is different from the activity exhibited by WT PedD, which is unable to catalyse either transfer or hydrolysis of an acyl chain (Fig. 5.18). These results indicate that the point mutation has changed the activity of PedD towards that of PedC, with the factors dictating the ability catalyse hydrolysis yet to be identified.

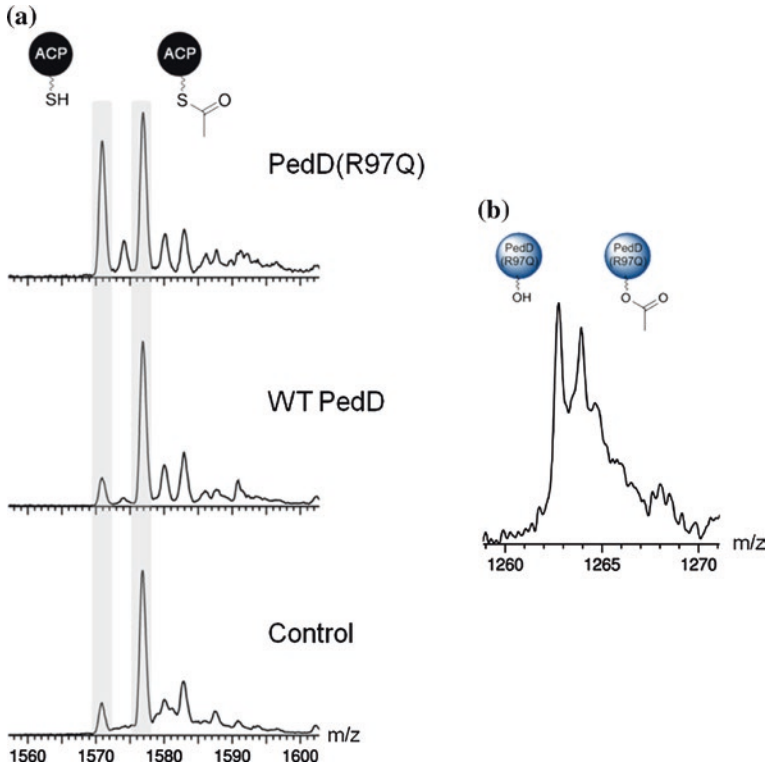



**Fig. 5.18** **a** Stacked nESI-MS spectra of the $7^+$ charge state of acetyl-ACP and deacetylated ACP following 5 min incubations with WT PedD and PedD(R97Q). PedD(R97Q) is able to deacylate acetyl-ACP, whereas WT PedD is not. **b** nESI-MS spectrum of the $34^+$ charge state of PedD(R97Q) following 5 min incubation with acetyl-ACP. The spectrum state shows clear attachment of an acetyl unit to the active site serine, corresponding to a mass shift of +42 Da

## 5.3 Conclusions

In summary, a simple method for the synthesis of a range of acyl-ACPs is presented, and successfully applied in assays to enable the study of KS, AH and AT domain specifcity as three examples for the broad range of potential applications in PKS studies.

The novel synthetic methodology reported allows a variety of acyl-ACPs to be accessed from simple SNAC thioesters via a transthioesterification reaction. The approach reported is rapid and less expensive than the commonly employed enzymatic loading of *apo*-ACPs with acyl-CoAs, and reduces the complexity of the synthesis. Akin to other reported methods for chemical loading of ACPs, this approach requires a *holo*-ACP deficient in any cysteine residues, as attachment to such residues can also occur resulting in complications when quantifying the amount of acyl-ACP.

Herein, acyl-ACPs are shown to be suitable substrate mimics for PKS enzymatic domains. Successful loading of PsyA KS1 with acetyl-ACP was observed via a 'de-acylation' assay, monitoring the disappearance of acetyl-ACP signals upon incubation with KS1. Due to the truer representation of the substrate, this method may be an improvement upon current approaches to test the acylation specificity of KS domains. Additionally, a full set of acyl-ACPs was used to probe the substrate specificity of an AH domain (PedC), fully characterising the hydrolytic specificity profile, with clear preference shown towards acetyl-ACP.

The use of these substrate probes also guided the successful mutation of a known AT domain, PedD, to exhibit AH-like activity. Investigation of the PedD(R97Q) mutant identified the key role of the active site Arg residue in binding and successful transfer of a malonyl unit from malonyl-CoA. Furthermore, PedD(R97Q) was able to de-acylate, but not hydrolyse, acyl-ACPs suggesting that other factors in the binding pocket govern the ability to catalyse hydrolysis.

## References

1. A.T. Keatinge-Clay, The structures of type I polyketide synthases. Nat. Prod. Rep. **29**, 1050–1073 (2012)
2. J. Crosby, M.P. Crump, The structural role of the carrier protein - active controller or passive carrier. Nat. Prod. Rep. **29**, 1111–1137 (2012)
3. J. Moldenhauer et al., The final steps of bacillaene biosynthesis in Bacillus amyloliquefaciens FZB42: Direct evidence for beta, gamma dehydration by a trans-acyltransferase polyketide synthase. Angew. Chem. Int. Ed. **49**, 1465–1467 (2010)
4. C.T. Calderone, W.E. Kowtoniuk, N.L. Kelleher, C.T. Walsh, P.C. Dorrestein, Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis. Proc. Natl. Acad. Sci. U.S.A. **103**, 8977–8982 (2006)
5. P. Pöplau, S. Frank, B.I. Morinaka, J. Piel, An Enzymatic domain for the formation of cyclic ethers in complex polyketides. Angew. Chem. Int. Ed. **50**, 13215–13218 (2013)

6. B. Kusebauch, B. Busch, K. Scherlach, M. Roth, C. Hertweck, Functionally distinct modules operate two consecutive alpha, beta → beta, gamma double-bond shifts in the rhizoxin polyketide assembly line. Angew. Chem. Int. Ed. **49**, 1460–1464 (2010)
7. R.J. Cox et al., Post-translational modification of heterologously expressed Streptomyces type II polyketide synthase acyl carrier proteins. FEBS Lett. **405**, 267–272 (1997)
8. J. Crosby et al., Acylation of *Streptomyces* type II polyketide synthase acyl carrier proteins. FEBS Lett. **433**, 132–138 (1998)
9. T.S. Hitchman, J. Crosby, K.J. Byrom, R.J. Cox, T.J. Simpson, Catalytic self-acylation of type II polyketide synthase acyl carrier proteins. Chem. Biol. **5**, 35–47 (1998)
10. C. Arthur et al., Synthesis and characterisation of acyl carrier protein bound polyketide analogues. ChemBioChem **3**, 253–257 (2002)
11. J.Q. Wu, K. Kinoshita, C. Khosla, D.E. Cane, 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**, 16301–16310 (2004)
12. N.A. Schnarr, A.Y. Chen, D.E. Cane, C. Khosla, Analysis of covalently bound polyketide intermediates on 6-deoxyerythronolide B synthase by tandem proteolysis-mass spectrometry. Biochemistry **44**, 11836–11842 (2005)
13. M. Jenner et al., Substrate specificity in ketosynthase domains from *trans*-AT polyketide synthases. Angew. Chem. Int. Ed. **52**, 1143–1147 (2013)
14. C. Kohlhaas et al., Amino acid-accepting ketosynthase domain from a *trans*-AT polyketide synthase exhibits high selectivity for predicted intermediate. Chem. Sci. **4**, 3212–3217 (2013)
15. P.C. Dorrestein et al., Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. Biochemistry **45**, 12756–12766 (2006)
16. K. Jensen et al., Polyketide proofreading by an acyltransferase-like enzyme. Chem. Biol. **19**, 329–339 (2012)
17. R.H. Cichewicz, F.A. Valeriote, P. Crews, Psymberin, a potent sponge-derived cytotoxin from *Psammocinia* distantly related to the pederin family. Org. Lett. **6**, 1951–1954 (2004)
18. K.M. Fisch et al., Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. Nat. Chem. Biol. **5**, 494–501 (2009)
19. L. Gu et al., GNAT-like strategy for polyketide chain initiation. Science **318**, 970–974 (2007)
20. J.C. Kwan, E.W. Schmidt, Cleaning up polyketide synthases. Chem. Biol. **19**, 309–311 (2012)
21. E. Yeh, R.M. Kohli, S.D. Bruner, C.T. Walsh, Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. ChemBioChem **5**, 1290–1293 (2004)
22. D. Schwarzer, H.D. Mootz, U. Linne, M.A. Marahiel, Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. Proc. Natl. Acad. Sci. U.S.A. **99**, 14083–14088 (2002)
23. J. Moldenhauer, X.H. Chen, R. Borriss, J. Piel, Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. Angew. Chem. Int. Ed. **46**, 8195–8197 (2007)
24. B. Kusebauch, B. Busch, K. Scherlach, M. Roth, C. Hertweck, Polyketide-chain branching by an enzymatic michael addition. Angew. Chem. Int. Ed. **48**, 5001–5004 (2009)
25. B. J. Dunn, C. Khosla, Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. J. R. Soc. Interface. vol 10 (2013)
26. B.J. Dunn, D.E. Cane, C. Khosla, Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. Biochemistry **52**, 1839–1841 (2013)
27. S.A. Bonnett et al., Acyl-CoA subunit selectivity in the pikromycin polyketide synthase pikaiv: steady-state kinetics and active-site occupancy analysis by FTICR-MS. Chem. Biol. **18**, 1075–1081 (2011)
28. J. Piel, A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. Proc. Natl. Acad. Sci. U.S.A. **99**, 14002–14007 (2002)

29. C. Oefner, H. Schulz, A. D'Arcy, G.E. Dale, Mapping the active site of Escherichia coli malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography. Acta Crystallogr. Sect. D-Biol. Crystallogr. **62**, 613–618 (2006)
30. Y.Y. Tang, C.Y. Kim, Mathews II, D.E. Cane, C. Khosla, The 2.7-angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase Proc. Natl. Acad. Sci. U.S.A.**103**, 11124–11129 (2006)
31. F.T. Wong, X. Jin, I.I. Mathews, D.E. Cane, C. Khosla, Structure and mechanism of the *trans*-acting acyltransferase from the disorazole synthase. Biochemistry **50**, 6539–6548 (2011)
32. A.T. Keatinge-Clay et al., Catalysis, specificity, and ACP docking site of *Streptomyces coelicolor* malonyl-CoA : ACP transacylase. Structure **11**, 147–154 (2003)

# Chapter 6
# Substrate Specificity of Ketosynthase Domains Part III: Elongation-Based Substrate Specificity

## 6.1 Introduction

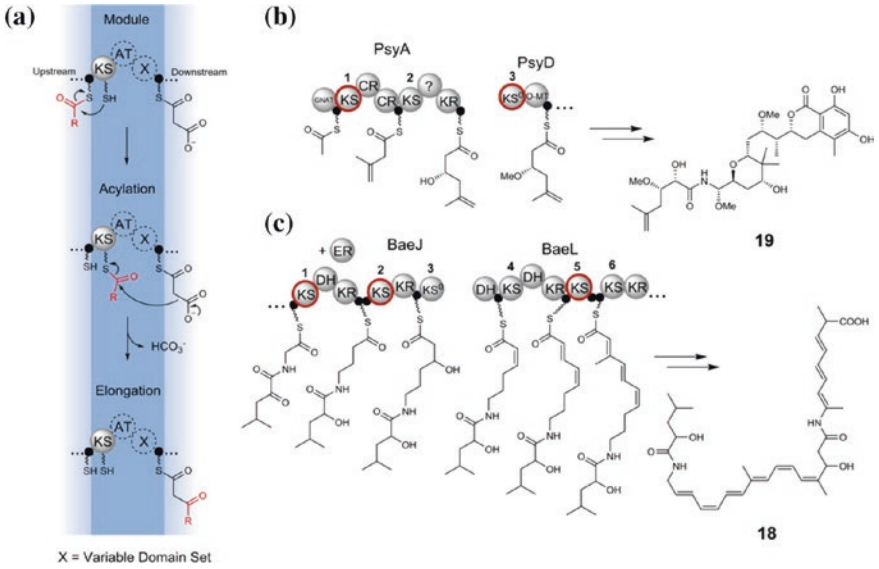
In order to maintain the structural integrity of a polyketide product it is important that at least some catalytic domains within a PKS possess specificity for particular intermediates. Since it is the KS domains that initially receive the product of the previous PKS module, through transfer to their active site Cys residue, it has been proposed that these domains fulfil a 'gatekeeper' role by preferentially accepting the correctly modified intermediate [1]. By controlling factors such as chain length and oxidation level at the β-position [2], this property helps to preserve product fidelity, and is an important consideration for potential engineering strategies [3].

Results described in Chaps. 3 and 4 highlight the specificity profiles of KS domains from *trans*-AT clusters with regard to the initial acylation step of catalysis. Although some specificity was observed, the tolerance profile was not as well defined as the phylogenetic grouping might suggest (Fig. 1.12). It was therefore postulated that further specificity might be afforded at the Claisen-type elongation step of the KS-catalysed reaction, where the acyl chain is transferred from the KS to ACP-bound malonate to form a new C-C bond (Scheme 6.1a) [4].

Previous studies on the elongation capabilities of *cis*-AT PKS modules revealed that, although unnatural substrates were apparently able to acylate the KS, no elongation was observed [5, 6]. For example, the data shown in Fig. 6.1 highlights four KS-AT-KR-TE modules: EryM5, EryM6, RifM7 and PicM3 and their known biosynthetic substrates. All of these modules were successfully acylated by the (2*S*, 3*R*)-[1-$^{14}$C]2-Methyl-3-hydroxypentanoic *N*-acetylcysteamine thioester, however, only EryM5 and EryM6 were able to catalyse the elongation reaction (Fig. 6.1). A possible explanation for this observation is that although aberrant substrates are able to successfully acylate the KS, the correct organisation of active site residues may be required for a successful condensation reaction, with the active site ordering thought to be promoted by acylation with the 'natural' substrate [5].



M. Jenner, *Using Mass Spectrometry for Biochemical Studies on Enzymatic Domains from Polyketide Synthases*, Springer Theses,
DOI 10.1007/978-3-319-32723-5_6




**Scheme 6.1** **a** The catalytic cycle of a KS domain. Acylation by an upstream ACP loads the KS with an acyl intermediate, followed by the elongating Claisen condensation between the acyl intermediate and a malonyl unit. In *cis*-AT PKSs the AT domain is integral to the PKS, whereas in *trans*-AT systems the activity is supplied by a free-standing AT domain. The X represents optional reductive domains which can alter the functionality at the β-keto position, producing β-hydroxyl, olefinic and fully saturated intermediates. **b**, **c** Partial proposed biosynthetic schemes for psymberin and bacillaene, with the KS domains used in this study highlighted in *red*

Although, given the multidomain structure of the modules used and the nature of the assay, it is impossible to say with confidence where the stalling is occurring. If the KS specificity hypothesis is correct, it raises the question as to how a PKS deals with acylation of KS domains by the 'incorrect' substrate, if indeed such an event happens during biosynthesis.

In this chapter, new insights into substrate specificity during the elongation step of KS-catalysed polyketide biosynthesis are reported. Using a range of KS domains from the psymberin (Psy) and bacillaene (Bae) PKSs, each with different predicted acyl intermediates (Scheme 6.1a, b), together with MS-based methodology, the substrate tolerance at the elongation step is shown to be more demanding than the preceding acylation. A mechanism, based on reversible KS acylation, is proposed to rationalise this phenomenon, and is supported by experimental evidence. The rationally-designed mutant of BaeL KS5 described in Chap. 3, which is able to accept a β-Me branched acyl unit, is also shown to elongate this bulky acyl group in addition to 5- and 6-membered rings, demonstrating scope for engineering new polyketides.

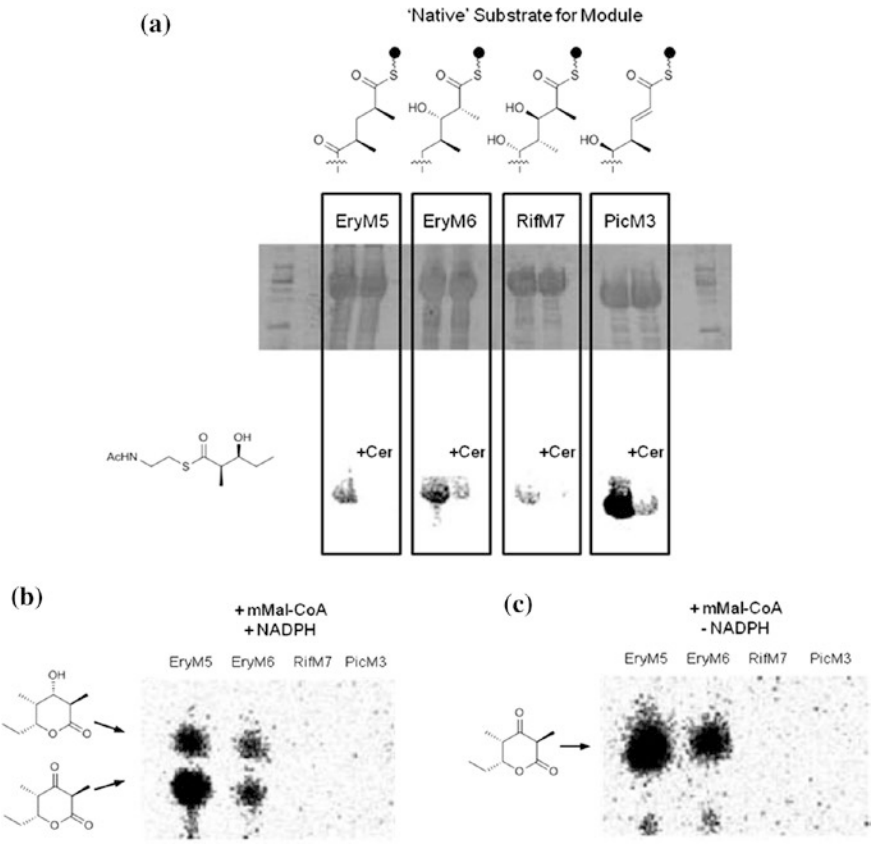



**Fig. 6.1** Elongation studies on *cis*-AT modules. **a** SDS-PAGE gel of module X-TE labelled with $^{14}$C N-acteylcysteamine substrate and subsequent radio image of module X-TE labelling. KS inhibitor, cerulenin, clearly prevents acylation. **b** Radio-TLC for in vitro elongation assay conducted in the presence of NADPH. **c** Radio-TLC of in vitro elongation assay in the absence of NADPH. Adapted from [5]

## 6.2 Results and Discussion

### *6.2.1 Synthesis of Malonyl-ACP*

In order to study the specificity of KS-catalysed chain elongation a new assay was required. This was achieved by incubation of a pre-acylated recombinant KS domain with one equivalent of malonyl-ACP. In principle, the β-keto elongated product would be formed on the PPant moiety of the ACP, and could be accurately detected by ESI-MS. Malonylation of the ACP PPant thiol was achieved by incubation of *holo*-PsyA-ACP3(Δ37, 38) with 50 equivalents of malonyl-CoA as detailed in Sect. 2.2.9.1. Reactions were monitored by ESI-MS for the presence

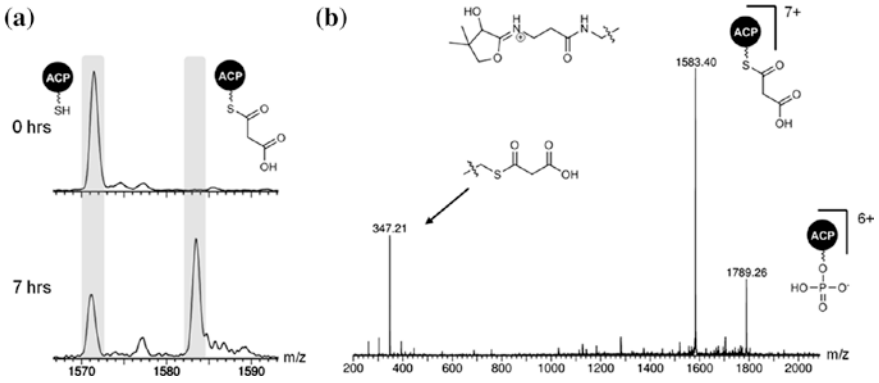



**Fig. 6.2** **a** nESI mass spectra (ESI-MS) of the $7^+$ charge state for *holo*-ACP following incubation with malonyl-CoA. **b** $MS^2$ spectrum of the isolated and activated $7^+$ charge state of malonyl-ACP yielding a species at *m/z* 347.21corresponding to malonyl-PPant. Decomposition of malonyl-PPant to acetyl-PPant (*m/z* 303.21) and PPant-SH (*m/z* 261.21) following activation of malonyl-ACP

of the malonyl-ACP product (Fig. 6.2a). Generally, malonylation reactions were complete after 7 h at 25 °C, yielding ~70–75 % malonyl-ACP. The PPant ejection assay was employed to check that the malonyl moiety was attached to the PPant thiol, resulting in predominantly an ion at *m/z* 347.2 corresponding to malonyl-PPant (Fig. 6.2b). Minor breakdown products of the activation process were also observed, including acetyl-PPant and PPant-SH. The former arising from decarboxylation of the malonyl unit, and the latter from loss of either acetyl or malonyl.

### *6.2.2 Synthesis of Alkyl-ACP*

An alkyl-ACP was produced as an internal standard for the elongation reaction, enabling quantification of elongated material. It was important that the internal standard was structurally robust upon collision induced activation and produced yields of >95 % to ensure accurate quantification. Attachment of an alkyl moiety to the PPant chain, forming a thioether, would provide the necessary stability upon activation, and the alkylation reaction is irreversible allowing higher yields. Using iodoacetamide as the alkylating agent would generate a PPant ejection ion at *m/z* 318.2, which would not overlap in mass with any potential elongated products. Alkylation of the ACP PPant thiol was achieved by incubation of *holo*-PsyA-ACP3(Δ37, 38) with 20 equivalents of iodoacetamide, in the absence of light, as detailed in Sect. 2.2.9.2. Reactions were monitored by ESI-MS for the presence of the alkyl-ACP product (Fig. 6.3a). Typical yields were >95 % alkyl-ACP and during MS/MS analysis the expected *m/z* 318.2 ion was observed (Fig. 6.3b).

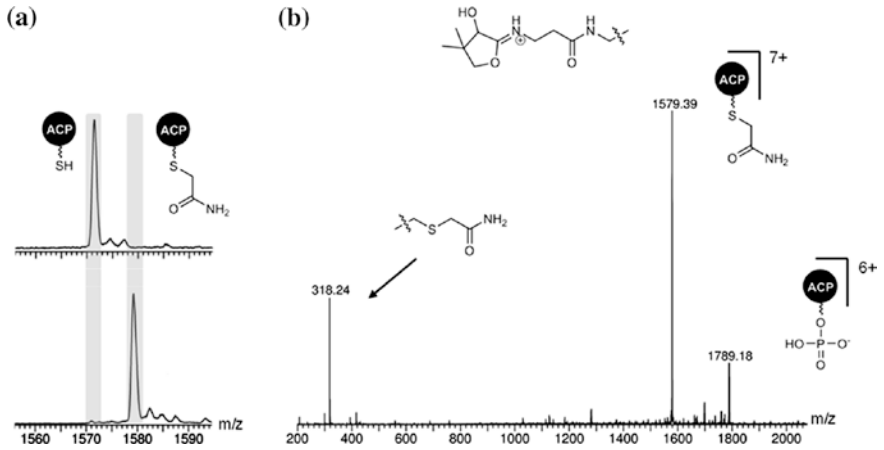



**Fig. 6.3** **a** nESI mass spectra (ESI-MS) of the $7^+$ charge state for *holo*-ACP following incubation with iodoacetamide. **b** MS/MS spectrum of the isolated and activated $7^+$ charge state of malonyl-ACP yielding a species at *m/z* 318.24 corresponding to alkyl-PPant

### *6.2.3 MS/MS Elongation Assay: Proof of Principle*

With the ability to reliably synthesise malonyl-ACP and an internal standard established, initial proof of principle elongation studies were performed. The proposed assay, outlined in Scheme 2.6, consisted of incubation of pre-acylated KS with one equivalent of malonyl-ACP. In theory, the elongated product would be formed on the PPant moiety of the ACP, and could be accurately detected by the PPant ejection assay. In order to quantify the amount of elongated material, the internal standard was added to a final concentration of 10 μM before desalting the reaction into 80:20 MeCN:$H_2O$, 0.1 % TFA. During MS analysis, the $7^+$ charge state of the ACP was isolated using a wide *m/z* window encompassing all acyl/alkyl variations of the ACP within the reaction. Activation by CID ejects the acyl/alkyl moieties present in the reaction, allowing accurate identification of PPant species.

Using previous knowledge obtained from SNAC acylation data (see Fig. 3.5), PsyA KS1 was thought to be a good model KS to test this assay on as it appeared to acylate well with the 'natural' acetyl-SNAC substrate (**20**). Furthermore, the elongated product for this acyl-KS, acetoacetyl-ACP, has been previously observed using SNAC-based acylation of ACPs, and the PPant ejection conditions yields an ion at *m/z* 345.2. Upon incubation of malonyl-ACP with acetyl-KS1 and subsequent MS analysis under the aforementioned conditions, the elongated acetoacetyl-PPant species was successfully identified, with an ion at *m/z* 345.2 (Fig. 6.4). Due to the wide window of activation, a range of other PPant species are observed in the MS/MS spectrum. These included; un-modified PPant (*m/z* 261.2), acetyl-PPant (*m/z* 303.2), malonyl-PPant (*m/z* 347.2) and the disulphide-linked *N*-acetylcysteamine-PPant (*m/z* 378.2).

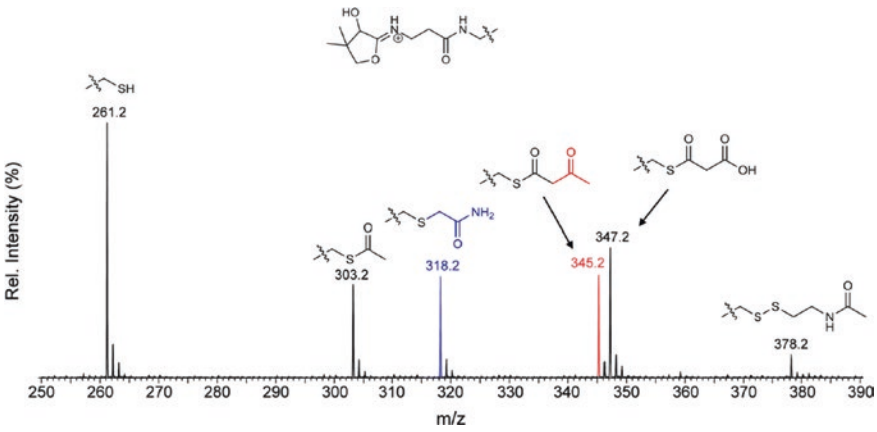



**Fig. 6.4** MS/MS spectrum of the activated 7 + PsyA ACP3 species following an acetyl elongation reaction. The production of acetoacetyl-ACP is shown in red (PPant ion *m/z* 345.2). Additional ions observed in the ejection assay are: un-modified PPant (*m/z* 261.2), acetyl-PPant (*m/z* 303.2), malonyl-PPant (*m/z* 347.2) and the disulphide-linked *N*-acetylcysteamine-PPant (*m/z* 378.2). Internal standard shown in blue (PPant ion *m/z* 318.2) was added after the elongation reaction at a final concentration of 10 μM

The acetyl-PPant unit is likely to have arisen from the excess acetyl-SNAC (**20**) remaining in the reaction mixture from the KS acylation reaction, acylating any free PPant thiol. It was proposed that this species may also be arising from activation of the elongated acetoacetyl-PPant, however, the MS activation conditions used in the assay were optimised using acetoacetyl-ACP ensuring that only the *m/z* 345.2 is ejected. The disulphide-linked *N*-acetylcysteamine-PPant species is a result of a relatively minor side reaction between the free thiols of *N*-acetylcysteamine and the PPant arm.

In order to confirm that the production of the acetoacetyl-PPant was a result of PsyA KS1 catalysis, a control reaction was performed under the same reaction conditions in the absence of PsyA KS1. The MS analysis of this reaction yielded no peak at *m/z* 345, demonstrating that the production of acetoacetyl-ACP is solely due to KS-catalysed condensation of the malonyl-ACP with the acetyl unit (Fig. 6.5).

### *6.2.4 Calibration of Internal Standard to Acyl-ACP*

Quantification of elongated material in reactions required the addition of alkyl-ACP internal standard, which generates an ion at *m/z* 318.2 following PPant ejection. In order to ascertain whether the alkyl-ACP could be used as an accurate measure of the concentration of elongated material, a calibration between alkyl- and acetyl-ACP was performed. The calibration was conducted over a range of concentrations

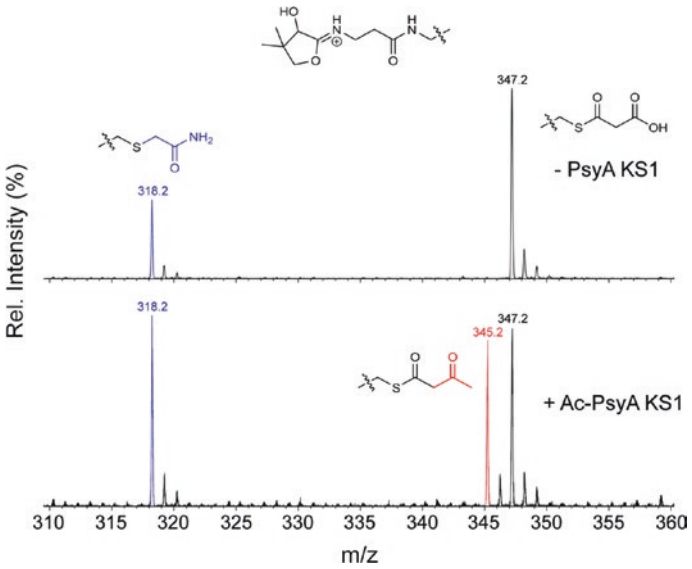



**Fig. 6.5** MS/MS spectra of the activated 7 + PsyA ACP3 species following acetyl elongation reactions in the absence (*top*), and presence (*bottom*) of PsyA KS1. The production of acetoacetyl-ACP (PPant ion *m/z* 345.2) can be solely attributed to the presence of the KS domain. Internal standard (PPant ion *m/z* 318.2) was added after the elongation reaction at a final concentration of 10 μM

of acetyl-ACP (0.5–20 μM) with the alkyl-ACP fixed at 20 μM. The ratio of acetyl- to alkyl-ACP was then calculated to determine the concentration of acetyl-ACP reported by the presence of 20 μM alkyl-ACP. A calibration plot was constructed using the 'theoretical' and 'measured' concentrations of acetyl-ACP.

The calibration yielded a linear, 1:1 relationship over the range of 0.5–20 μM, between theoretical and measured values of the acetyl-ACP, using alkyl-ACP as an internal standard (Fig. 6.6). This result meant that direct quantification of elongated material in the reaction was possible by addition of alkyl-ACP as an internal standard.

### *6.2.5 Assessing Beta-Keto-ACP Stability*

The product of KS-catalysed elongation is a β-keto-ACP thioester, which due to its intrinsic instability has the potential to readily hydrolyse off the ACP-PPant chain. Although the acetoacetyl-ACP was successfully detected in Sect. 6.2.3, the reaction was allowed to proceed over a 16 h period, which may allow hydrolysis of the acetoacetyl-ACP to occur resulting in an underestimation of elongated material. Therefore, to assess the stability of the β-keto-ACP species over the 16 h time

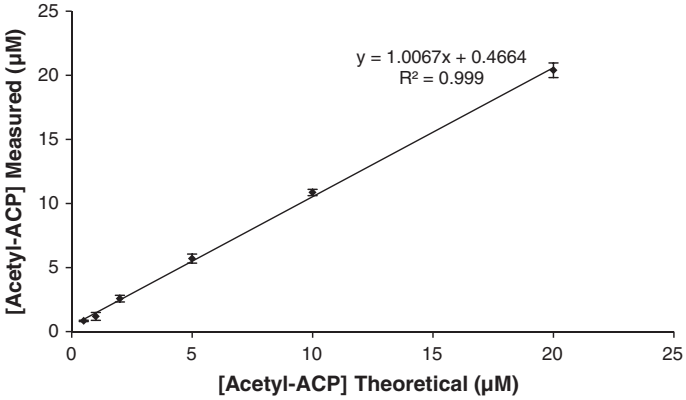



**Fig. 6.6** Calibration of theoretical [Acetyl-ACP] against measured [Acetyl-ACP] from PPant ejection assay over a 0.5–20 μM range

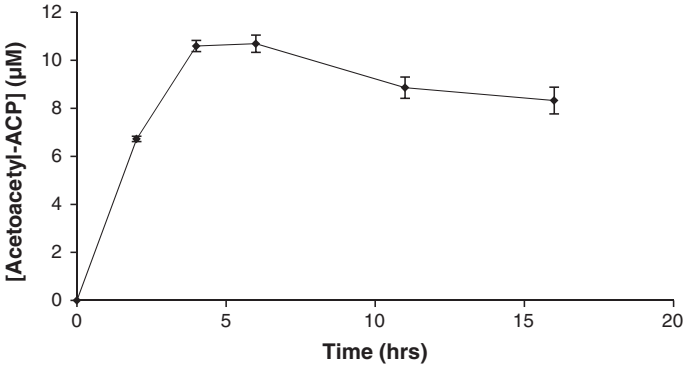



**Fig. 6.7** Time-course assessment of β-keto-ACP stability during a 16 h period. The concentration of acetoacetyl-ACP was quantified using the alkyl-ACP internal standard

period, an investigative time-course assessment of the acetyl-PsyA KS1 elongation was conducted, analysing aliquots of the reaction at regular time intervals. The plot shown in Fig. 6.7 shows a clear peak in acetoacetyl-ACP concentration after 6 h, followed by a gradual decrease, presumably due to hydrolysis. By the endpoint of the reaction the difference between the 6 h point and the 16 h point was ~20 %. Although measurement the 16 h time-point may result in a slight underestimation of the total elongated product, for acyl chains which may not be as good substrates for the KS, more time may be required for detection of elongated products. Therefore, in order to give less potent substrates the best chance of detection, the 16 h assay time was used.

### *6.2.6 KS Elongation Specificity Profiles*

#### 6.2.6.1 Psy KS-Catalysed Elongation of Acyl Chains

PsyA KS1 is located in phylogenetic clade VI (see Fig. 1.12) with other acetyl-accepting KS domains, and was hypothesised to be highly specific, due to its position at the start of the biosynthetic cluster (Scheme 6.1b). However, previous work in Chap. 3 has shown that functionally, PsyA KS1 is capable of accepting a range of substrates at the acylation stage. This was tentatively attributed to the lack of evolutionary pressure upon the domain to confer specificity towards an acetyl unit, due to the presence of a highly specific upstream GNAT domain, which would supply only acetyl units to KS1 [7]. Using the elongation assay described, PsyA KS1 showed a distinct preference towards the acetyl unit, with the longer and bulkier carbon chains clearly disfavoured (Fig. 6.8a). There was no observable elongation for the β-methyl substrate despite reasonable acylation of the KS observed previously, or for the oxazole and 2-methyl-tetrahydropyran substrates. It is worth noting that initial studies using β-OH SNACs also yielded no observable elongated material. This may have been as a result of KS specificity, however the hypothetical β-keto-δ-hydroxy-ACP thioester product is known to be unstable with respect to cyclisation, which breaks the thioester bond and releases a dihydropyranone ring from the ACP, rendering it undetectable in this assay [5]. Overall, however, these results suggest that, not only does PsyA KS1 impose specificity at the elongation step of its catalytic action, but that the specificity is in excellent agreement with sequence-based phylogenetic predictions [8, 9].

The non-elongating domain, PsyD $KS3^0$, was employed as an additional control for these experiments. Non-elongating KS domains are a common feature of *trans*-AT PKSs, and lack an essential active-site His residue, which is required for decarboxylation of malonate [10]. PsyD $KS3^0$ was therefore expected to yield no product during an elongation assay. Incubation of malonyl-ACP with PsyD $KS3^0$ acylated with a range of groups generated no observable β-ketoacyl products (**48**–**54**), despite effective acylation of the KS domain.

#### 6.2.6.2 Bae KS-Catalysed Elongation of Acyl Chains

In contrast to PsyA KS1, BaeJ KS2 and BaeL KS5 are predicted to accept and elongate much longer acyl chains. In terms of phylogenetic grouping, BaeJ KS2 is located in clade V (see Fig. 1.12) with other similar KS domains predicted to accept either olefinic or fully saturated, unbranched acyl chains. BaeL KS5 is located in clade IX (see Fig. 1.12), and is believed to accept exclusively unbranched, α, β-unsaturated intermediates. Acylation studies, carried out on BaeJ KS2 and BaeL KS5 previously (see Chap. 3), indicated that unbranched acyl chains were clearly favoured substrates [11]. In the case of BaeL KS5, a bulky methionine residue in the binding pocket was found to prevent acylation by

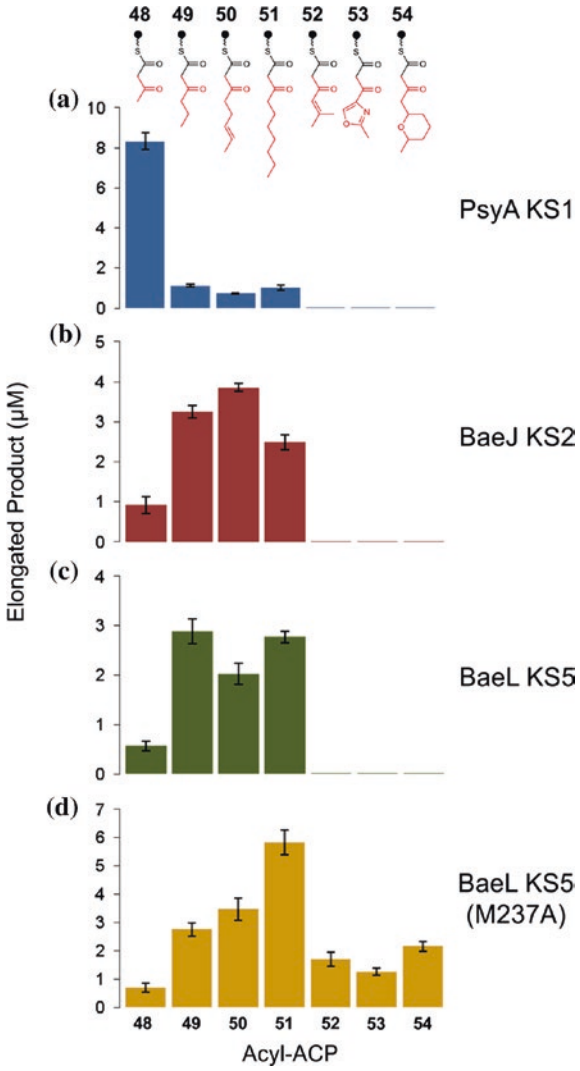



**Fig. 6.8** Elongation specificity profiles for Psy and Bae KS domains, generating β-ketoacyl products (**48**–**54**). PsyA KS1 exhibits specificity for its predicted acetyl substrate, whilst BaeJ KS2 and BaeL KS5 prefer longer chains, as predicted by phylogenetic analysis. The rationally designed BaeL KS5 mutant is able to elongate β-methyl branched chains in contrast to the WT enzyme. MS/MS spectra for each substrate and KS domain can be found in the appendix

branched substrates. Recently, an X-ray crystal structure of the *Bacillus subtilis* homologue of BaeJ KS2 was published with the native intermediate bound in the active site. This structure indicated favourable interactions between residues in the active site and the acyl chain, and provided a valuable insight into the potential residues involved in dictating substrate specificity [11].

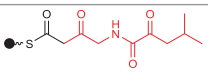
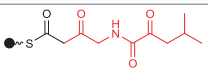
Elongation assays conducted on BaeJ KS2 revealed a remarkably different substrate profile from that seen for PsyA KS1, with the longer acyl chains preferentially elongated to form their β-ketoacyl equivalents **49–51** (Fig. 6.8b). In contrast, the short acetyl chain—being the preferred substrate of PsyA KS1—was

elongated only poorly by BaeJ KS2, to yield little of the acetoacetyl-ACP (**48**). The X-ray structure of BaeJ KS2, acylated with its natural substrate, shows that the entire 10-atom chain is accommodated within a relatively hydrophobic binding pocket. Presumably, favourable interactions between longer carbon-chains and this pocket account for the observed substrate preference of BaeJ KS2. A similar elongation profile was produced by BaeL KS5, with clear preference towards longer acyl chains (Fig. 6.8c). In agreement with previous acylation data, no elongation was observed for a β-Me substrate, as evidenced by the absence of **52**. This was expected as, without acylation of the KS domain, no elongation can occur. Also, similar to BaeJ KS2 no elongation products were observed for the oxazole and 2-methyl-tetrahydropyran substrates, which would yield elongated products **53** and **54**. Additionally, a crotonyl-ACP analogue of the natural, unbranched α, β-unsaturated intermediate of BaeL KS5 was tested, however no elongation was observed due to extensive 1, 4-conjugate addition to the SNAC by the KS domain. Overall, these data demonstrated that both BaeJ KS2 and BaeL KS5 possessed a preference for substrates bearing longer acyl chains, and that both KSs were unable to catalyse elongation of β-branched chains, as predicted by phylogenetic analysis.

Previous work in Chap. 3 showed that the active-site mutant BaeL KS5(M237A) was, in contrast to the wild type (WT), able to accept a β-methyl branched acyl chain at the acylation stage. Whether this KS variant had the ability to process the bulkier substrate fully, yielding elongated product **52**, remained an open question. The elongation profile of BaeL KS5(M237A) revealed a similar trend to that of WT BaeL KS5 with regard to the straight chain acyl units, and preferential elongation of longer acyl chains was observed. However, the mutant was also found to be competent at elongating the β-methyl branched chain, as evidenced by the detection of β-ketoacyl-ACP **52** in the assay (Fig. 6.8d). Furthermore, the elongated products **53** and **54**, corresponding to the oxazole and 2-methyl-tetrahydropyran substrates respectively, were also detected at similar levels to that of **52**. This suggested that the single Met → Ala point mutation in the active site of KS5 not only permitted acylation by the bulkier acyl chains, as observed in Chapter 3, but that it also permitted the elongation step to be catalysed for these substrates. The successful, functional modification of BaeL KS5 substrate specificity highlights the attractive possibility of creating engineered PKSs harbouring KS domains with 'non-native' specificity profiles.

As a further point of investigation, BaeJ KS1 was also submitted to the elongation assay. Phylogenetically, BaeJ KS1 is located in the relatively small but exclusive clade XVI, with other KS domains occurring immediately downstream of an NRPS module, and therefore must accept and elongate 2-amidoacetyl intermediates (see Figs. 1.12 and 1.16). The data presented in Chap. 4 clearly shows the selectivity of BaeJ KS1 towards its full-length, glycine-derived natural substrate (**27**) over straight chain acyl units. However, in order to observe the subtle substrate selectivity of BaeJ KS1 low SNAC concentrations were used (0.5 mM), and furthermore, data from preliminary experiments on BaeJ KS1 revealed that, at high concentrations of SNAC (2–4 mM), successful acylation of BaeJ KS1

**Table 6.1** BaeJ KS1 elongation specificity profile for β-ketoacyl products **48–56**

| | Acyl-ACP | Elongated Product (μM) |
|---|---|---|
| | **48–54** | *N.D.* |
| **55** |  | 4.1 ± 0.23 |
| **56** |  | 0.6 ± 0.10 |

was observed. Therefore, under these forcing conditions it was possible to obtain an acylated form of BaeJ KS1 for all the substrates used in this assay thus far, in addition to the full-length glycine and alanine derived substrates, **27** and **28**. The elongation profile of BaeJ KS1 revealed categorical preference towards 2-amidoacetyl substrates, as evidenced by the detection of **55** and **56**, with preference shown towards the glycine substrate. There was no observable elongated product for any acyl chains which did not contain the 2-amidoacetyl functionality (Table 6.1).

These results indicate that BaeJ KS1 harbours selectivity towards substrates containing 2-amidoacetyl functionality at both the acylation and elongation steps, with the selectivity imposed at the condensation stage being more stringent. These observations also show good agreement with the phylogenetic positioning of BaeJ KS1, where the KS domains within clade XVI are predicted to exclusively elongate intermediates containing 2-amidoacetyl functionality. The MS/MS spectra for these experiments are shown in the Appendix.

### *6.2.7 Monitoring Acyl Transfer Within PsyA ACP1-KS1*

The elongation assay performed on PsyA KS1 demonstrated a clear preference for the natural acetyl substrate over longer acyl chains. This finding contrasted with the rather broad substrate profile previously seen by the same domain in the acylation step. No discernible difference between the rate of acylation of PsyA KS1 by either acetyl or butyryl-SNAC was seen, for example, whilst the degree of elongation of these two thioesters differed by a factor of eight (Fig. 6.8a). To investigate whether the apparent lack of specificity during acylation was due to the use of simple SNAC substrate analogues, in an artificially bimolecular reaction, the PsyA ACP1-KS1 didomain was cloned to produce a more ‘native’ representation of the acyl chain loading. This protein was successfully expressed in an *apo*-form, allowing direct loading of the ACP with acyl-PPant chains using the PPant transferase Svp from *Streptomyces verticillus,* to probe the KS acyl transfer step in a native environment.

#### 6.2.7.1 Purification of PsyA ACP1-KS1 and Svp PPTase

The PsyA ACP1-KS1 didomain was amplified from a fosmid containing PsyA DNA, and subsequently cloned into the pET28b expression vector, as described in section "Cloning of PsyA ACP1-KS1 Didomain". The didomain was expressed and purified as a His-Tag fusion protein as described in Sect. 2.2.1. The *Streptomyces verticillus* Svp PPTase expression plasmid was kindly provided by Dr. Anna Vagstad. The PPTase was expressed and purified by Dr. José Afonso according to a previously published protocol [12]. Post-purification, the didomain and PPTase were analysed by SDS-PAGE and by mass spectrometry. An SDS-PAGE gel and mass spectra of purified PsyA ACP1-KS1 and Svp PPTase are shown in Fig. 6.9.

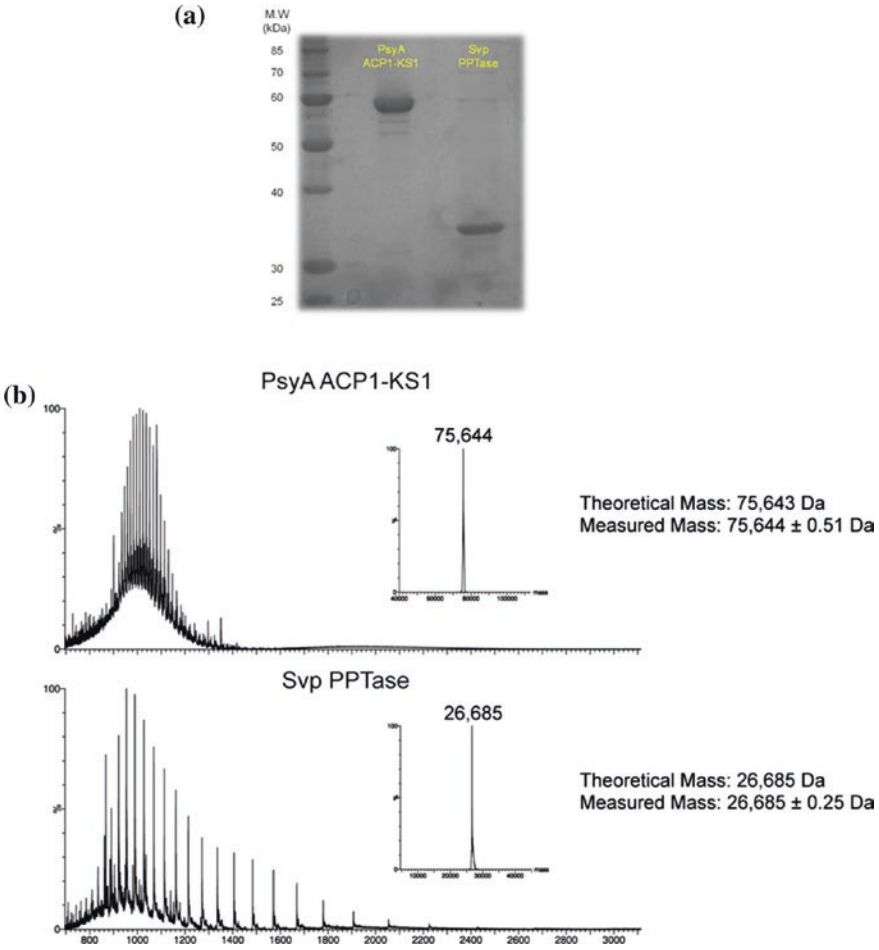



**Fig. 6.9** **a** 12 % SDS-PAGE gel of PsyA ACP1-KS1 and Svp PPTase. **b** nESI mass spectra of PsyA ACP1-KS1 and Svp PPTase sprayed from 80:20 MeCN:$H_2O$ 0.1 % TFA. Theoretical and measured molecular weights are annotated on the spectra

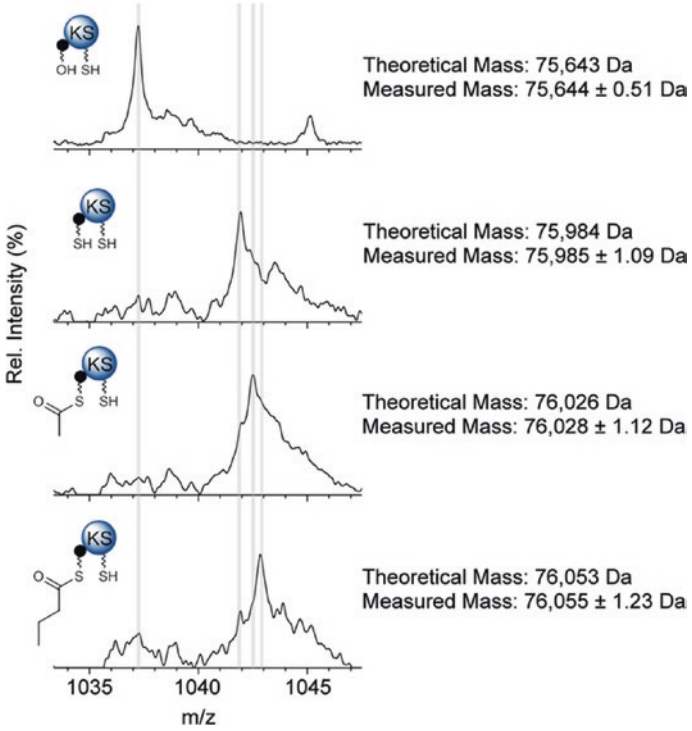



**Fig. 6.10** Stacked nESI-MS spectra of the $73^{+}$ charge state of PsyA ACP1-KS1 following PPant, acetyl-PPant and butyryl-PPant loading using the Svp PPTase. In all cases, complete conversion to the pantetheinylated state of PsyA ACP1-KS1

### 6.2.7.2 Examining Acyl Transfer in PsyA ACP1-KS1

The PPant transferase Svp from *Streptomyces verticillus* was successfully employed to transfer the PPant chain from both acetyl- and butyryl-CoA to the conserved serine of ACP1, resulting in the required *holo*-ACP1-KS1 didomain with an acylated PPant arm. The MS/MS fragmentation of both the acetyl- and butyryl-didomains revealed PPant ejection ions consistent with acyl transfer from ACP1 to KS1 (Fig. 6.10). Examination of the relative intensities of the fragment ions at *m/z* 303.2 (acetyl), *m/z* 331.2 (butyryl) and *m/z* 261.2 (unacylated) indicated that both acyl chains transferred to the KS active site to a similar degree (Fig. 6.11).

To control against acyl hydrolysis from the PPant chain, which could result in detection of the unacylated ion, the KS1 active site Cys residue was first alkylated with iodoacetamide to render it inactive. Conducted under mild conditions, this procedure cleanly alkylated the didomain once (Fig. 6.12a). Following acetyl-PPant loading, MS/MS analysis of the didomain showed a fragment ion at *m/z*

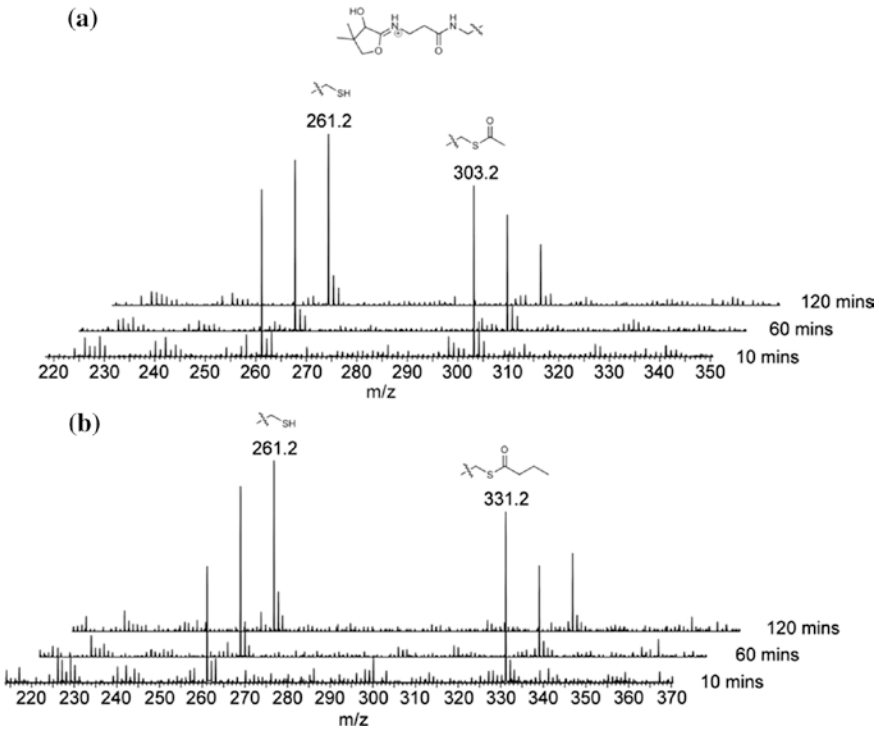



**Fig. 6.11** MS/MS spectra of the activated 73 + PsyA ACP1-KS1 species following loading with **a** Acetyl-CoA and **b** Butyryl-CoA. The relative intensities of the fragment ions at *m/z* 303.2 (acetyl), *m/z* 331.2 (butyryl) and *m/z* 261.2 (unacylated) indicated that both acyl chains transferred to a similar degree over the 120 min experiment

303.2 exclusively, demonstrating the intrinsic stability of the acetyl thioester in the absence of KS activity, and confirming that the fragment ion at *m/z* 261.2 seen above was indeed due to intramolecular transfer to the KS1 domain (Fig. 6.12b). Moreover, this control ruled out non-specific acyl transfer from the PPant arm to any Cys residues within PsyA ACP1-KS1 following denaturation and MS analysis of the didomain. Thus, the substrate specificity of KS1 during initial acylation is lower than that seen at elongation.

### *6.2.8 Reversible Transfer of Acyl Chains in PsyA ACP1-KS1*

The mechanism of KS-catalysed polyketide chain elongation is comprised of two distinct steps: first acyl transfer to the KS active site by the donating ACP of the preceding PKS module, and, second, the decaboxylative Claisen-type elongation

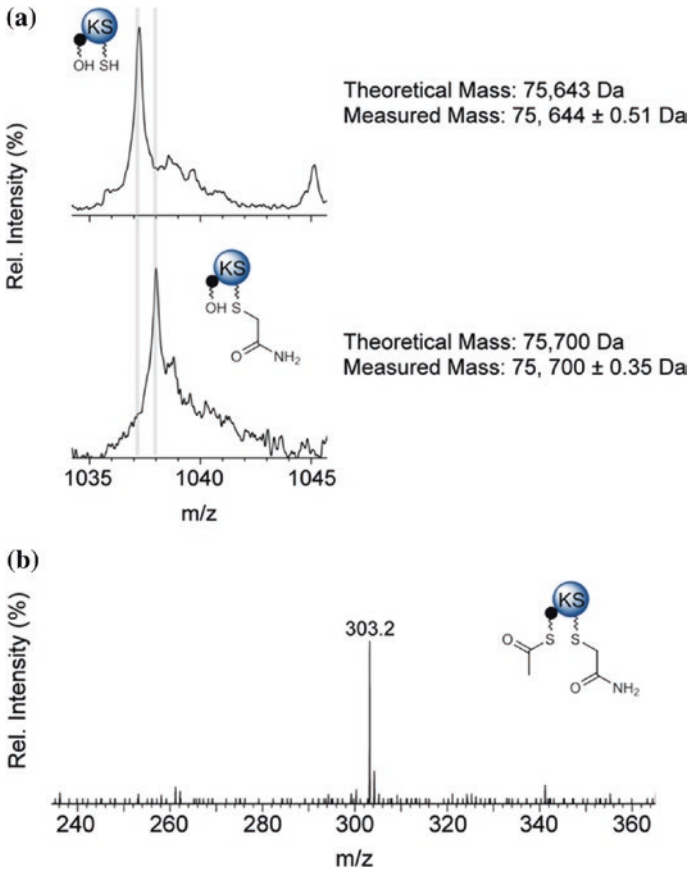
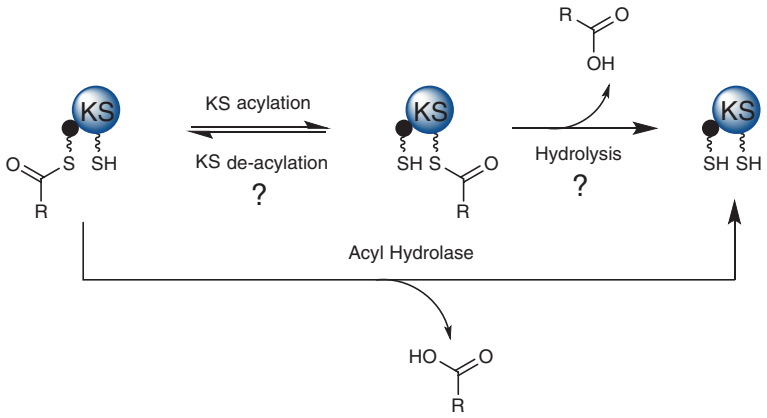



**Fig. 6.12** **a** Stacked nESI-MS spectra of the $73^{+}$ charge state of *apo*-PsyA ACP1-KS1 following alkylation with iodoacetamide, yielding a mass increase of +57 Da, good agreement with theoretical mass shift for this modification. **b** $MS^2$ spectrum of the activated 73 + Alkyl-PsyA ACP1-KS1 species following loading with acetyl-CoA. The fragment ion at *m/z* 303.2 was exclusively observed, demonstrating the intrinsic stability of the acetyl thioester in the absence of KS activity, and confirming that the fragment ion at *m/z* 261.2 seen in acyl-transfer experiments was indeed due to intramolecular transfer to KS1

reaction itself. The data presented here convincingly demonstrate that the KS domains tested confer higher specificity at the elongation step than they do during initial acylation of the enzyme. Some indication of this phenomenon has been reported in *cis*-AT systems, where an unnatural substrate was apparently able to acylate the KS domain of PicM3, but no elongated product was observed to be released by a downstream thioesterase [5]. It was suggested that elongation was unable to proceed due to lack of appropriate amino acid residue organisation in the active site, resulting in the intermediate being effectively ‘stalled’ on the KS domain.




**Scheme 6.2** Hypothetical movement of an acyl chain between a KS and the upstream ACP. Once the acyl-KS is formed, the acyl moiety can potentially be removed by de-acylation by the downstream ACP and subsequent hydrolysis by an acyl hydrolase. Alternatively, direct hydrolysis from the KS active site can occur

The results presented thus far show that substrate specificity at the second elongation step exceeds that of acyl loading also highlight the potential problem of KS enzymatic stalling. In order to restore enzymatic activity either the acyl chain must hydrolyse off the active site Cys, or the upstream ACP must de-acylate the KS in a reversal of the initial acylation step. Once on the upstream ACP, the acyl chain may hydrolyse off the PPant arm, possibly catalysed by an acyl hydrolase (AH) domain (Scheme 6.2) [13]. In theory, both mechanisms are plausible, but there is a lack of any clear evidence in the literature for reversibility of the KS acylation step, other than the not unsurprising observation that a large excess of SNAC thiol was able to remove an acyl group from the KS active site by transthioesterification [6].

In an effort to address the question of potential reversibility in KS acylation, a pre-acylated KS domain possessing an upstream ACP was required. In order to test the ability of a tethered upstream ACP domain to deacylate the KS active site, the *apo*-ACP1-KS1didomain was employed once more. First, acetyl-SNAC thioester was used to acylate the KS active site Cys residue. This reaction proceeded smoothly with a single acylation event (Fig. 6.13), and the excess acetyl-SNAC was removed by ultrafiltration. Next, PPant transferase Svp from *Streptomyces verticillus* was employed successfully to transfer the PPant chain from CoASH to the conserved serine of the ACP, resulting in the required *holo*-ACP1-KS1didomain with a pre-acylated KS active site. MS/MS analysis of this species using PPant ion ejection resulted in the spectrum shown in Fig. 6.14. In addition to the fragment ion at *m/z* 261.2, corresponding to the unmodified PPant chain originally loaded onto the ACP, an ion at *m/z* 303.2 was also clearly detected. Being the

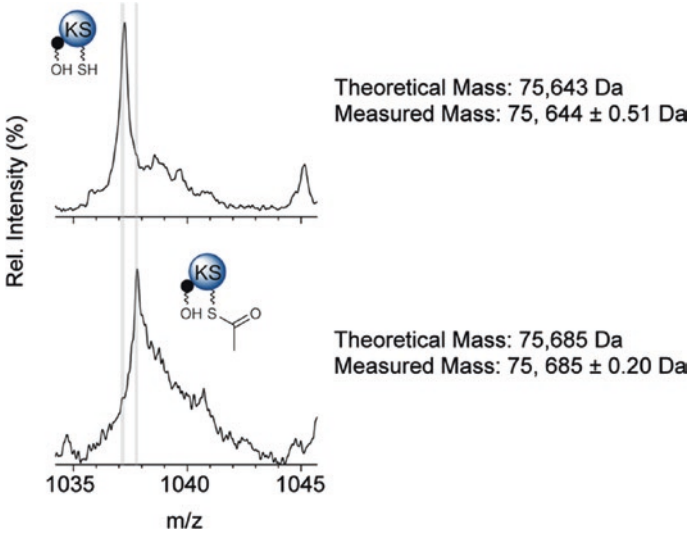



**Fig. 6.13** Stacked nESI-MS spectra of the $73^{+}$ charge state of *apo*-PsyA ACP1-KS1 following acylation with acetyl-SNAC, yielding a mass increase of +42 Da, good agreement with theoretical mass shift for this modification

signature for acetyl-ACP, the presence of this fragment provided strong evidence for transfer of the acetyl unit from KS1 to the nominal donor ACP, which, in turn, demonstrated the inherent reversibility of ACP↔KS acyl transfer.

Given the significance of this result, two control measurements were conducted in addition to those already described above for the 'forward' acyl transfer reaction. Firstly, it was crucial to rule-out the presence of any residual acetyl-SNAC activity being carried through from the initial KS acylation step. This could potentially lead to a population of acetyl-ACP within the didomain formed from acetyl-SNAC transthioesterification and not by intramolecular transfer from the KS itself. The acylation, and Svp-catalysed PPant attachment were carried out as above, but using an inactive batch of *apo*-ACP1-KS1 didomain, which had been pre-alkylated with iodoacetamide on the KS active site Cys residue (see Fig. 6.12a). Unable to undergo acylation on the KS domain, this species should show only an unacylated PPant ejection ion under MS/MS activation, and this proved to be the case (Fig. 6.15). Therefore it was possible to deduce that no acetyl-SNAC activity remained during the PPant loading step.

A second control was conducted to eliminate the possibility of intermolecular acyl transfer from the KS active site to CoASH before PPant loading. Such a reaction would result in direct loading of acetyl-PPant on to the ACP of the didomain, without the need for intramolecular transfer. Incubation of acetylated *apo*-ACP1-KS1 with CoASH in the quantities employed in the PPant loading step produced

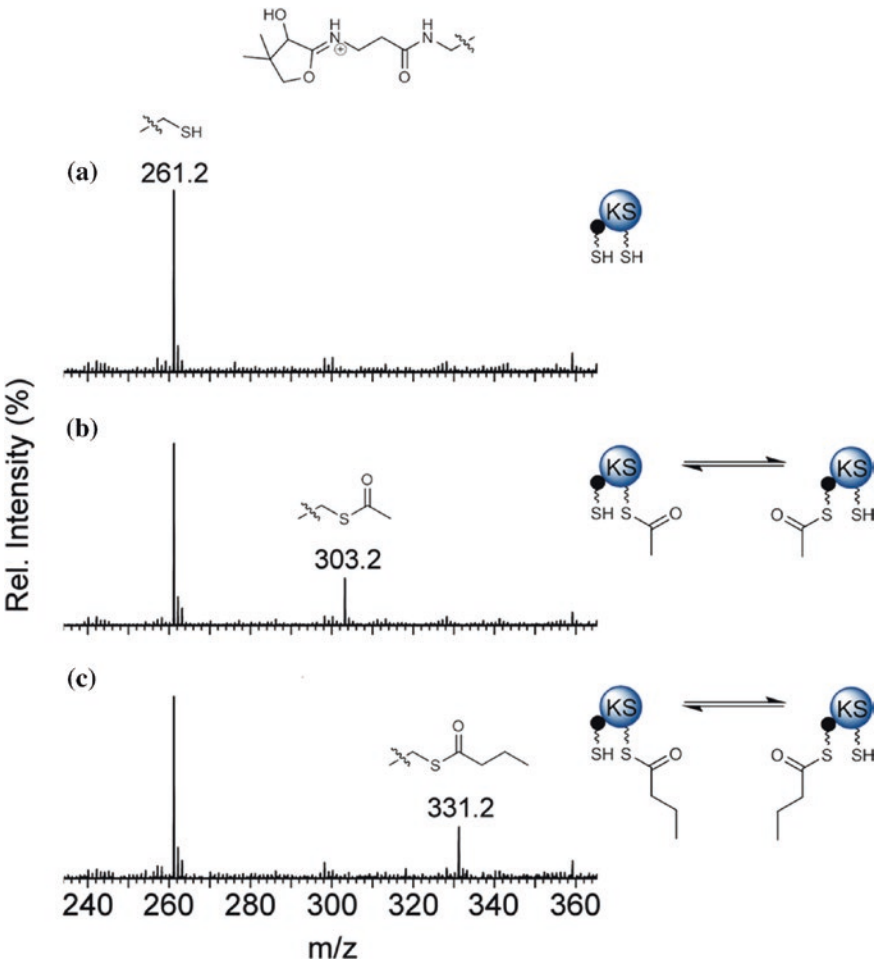



**Fig. 6.14** Reversibility of acyl transfer between ACP and KS. MS/MS spectra of activated 73 + PsyA ACP1-KS1 species from **a** unacetylated **b** acetylated and **c** butyrylated forms following PPant loading. PPant loading of acetyl- and butyryl-PsyA ACP1-KS1gives rise to PPant ions at *m/z* 303.2 and 331.2 indicating deacetylation of the KS active site by the ACP1 PPant thiol

no detectable transfer of the acetyl group from the didomain (even over a 1 h period), as evidenced by MS analysis. This finding was unsurprising, as, from previous experience, this type of intermolecular transfer is very slow and requires a large excess of thioester donor and extended incubation times. These controls, together with those described in Sect. 6.2.7, point conclusively to the existence of reversible acyl transfer within the ACP1-KS1 didomain.

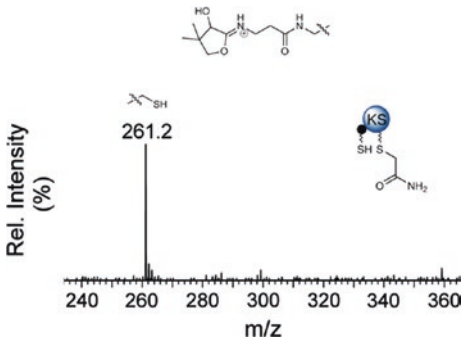



**Fig. 6.15** MS/MS spectrum of the activated 73 + alkyl-PsyA ACP1-KS1 species following the reaction controlling for any residual acetyl-SNAC activity. Only *m/z* 261.2 is observed, indicating no acylation of the ACP1 PPant chain from residual acetyl-SNAC

## 6.3 Conclusions

A key feature of polyketide biosynthesis is that individual PKSs are generally responsible for the production of a specific compound, or subtle variants thereof. In order for product fidelity to be maintained, substrate specificity for defined intermediates must be exhibited by some, or all, of the PKS catalytic domains. The KS domain has long been thought to perform a 'gatekeeper' role in polyketide biosynthesis, thereby imposing strict substrate specificity upon the intermediates [2, 14, 15]. In modular PKSs, KS domains receive the polyketide intermediate directly from the upstream module's ACP by acylation of the active site, followed by a condensation reaction with a malonyl unit from the downstream ACP. By exhibiting substrate specificity, the KS domain may ensure that the correct degree of reductive—and other—processing has occurred within the preceding module. In principle, KS specificity could manifest at the initial acylation step, at the C-C forming elongation step, or both.

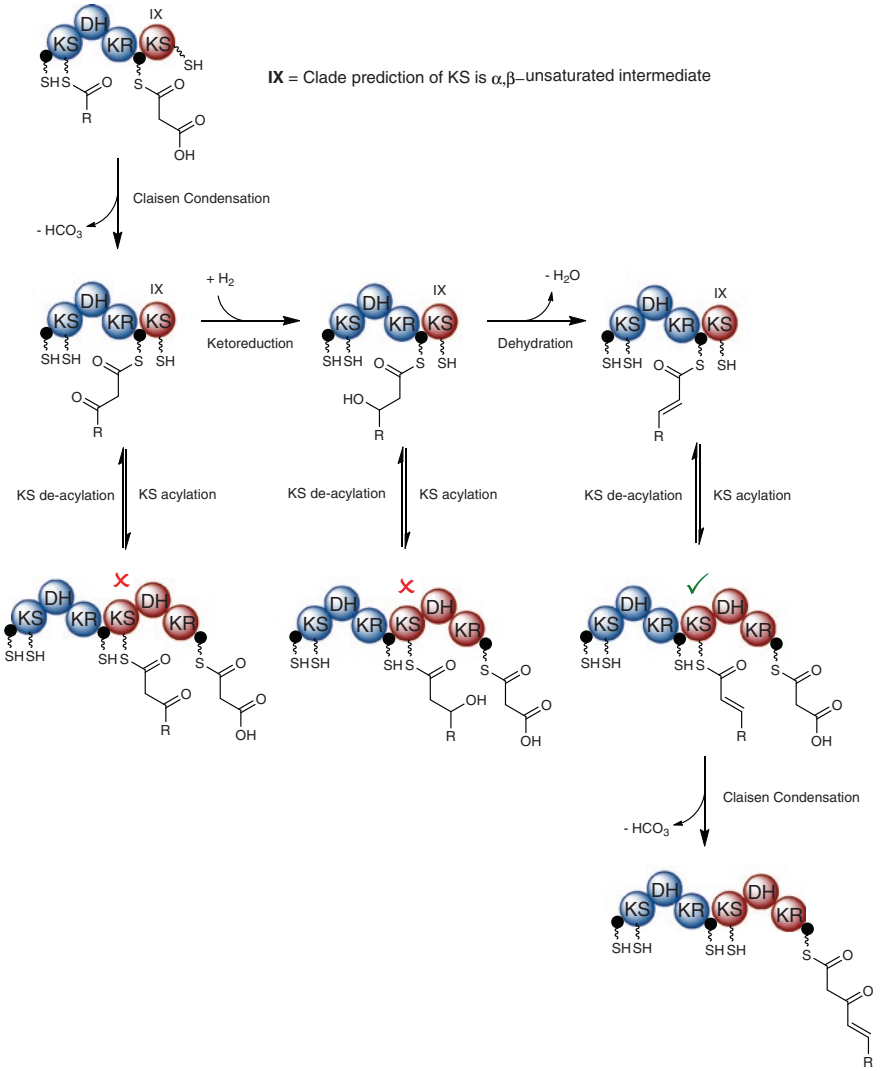
As discussed previously, phylogenetic analysis based on primary sequence data, has revealed that KS domains from *trans*-AT PKSs group into distinct clades according to their predicted substrate (see Fig. 1.12). This feature implies that KS domains in this important group of PKSs do exhibit specificity, and that it can be exploited to assign PKSs to their products in cases where colinearity rules break down [8, 16]. The work presented in Chaps. 3 and 4 examined substrate tolerance during the KS acylation step using simple SNAC thioesters, revealing that some specificity does indeed exist. However, the specificity is not as stringently defined as phylogenetic analysis would predict.

The work presented in this chapter shows that a high level of specificity is observed during the KS-catalysed elongation step, generating specificity profiles in line with phylogenetic predictions. For example, PsyA KS1 was shown to possess a relatively relaxed substrate tolerance using simple SNAC thioesters, with little preference regarding substrate length and branching (see Fig. 3.5). The elongation profile of PsyA KS1 is markedly different to that of the elongation step, with the natural acetyl substrate clearly favored (Fig. 6.8a). These observations reveal strikingly different levels of specificity at the acylation and elongation steps.

The relatively low specificity of the acylation step is further confirmed using the PsyA ACP1-KS1 as a model system for a tethered substrate. Here, no discernible difference in acyl group transfer was observed following loading with acetyl- and butyryl-CoA, despite a clear difference between these acyl chains at the elongation step (Fig. 6.11). This result is somewhat unsurprising as thioester bonds are high energy linkages, which form and break rapidly in the presence of a thiol, or thiolate, by a process of transthioesterification. The rate of this reaction greatly exceeds that of hydrolysis, as thiol groups are significantly more nucleophilic than water [17]. Thus, acyl transfer between an ACP and the immediate downstream KS may be difficult to control, leading to relatively limited enzymatic selectivity, and a potentially reversible process.

The KS-catalysed chain elongation step involves a Claisen-type condensation with a malonyl unit on the downstream ACP. This reaction forms a new C–C bond and is accompanied by a formal decarboxylation, although evidence from fatty acid biosynthesis indicates that $CO_2$ is actually lost as bicarbonate [18, 19]. Whilst this process is formally reversible, in practice it is likely to be irreversible, requiring ATP-driven enzymatic activity reminiscent of acetyl-CoA carboxylase [20]. The apparent difference in specificity observed in KS acylation and elongation steps does raise a potential problem: that of acylation occurring at the KS active site by a substrate which is unable to be elongated. This situation could arise by, for example, incomplete reductive processing by the previous module. However, providing that KS acylation is reversible, this problem can be overcome, since the acyl chain can be transferred back to the upstream module for complete processing. Only then will subsequent KS acylation lead to effective elongation. The rate constant for back transfer from KS to upstream ACP has been measured for the *cis*-AT DEBS module 2 and found to be an order of magnitude lower than either the forward reaction, or the subsequent elongation step, leading to the conclusion that acylation is practically irreversible [6]. It should be pointed out, however, that these observations were made for a productive substrate under steady state conditions, and not for a stalled substrate incapable of elongation. Observations from the PsyA ACP1-KS1 didomain do, indeed, demonstrate that KS acylation is a reversible process, and that this explanation is plausible. It is therefore possible to imagine a competition between processing of the acyl chain within a module and transfer to the downstream KS (Scheme 6.3).

The results presented also give valuable insights into the substrate flux within multimodular PKSs. If a multitude of reversible reactions do indeed exist during biosynthesis, as proposed in Scheme 6.2, then the driving force effectively ‘pushing’ intermediates through the biosynthetic assembly line is the concentration of extender unit (typically malonyl-CoA) within the cell. Modular type I PKSs which have had their TE domain rendered inactive, have been shown to harbour ACPs saturated with intermediates, forming a ‘molecular traffic-jam’ [21]. The data presented here support these observations, as with no release mechanism available, acyl intermediates will be in equilibrium between the upstream ACP and the KS, a proportion of which will be located on the PPant arm of the ACP.




**Scheme 6.3** Hypothetical biosynthetic example of how reversibility of the acylation step and high specificity of the condensation stage promotes elongation of the correct acyl intermediate. Here, incomplete reductive processing by the previous module results in acylation of the downstream KS with a substrate incapable of being elongated. However, the reversible nature of the acylation allows complete reductive processing to be conducted, ultimately yielding an α, β-unsaturated intermediate—the correct substrate of KS domains in clade IX—which can be successfully elongated

Recent work on pikromycin module 5 has provided new structural insights into the dynamics of the ACP domain within a module, and in the subsequent loading the KS domain of the downstream module. Using a crystal structure and cryo-EM the authors show that the KR domain undergoes a large conformational change upon acylation of the KS domain, and that the position of the ACP domain within the module is dependent upon the nature of the acyl chain [22, 23]. This mechanism is proposed to be a major contribution in the delivery of the correctly modified acyl chain to the downstream module, meaning that acylation of the KS with the incorrect substrate may be prevented. Therefore the ACP may be incapable of acylating the downstream KS6 until full substrate processing has occurred, which would prevent acylation of the downstream KS with a non-productive intermediate, and reduce stalling [24]. Whether a similar mechanism exists in modules lacking KR domains, or in *trans*-AT PKSs is yet to be shown. The findings presented here, using the PsyA ACP1-KS1 didomain, demonstrate that KS acylation in this *trans*-AT PKS didomain is a reversible process (in the absence of a downstream malonyl-ACP), and that this mechanism may allow unloading of non-productive substrates from the KS domain.

In summary, it is clear from this work that KS domains from *trans*-AT PKSs do possess substrate specificity compatible with that proposed by phylogenetic analysis, and that this specificity is seen principally in the elongation step of the catalytic process. Encouragingly, mutagenesis of BaeL KS5 suggests that this specificity can be tuned to allow the elongation of non-natural substrates.

## References

1. D.C. Gay et al., A close look at a ketosynthase from a *trans*-acyltransferase modular polyketide synthase. Structure **22**, 444–451 (2014)
2. B. Busch et al., Multifactorial control of iteration events in a modular polyketide assembly line. Angew. Chem. Int. Ed. **52**, 5285–5289 (2013)
3. M. Till, P.R. Race, Progress challenges and opportunities for the re-engineering of *trans*-AT polyketide synthases. Biotechnol. Lett. **36**, 877–888 (2014)
4. R.J. Heath, C.O. Rock, The Claisen condensation in biology. Nat. Prod. Rep. **19**, 581–596 (2002)
5. K. Watanabe, C.C.C. Wang, C.N. Boddy, D.E. Cane, C. Khosla, Understanding substrate specificity of polyketide synthase modules by generating hybrid multimodular synthases. J. Biol. Chem. **278**, 42020–42026 (2003)
6. J.Q. Wu, K. Kinoshita, C. Khosla, D.E. Cane, 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**, 16301–16310 (2004)
7. L. Gu et al., GNAT-like strategy for polyketide chain initiation. Science **318**, 970–974 (2007)
8. T. Nguyen et al., Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nat. Biotechnol. **26**, 225–233 (2008)
9. K.M. Fisch et al., Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. Nat. Chem. Biol. **5**, 494–501 (2009)
10. T. Gulder, M. Freeman, J. Piel, The catalytic diversity of multimodular polyketide synthases: natural product biosynthesis beyond textbook assembly rules. Top. Curr. Chem. 1–53 (2011)

11. D.C. Gay et al., A close look at a ketosynthase from a *trans*-acyltransferase modular polyketide synthase. Structure **22**, 444–451 (2014)
12. C. Sanchez, L.C. Du, D.J. Edwards, M.D. Toney, B. Shen, Cloning and characterisation of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. Chem. Biol. **8**, 725–738 (2001)
13. K. Jensen et al., Polyketide proofreading by an acyltransferase-like enzyme. Chem. Biol. **19**, 329–339 (2012)
14. A.T. Keatinge-Clay, The structures of type I polyketide synthases. Nat. Prod. Rep. **29**, 1050–1073 (2012)
15. Y.-M. Zhang, J. Hurlbert, S.W. White, C.O. Rock, 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**, 17390–17399 (2006)
16. T. Hochmuth, J. Piel, Polyketide synthases of bacterial symbionts in sponges- Evolution-based applications in natural products research. Phytochemistry **70**, 1841–1849 (2009)
17. P.J. Bracher, P.W. Snyder, B.R. Bohall, G.M. Whitesides, The relative rates of thiol-thioester exchange and hydrolysis for alkyl and aryl thioalkanoates in water. Origins Life Evol. B **41**, 399–412 (2011)
18. A. Witkowski, A.K. Joshi, Y. Lindqvist, S. Smith, Conversion of a beta-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. Biochemistry **38**, 11643–11650 (1999)
19. A. Witkowski, A.K. Joshi, S. Smith, Mechanism of the beta-ketoacyl synthase reaction catalyzed by the animal fatty acid synthase. Biochemistry **41**, 10877–10887 (2002)
20. M.D. Lane, J. Moss, S.E. Polakis, Acetyl coenzyme A carboxylase. Curr. Top. Cell. Reg. **8**, 139–195 (1974)
21. J. Moldenhauer, X.H. Chen, R. Borriss, J. Piel, Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. Angew. Chem. Int. Ed. **46**, 8195–8197 (2007)
22. S. Dutta et al., Structure of a modular polyketide synthase. Nature **510**, 512–517 (2014)
23. J.R. Whicher et al., Structural rearrangements of a polyketide synthase module during its catalytic cycle. Nature **510**, 560–564 (2014)
24. P.F. Leadlay, Structural biology: Enzyme assembly line pictured. Nature **510**, 482–483 (2014)

# Appendix

## Protein Sequences

### PsyA KS1

```
         10         20         30         40         50         60
MKHHHHHHHH GGLVPRGSHG GSEFDQSARA EKGVAIVGMA CRLPGGITTP EALWTVLAEG

         70         80         90        100        110        120
RDVVGTVPAG RWVWPQETGP EHGDPGIDCG GFLDDIARFD AKLFRISPRE AKVMDPQQRL

        130        140        150        160        170        180
LLELAWSAFE DAGYSKDAVE GTKTGVFVGA SGSDYRLLLE QHRVNIEPVM GTGTAVAVLP

        190        200        210        220        230        240
NRISYFFDLQ GPSLLIDTAC SSSLVAIHEA VQALRAGSCE QALVGGINIM CHPAMTLAYY

        250        260        270        280        290        300
KAGMLSPDGR CKTFDAEANG YVRSEGAIVM MLKPLSAAQR DGDRIYAVVK GSACNHGGQA

        310        320        330        340        350        360
GGLTVPNPQQ QTALLRAAWA SARVTPDQLG YLEAHGTGTS LGDPIEVKGM QDAFRADDNI

        370        380        390        400        410        420
AAATTCYLGS VKSNLGHLEA AAGIAGLMKL ALCLYHRQLV SSLHVHTVNP KLGLEQTPFQ

        430        440        450        460        470        480
IAQQVMAWPT LKSGQPSLTG VSSFGSGGTN AHVVVEGVEQ VGPARAERPV VIRLSAPNVE

        490        500        510        520        530        540
QLAIYARCLR DYLQGLPERA RPPLSALAYT LSRRQPMAVS ASYWARDEAS LVSGLADIAA

        550        560        570        580        590        600
GLVTSVEEER GLSFGEGPVI ALPGYPFAET SFWFDKPEAQ AAPARPAKVA LEDPVVIARR

        610
GLGIVSDVLT RS
```



M. Jenner, *Using Mass Spectrometry for Biochemical Studies on Enzymatic Domains from Polyketide Synthases*, Springer Theses,
DOI 10.1007/978-3-319-32723-5

## PsyA KS2

```
        10         20         30         40         50         60
MKHHHHHHHH GGLVPRGSHG GSEPSRVEMA ASIVEVKTSD DAPISDVEGD MPAIAVIGMS

        70         80         90        100        110        120
GQFPQANNVE ALWQNLVEGR DCISEVPLDR WPVDAYFDPT PQVPGKTYSR WMGVLEDADK

       130        140        150        160        170        180
FDPLFFSISP REAMAMDPQQ RLFLETCWSC VEDAGYAPSS LSGTRCGVFA GCGVSDYNQH

       190        200        210        220        230        240
LDADGLDAQR FMGGSTSILA ARISYELNLR GPSMAVDTAC SASLVAIAVA CDNLVAGACD

       250        260        270        280        290        300
TALAGGVCVM AGPAMHIMTS QARMLSPDGR CFTFDQRANG FVPGEGVGVV LLKRLADAER

       310        320        330        340        350        360
DGDRILGVLR GWGVNQDGKT NGITAPSGDS QTGLQRDVYE RYGIDPATIQ LVEAHGTGTK

       370        380        390        400        410        420
LGDPIEVEGL CQAFSSFTDQ RNYCALGSAK SNIGHLLMAA GVAGLIKTLL ALQHQTLPPT

       430        440        450        460        470        480
IHFEQLNEHI ALDDSAFYVN DRIRPWASQG ATPRRAAVSS FGFSGTNAHV VVEEYATDAR

       490        500        510        520        530        540
RSHSRAAGPF LVVLSAKQKE RLREAVQRLC DHLAAHPDAS IADLAYTLQV GRDAMIERVG

       550        560        570        580        590        600
FLVASCDDLM QQLADFLDGR ASGWRQGTVG KGEQAVVDAT GADEATLRAW ASGARIDWSG

       610        620        630        640
LYPDETRPTR LSLPTYPFAR ERYWACAALP GGEALEGDWT LQPLSTD
```

## PsyD KS3$^{0}$

```
        10         20         30         40         50         60
MKHHHHHHHH GGLVPRGSHG GSQLRWIEDK MRDAQGGSAG PEPIAIVGLS GMFPQCSDVR

        70         80         90        100        110        120
AFWRALDADQ ALLEELPTTR FPWRDWYDAT GENPDKSRSK VGGFLPDIAS FDPRFFGVLP

       130        140        150        160        170        180
DDAARMDPRQ RLLLMAVYHA LEDAGIDAGS LKKSRTGVFV AGEDNEYAQV LREAEVDLGD

       190        200        210        220        230        240
GFAQAANMLA NQISYFFDFA GPSEMINTMC SGGAVALHRA VSALRAREVE LAVVGAANVI

       250        260        270        280        290        300
LRPEPFVQLS RAKQLSTTAT VRSFGEGADG HLRAEGVASV LLKPLRAAEA AGDRIYAVIK

       310        320        330        340        350        360
HSAVNYNGQG GMSIAAPFVQ SHQEVIRACY DEAKVDPREV GYIEAQGMGN PVADLAEWHA

       370        380        390        400        410        420
CNNALRAMAQ EQGVALPKGN CRVSSLKPML GHMESTSAFG ALFKIIRSFQ THTVHQIVGF

       430        440        450        460        470        480
AKPNPELVVE QQPCRLMAAT EPWPAGPVPR LAGLHAYGIG GNNAHLLVEE YRDVRRPVRD

       490        500        510        520        530        540
EESGSEWVVL SARTKDRLQA VARWLADDME HDQPHLADVA YTLQTGRESM RERAAFIAHS

       550        560        570        580        590        600
IADLRAKLLA VAEGTTGEGL LYGSVAPRSR SRGQAAEQAI DPSEAWYTLG RRWVDGDVIA

       610        620        630        640
WDQARPADSA QRIALPGYPF EQRRCWFEAP PPVESGDADQ APNKLIDL
```

## PsyA ACP3

```
        10         20         30         40         50         60
MKHHHHHHHH GGLVPRGSHG GSTSSGELAT VVRTTLMQLL ELPTVDDDEA FQNYGLDSIS

        70         80         90
ATIFSNRLEQ VLGQPVLPHW LIDYPTVSAL AQQLEAVCV
```

## PedC AH

```
         10          20          30          40          50          60
MKDLQNIQNT HPVVWMFSGQ GSQYFQMGRQ LYEQDETFHA WMKSLDDNVR DYIGQSLLDI

         70          80          90         100         110         120
IYDTGHERSL PFDRLIHTHP ALFMVQYALA KSLLARGLPA PDFLIGASLG EFIAISLAGD

        130         140         150         160         170         180
THVENILFNL IKQARLFDEY CNAGAMLLVI DHIDTFSTTP AFSKDCELAG INFDHCFVVS

        190         200         210         220         230         240
GPRTGILQTR KSLTKQNIAC QLLPVSIAFH SSWMDEVHEI FIQQFPEQIC RRLHTPVISC

        250         260         270         280         290         300
ALPVPEQLTR FSSTYWWHVI RQPIAFHLAI NTFHQSSPNA VYLDLGPAGN MAAATKYNLP

        310         320         330         340
SSIHYRILPT MTPFGRDLEN IEIARLRLLE LDQRLEHHHH HH
```

## PedD AT

```
         10          20          30          40          50          60
MKHHHHHHHH GGLVPRGSHG GSEFKSYLFP GQGSQHLGMG EQLFDRFPNI IEAANDILGY

         70          80          90         100         110         120
SIKTLCLEDP QRQLRLTQYT QVALYVVNAL TYRQHLQQGG GLPDFVAGHS LGEYNALESA

        130         140         150         160         170         180
GVFSFEDGLR LVQKRGDLMS QAPRGAMAAI LGISADSVAG ILAEQGLTRI DIANYNAPTQ

        190         200         210         220         230         240
TIISGLEADI RDAQAVFESC QAMYVPLNTS GAFHSRYMQS ARDEFAQFLE AFEFRDPQIP

        250         260         270         280         290         300
VVANVTAKPY VGTEVVRTLA DQLTGSVRWL DSMRFLLDQG VTEFRELGPG DVLSKLVESI

        310         320         330         340         350         360
RSSAMSKPVS EFAAENSQQL VDEWNRTCPI GSRVRVKGYD DILVTKSRAV LLFGHRAAIY

        370         380
MENYQGYFAL SEVEPLIEQQ PLVEKVW
```

## BaeJ KS1

```
         10          20          30          40          50          60
MKHHHHHHHH GGLVPRGSHG TEQKPKEAFL APVPQKKVNA EDREHPAGEF PDYYEDSLAV

         70          80          90         100         110         120
IGISCEFPGA KDHYEFWNNI KEGKESITFF SKEELRRSGI SEELADHPGF VPAKSVLEGK

        130         140         150         160         170         180
EMFDPGFFGF SPKDAEYMDP QLRMLLLHSW KAIEDAGYIS KEIPETSVYM SASTNSYRSL

        190         200         210         220         230         240
LPEETTAQLE TPDGYVSWVL AQSGTIPTMI SHKLGLKGPS YFVHANCSSS LIGLHSAFQS

        250         260         270         280         290         300
LQSGEAKYAL VGGATLHTES SAGYVHQPGL NFSSDGHIKA FDADADGMIG GEGAGAVLLK

        310         320         330         340         350         360
KASDAVKDGD HIYALLRGIG VNNDGADKVG FYAPSVKGQA EVIQKVIDQT GIHPETIAYV

        370         380         390         400         410         420
EAHGTGTKLG DPIELSALQS VYGRYTDKKQ YCGIGSVKTN LGHLDTAAGM AGCIKVVMSL

        430         440         450         460         470         480
YHQEIAPSIN YKEPNPNLHL EDSPFFVAEE KKELTRENRA HRMALSSFGL GGTNTHAIFE

        490         500         510         520         530         540
QYPDASEAAD AAGPFIIPLS ARKKDRLKEY AKQLLAFLER KTDTDLADLA YTFQVGREAM

        550         560         570         580         590         600
EERAAFITSG TAELKRQLAD FINDKPAVTG CFRGEKQQAK DIAWLSDDDD SAELIEKWLA

        610         620         630         640         650         660
KGKGPKLCEM WSKGVAINWH KLYKDKHPKR ISLPVYPFAK EPYWPKKAEK NTSAAHTGVS


VLHPLVQQN
```

## BaeJ KS2

```
        10         20         30         40         50         60
MKHHHHHHHH GGLVPRGSHG GSAKEHPGRF GEKKKESPKK EQPKAQKPKM QRKKRFATVM

        70         80         90        100        110        120
NASAATQEPR RFDPVAIVGI SGRFPGAKDI EEFWRNLKEG KDSITTIPKE RWDWQAFDGD

       130        140        150        160        170        180
PNLEGNKTNI KWGGFIDGIA EFDPLFFGIS PREAQYLDPQ QRLLLTYAWR AIEDAGCKPE

       190        200        210        220        230        240
SLSGTNTGVF IGTGNTGYKD LFTRAGLAPE GHAATGSMIP SIGPNRLSYL LNLHGPSEPI

       250        260        270        280        290        300
ETACSSSLVA IHRAVSAIEN GECDMAIAGG INTILTEEAH ISYSKAGMLS KDGKCKTFSK

       310        320        330        340        350        360
DANGYVRGEG AGILMLKKLS DAERDGNPIY GVIRGTAENH GGRANTLTSP NPKLQADLLV

       370        380        390        400        410        420
KAYRKAGVDP STVTYIEAHG TGTELGDPIE INGLKAAFHE LAKTNQEPEV SGHRCGIGSV

       430        440        450        460        470        480
KSNIGHLELA AGVSGVMKVL LQMKHKTLVK SLHCETINPY IQLDDSPFYI VRENQEWTAA

       490        500        510        520        530        540
KDRNGNAIPR RAGVSSFGIG GVNAHIVIEE YVPKAAAQPE HSPENPAIIL LSAKSKEKLF

       550        560        570        580        590        600
EQAQQLQKAI RQTPYTDQDL AGVAYTLQTG RDEMEERLAI LAATMAELET KLEAFTKNEK

       610        620        630        640        650        660
NIAGLYTGQS HRNKDTFALF TADEDMDIVI DAWIKKGKFA KLAEVWVKGG VFNWNRLYEK

       670        680        690        700
GTPRLLSLPS YPFAKDTYWV PENPDKNKVH TEERMKRILT KQW
```

## BaeL KS5

```
        10         20         30         40         50         60
MKHHHHHHHH GGLVPRGSHG GSMKAHRDTL LEVTGMKTEE TPAAKPALSE KAKVKPVFKR

        70         80         90        100        110        120
NEGKKGFAQV RPDTRENQQL TEREDIAIIG ISGRYPQAEN LQEFWKNLSE GTDCITEIPN

       130        140        150        160        170        180
DRWDHSLYYD ADKDKEGKTY GKWGGFLKDV DKFDPQFFSI SPRDAKLMDP QERLFLQCVY

       190        200        210        220        230        240
ETMEDAGYTR KKLTEKSGDL LGANVGVYVG VMYEEYQLYG AEEQARGKSL ALTGNPSSIA

       250        260        270        280        290        300
NRASYVFGFN GPSMALDTMC SSSLTAIHLA CQSLRNGECE AAFAGGVNVS VHPNKYLMLG

       310        320        330        340        350        360
QNRFLSSKGR CESFGEGGDG YVPGEGVGAV LLKPLSKAKA DGDHIYGLIK GTAVNHDGKT

       370        380        390        400        410        420
NGYSVPNPNA QAAVIKQALK DAGTDPRAVS YIEAHGTGTS LGDPIEITGL TKAFSEQTQD

       430        440        450        460        470        480
KQFCAIGSAK SNIGHCESAA GIAGLTKVLL QMKHKQLAPS LHSRTLNPNI DFLATPFKVQ

       490        500        510        520        530        540
QTLEEWKRPV INENGVNKEL PRTAGLSSFG AGGVNAHIVI EEYSADEDKE TAFAAPHPSM

       550        560        570        580        590        600
IVLSAKNEQR LQKRAKRLLD ALRSGRYREA DLSRIAYTLQ VGREPMEERL GMIVSNLREL

       610        620        630        640        650        660
EEKLDEFTGG KESIDQLYRG QVKQNKDTMA LFTADEDMEK TIEAWLEKGK AAKVLELWVK

       670        680        690        700        710
GLPLNWDKLY QMGRPQKISL PAYPFAKDRY WIDTSADAAV KRPAETQSAP IAPAA
```

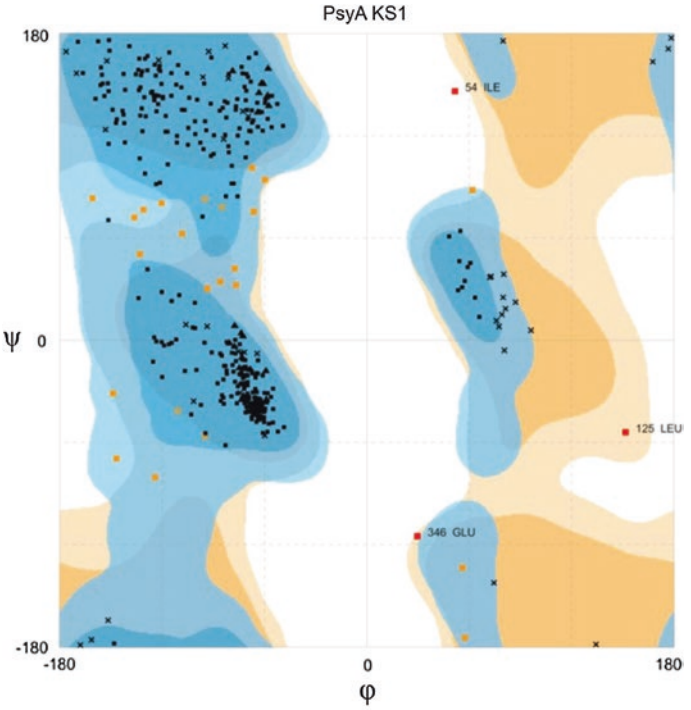
## PsyA ACP1-KS1

```
        10         20         30         40         50         60
GSSHHHHHHS SGLVPRGSHM TRQRRDGGTR SVNPTEPVIH GDFSPHLERL VRSLLLDETA

        70         80         90        100        110        120
FAWDRPLMEM GLDSADLLQL GERVASAFGV SPDPAFFFTH NTCKKILATL APSTKGREVI

       130        140        150        160        170        180
VDQSARAEKG VAIVGMACRL PGGITTPEAL WTVLAEGRDV VGTVPAGRWV WPQETGPEHG

       190        200        210        220        230        240
DPGIDCGGFL DDIARFDAKL FRISPREAKV MDPQQRLLLE LAWSAFEDAG YSKDAVEGTK

       250        260        270        280        290        300
TGVFVGASGS DYRLLLEQHR VNIEPVMGTG TAVAVLPNRI SYFFDLQGPS LLIDTACSSS

       310        320        330        340        350        360
LVAIHEAVQA LRAGSCEQAL VGGINIMCHP AMTLAYYKAG MLSPDGRCKT FDAEANGYVR

       370        380        390        400        410        420
SEGAIVMMLK PLSAAQRDGD RIYAVVKGSA CNHGGQAGGL TVPNPQQQTA LLRAAWASAR

       430        440        450        460        470        480
VTPDQLGYLE AHGTGTSLGD PIEVKGMQDA FRADDNIAAA TTCYLGSVKS NLGHLEAAAG

       490        500        510        520        530        540
IAGLMKLALC LYHRQLVSSL HVHTVNPKLG LEQTPFQIAQ QVMAWPTLKS GQPSLTGVSS

       550        560        570        580        590        600
FGSGGTNAHV VVEGVEQVGP ARAERPVVIR LSAPNVEQLA IYARCLRDYL QGLPERARPP

       610        620        630        640        650        660
LSALAYTLSR RQPMAVSASY WARDEASLVS GLADIAAGLV TSVEEERGLS FGEGPVIALP

       670        680        690        700
GYPFAETSFW FDKPEAQAAP ARPAKVALED PVVIARRGLG IVSDVLTRS
```

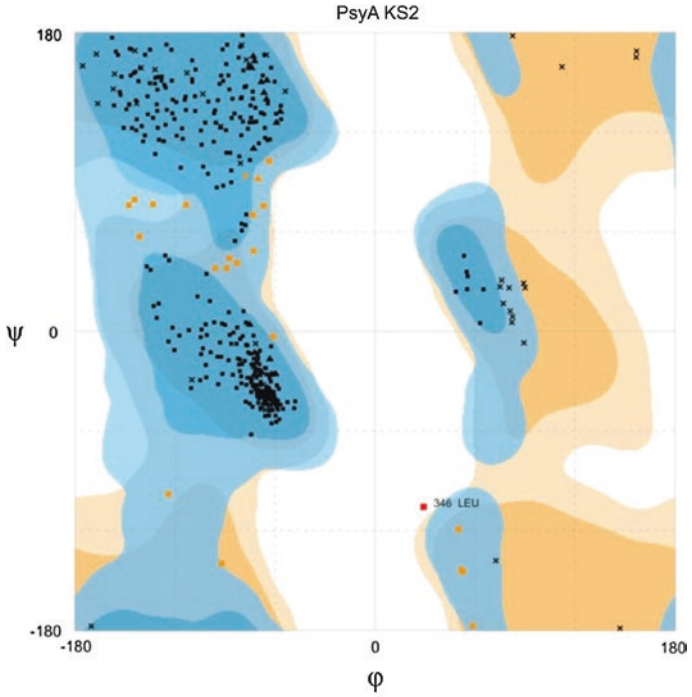
## Svp PPant Transferase (*Streptomyces verticillus*)

```
        10         20         30         40         50         60
MLAALLPSWA VTEHAFTDAP DDPVSLLFPE EAAHVARAVP KRLHEFATVR VCARAALGRL

        70         80         90        100        110        120
GLPPGPLLPG RRGAPSWPDG VVGSMTHCQG FRGAAVARAA DAASLGIDAE PNGPLPDGVL

       130        140        150        160        170        180
AMVSLPSERE WLAGLAARRP DVHWDRLLFS AKESVFKAWY PLTGLELDFD EAELAVDPDA

       190        200        210        220        230        240
GTFTARLLVP GPVVGGRRLD GFEGRWAAGE GLVVTAIAVA APAGTAEESA EGAGKEATAD

       250
DRTAVPRSHH HHHH
```

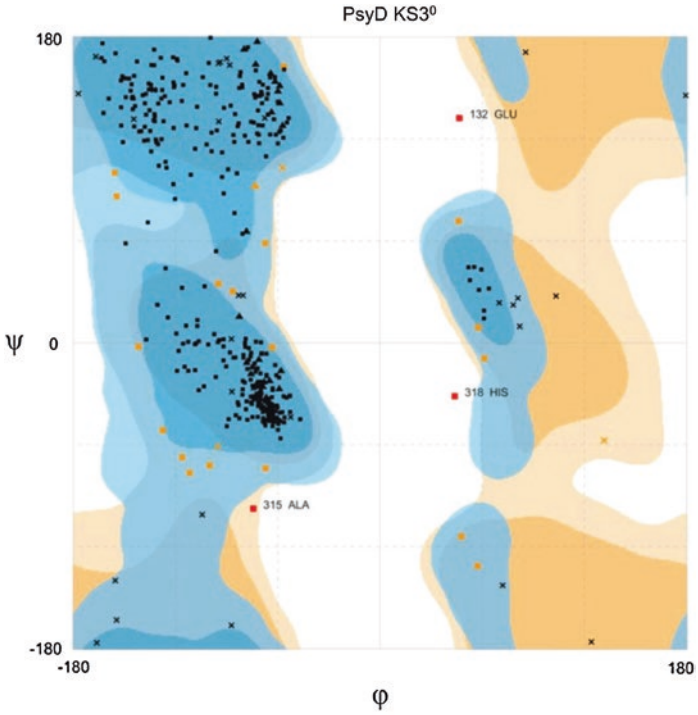
# Analysis of Homology Model Quality



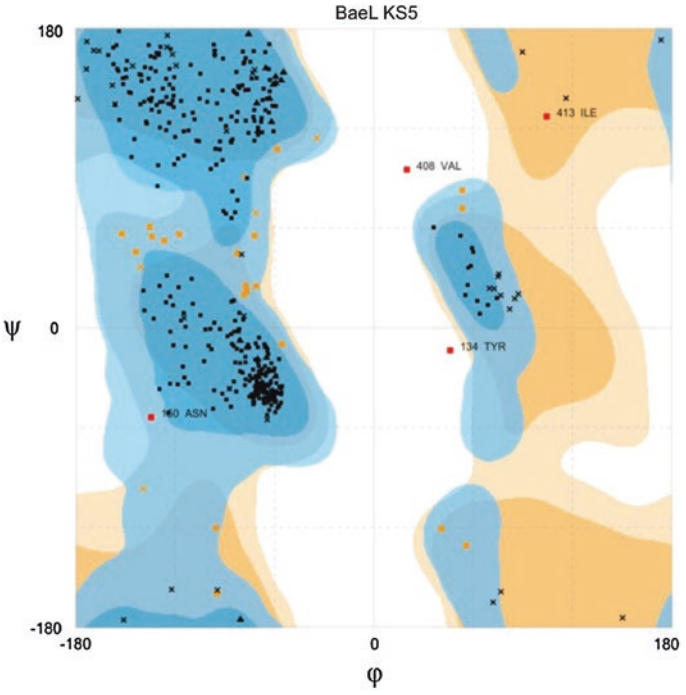

| | |
|---|---|
| Number of residues in favoured region (~98.0% expected) | : 395 (93.8%) |
| Number of residues in allowed region (~2.0% expected) | : 23 (5.5%) |
| Number of residues in outlier region | : 3 (0.7%) |



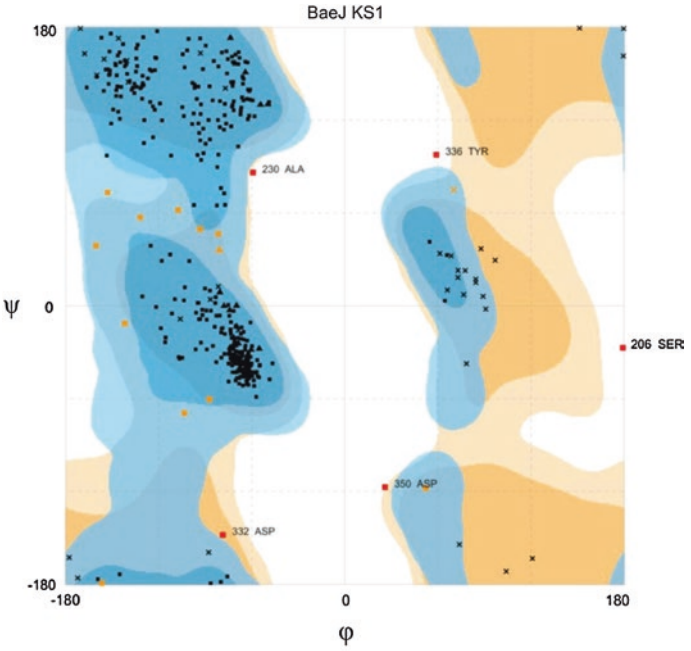

| | |
|---|---|
| Number of residues in favoured region (~98.0% expected) | : 396 (94.5%) |
| Number of residues in allowed region (~2.0% expected) | : 22 (5.3%) |
| Number of residues in outlier region | : 1 (0.2%) |



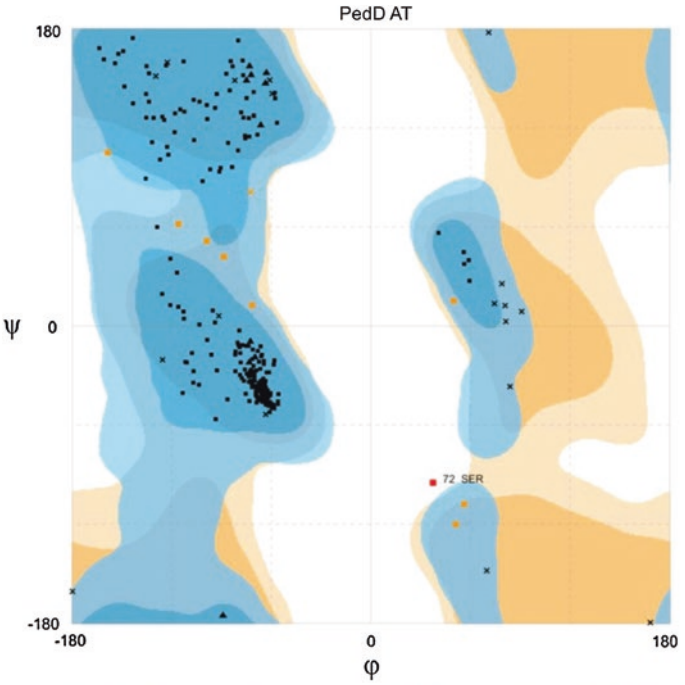

| | |
|---|---|
| Number of residues in favoured region (~98.0% expected) | : 402 (94.1%) |
| Number of residues in allowed region (~2.0% expected) | : 22 (5.2%) |
| Number of residues in outlier region | : 3 (0.7%) |



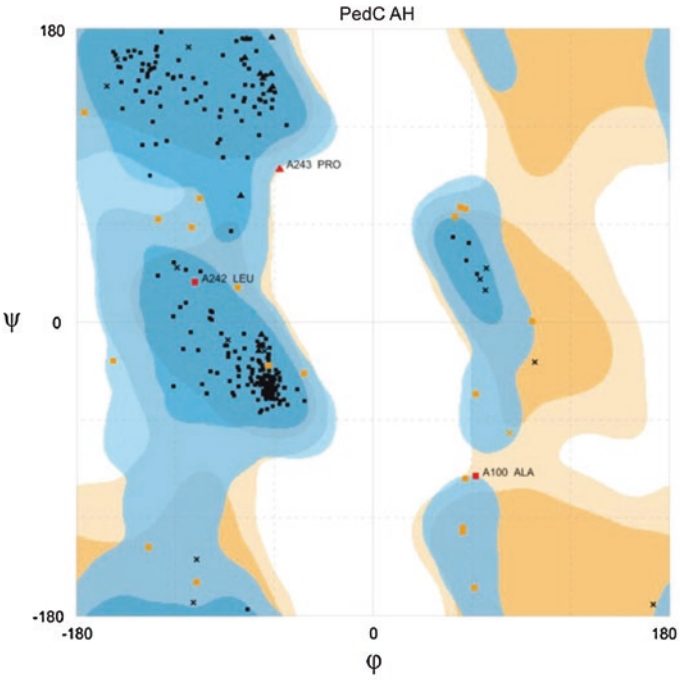

| | |
|---|---|
| Number of residues in favoured region (~98.0% expected) | : 406 (93.3%) |
| Number of residues in allowed region (~2.0% expected) | : 25 (5.7%) |
| Number of residues in outlier region | : 4 (0.9%) |




| | |
|---|---|
| Number of residues in favoured region (~98.0% expected) | : 406 (95.8%) |
| Number of residues in allowed region (~2.0% expected) | : 13 (3.1%) |
| Number of residues in outlier region | : 5 (1.2%) |

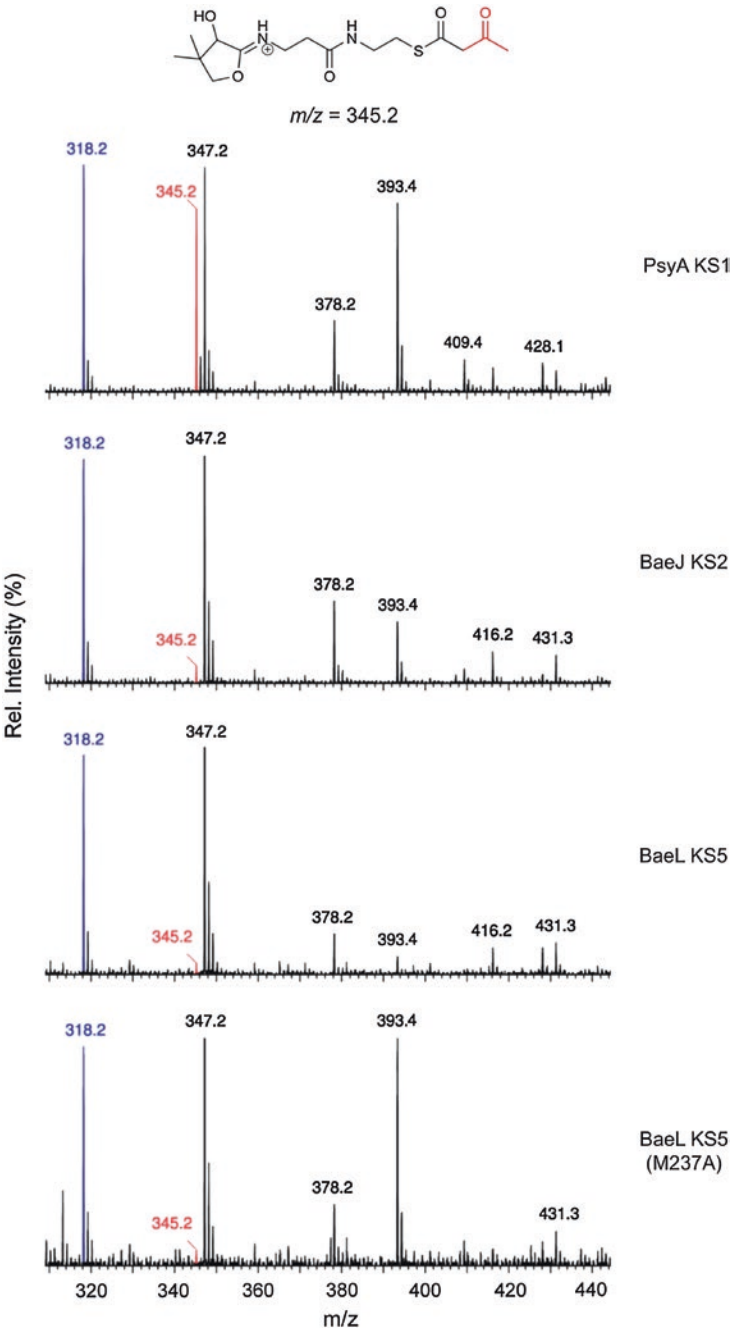


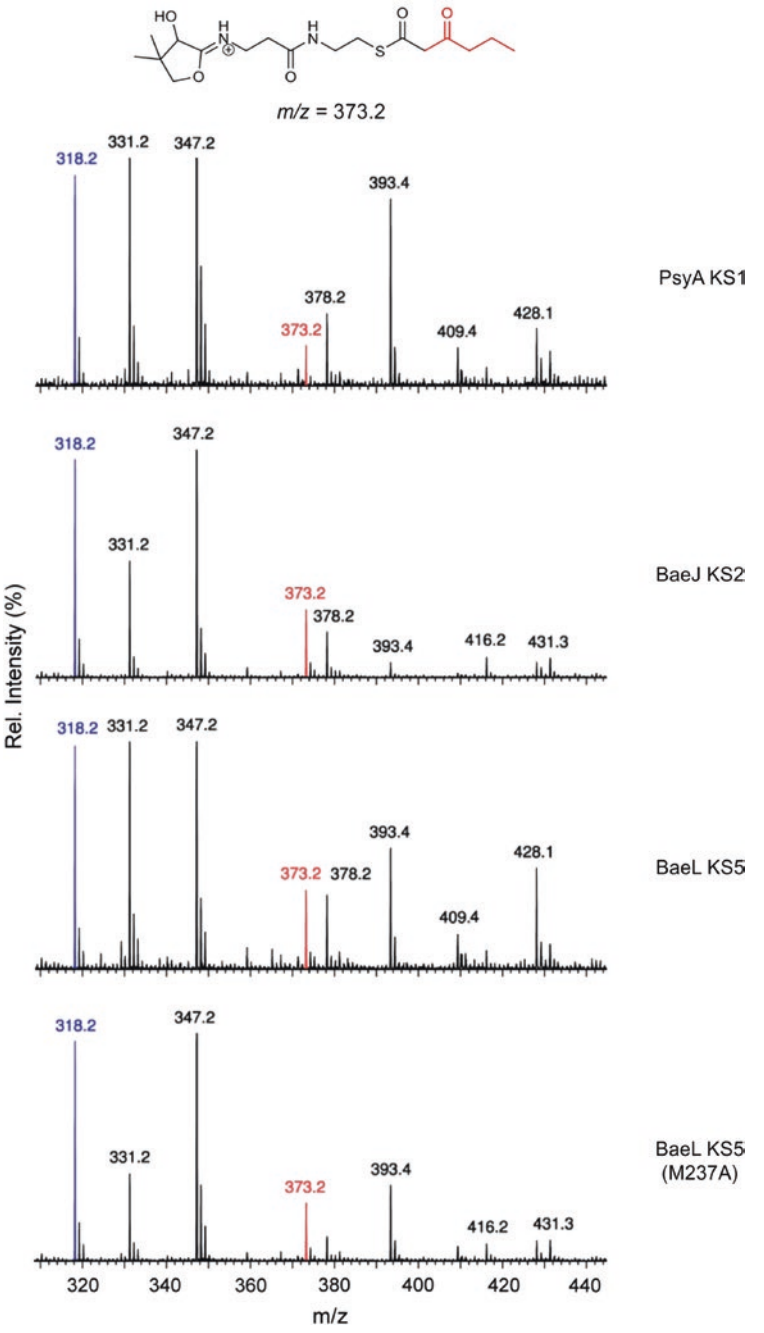

| | |
|---|---|
| Number of residues in favoured region (~98.0% expected) | : 251 (96.2%) |
| Number of residues in allowed region (~2.0% expected) | : 9 (3.4%) |
| Number of residues in outlier region | : 1 (0.4%) |

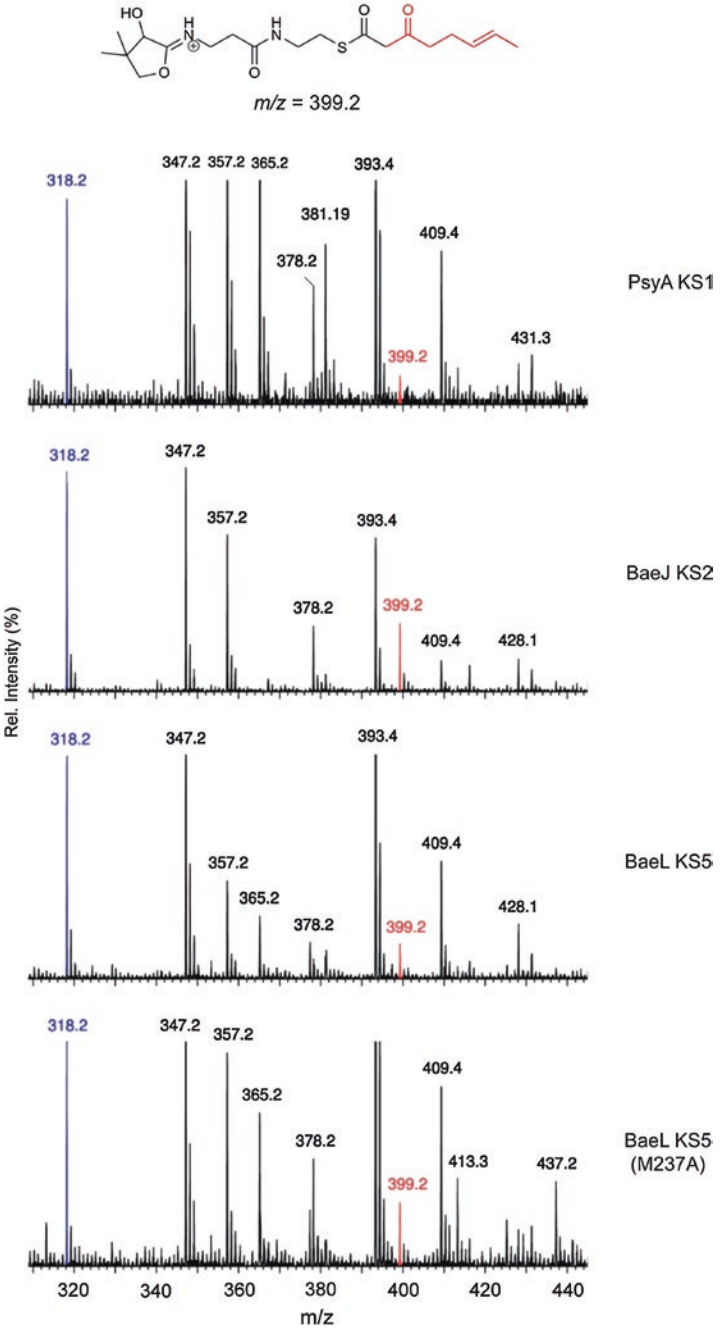


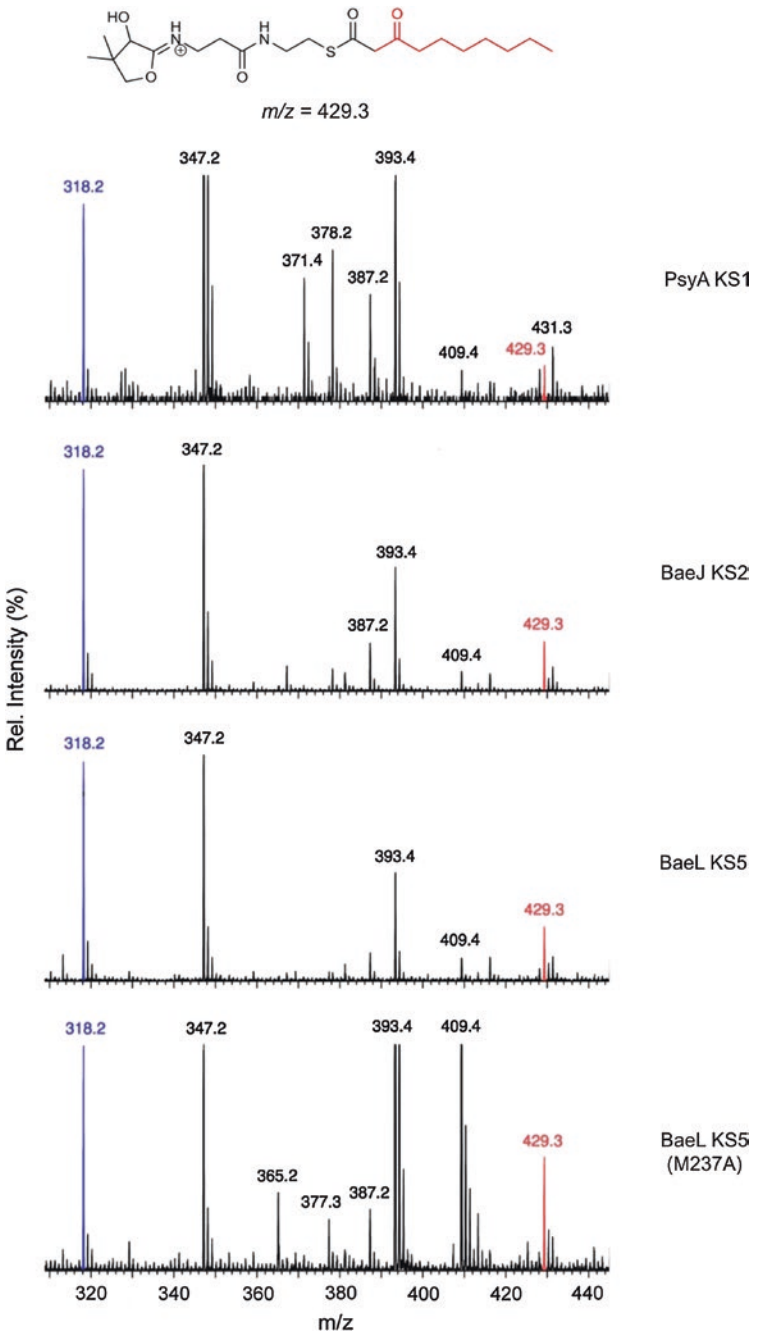

| | |
|---|---|
| Number of residues in favoured region (~98.0% expected) | : 266 (92.0%) |
| Number of residues in allowed region (~2.0% expected) | : 20 (6.9%) |
| Number of residues in outlier region | : 3 (1.0%) |

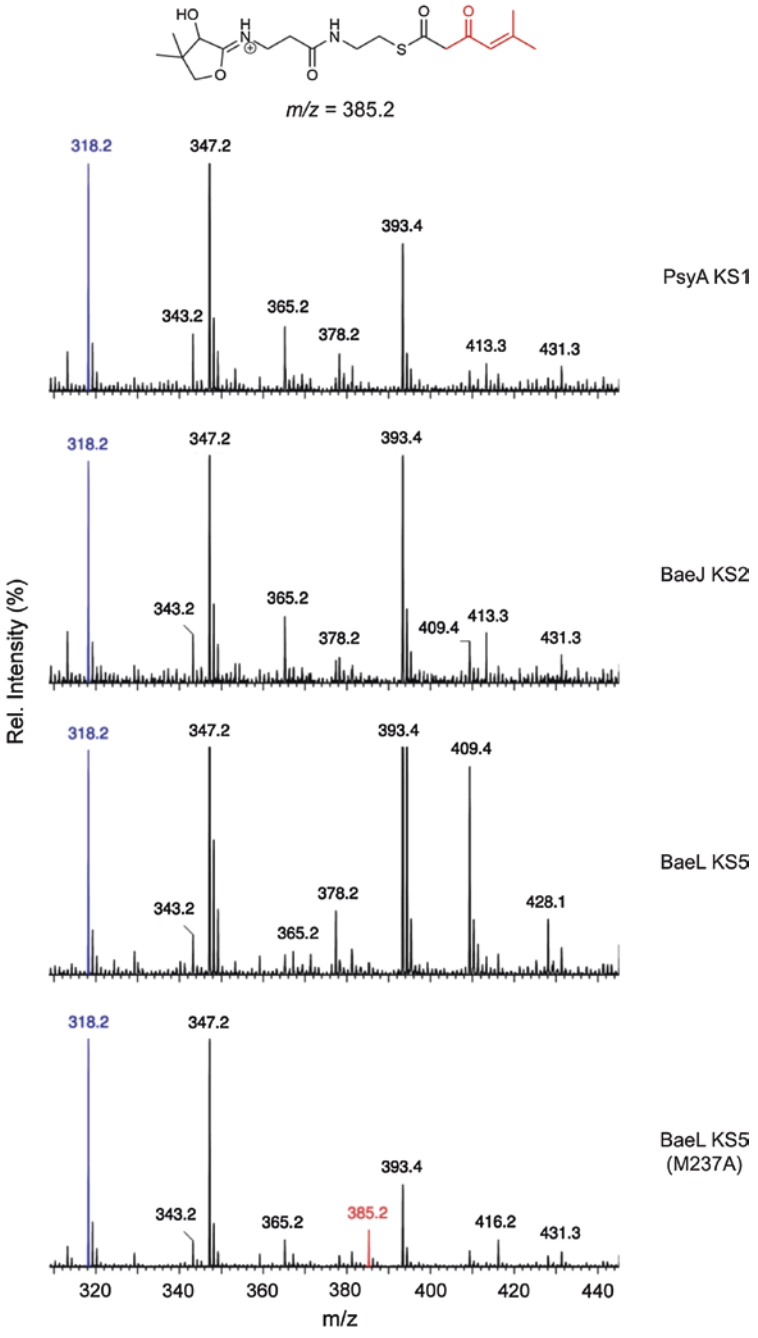
## MS/MS Spectra of Elongation Reactions

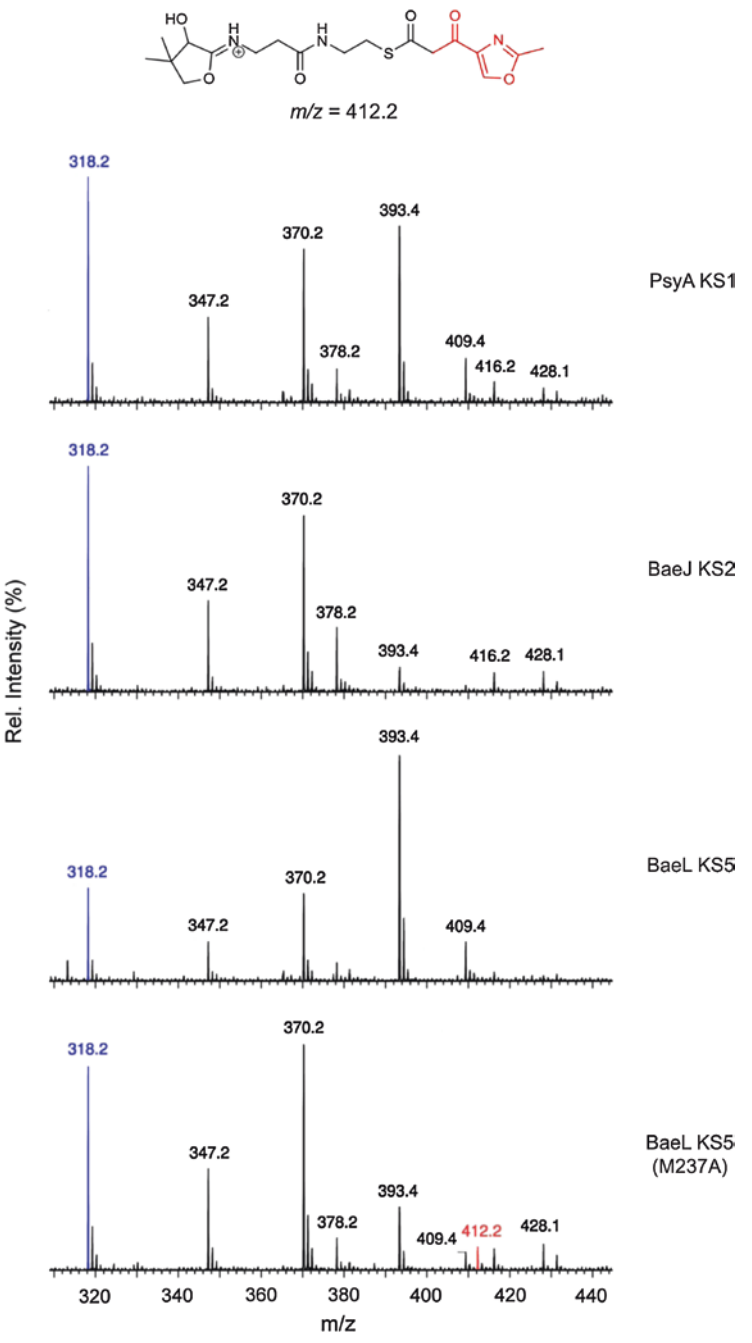


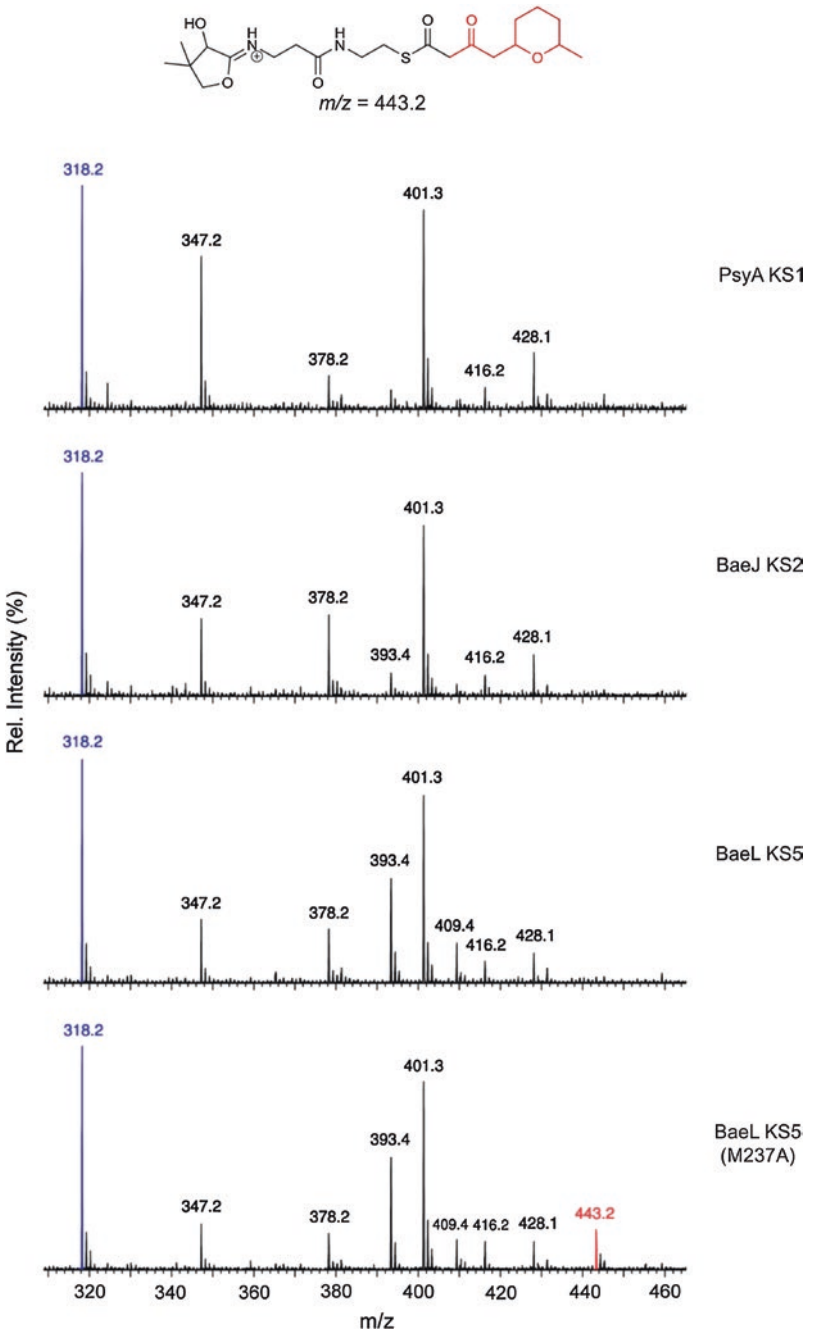

HO
H
N
H
N
O
O
S
O
O
m/z = 373.2
318.2
331.2
347.2
393.4
378.2
373.2
409.4
428.1
PsyA KS1
318.2
347.2
331.2
373.2
378.2
393.4
416.2
431.3
BaeJ KS2
318.2
331.2
347.2
393.4
373.2
378.2
428.1
409.4
BaeL KS5
318.2
347.2
331.2
373.2
393.4
416.2
431.3
BaeL KS5
(M237A)
Rel. Intensity (%)
320
340
360
380
400
420
440
m/z


m/z = 399.2
PsyA KS1
BaeJ KS2
BaeL KS5
BaeL KS5
(M237A)
Rel. Intensity (%)
m/z
318.2
347.2
357.2
365.2
393.4
381.19
409.4
378.2
399.2
431.3
428.1
413.3
437.2
320
340
360
380
400
420
440


m/z = 429.3
PsyA KS1
BaeJ KS2
BaeL KS5
BaeL KS5 (M237A)
Rel. Intensity (%)
m/z
318.2
347.2
393.4
378.2
371.4
387.2
409.4
431.3
429.3
365.2
377.3
320
340
360
380
400
420
440


m/z = 385.2
PsyA KS1
BaeJ KS2
BaeL KS5
BaeL KS5 (M237A)
Rel. Intensity (%)
m/z
318.2
347.2
393.4
343.2
365.2
378.2
413.3
431.3
409.4
428.1
385.2
416.2
320
340
360
380
400
420
440


m/z = 412.2
PsyA KS1
BaeJ KS2
BaeL KS5
BaeL KS5 (M237A)
Rel. Intensity (%)
m/z
318.2
347.2
370.2
378.2
393.4
409.4
412.2
416.2
428.1
320
340
360
380
400
420
440


m/z = 443.2
PsyA KS1
BaeJ KS2
BaeL KS5
BaeL KS5 (M237A)
Rel. Intensity (%)
m/z
318.2
347.2
378.2
393.4
401.3
409.4
416.2
428.1
443.2
320
340
360
380
400
420
440
460