

**The Cystobactamide pathway:  
Enzymological investigation of  
unusual non-ribosomal peptide  
biosynthesis mechanisms**

Dissertation  
zur Erlangung des Grades  
des Doktors der Naturwissenschaften  
der Naturwissenschaftlich-Technischen Fakultät  
der Universität des Saarlandes

von  
**Bastien Schnell**

Saarbrücken

2019

Tag des Kolloquiums: 02/07/2019

Dekan: Prof. Guido Kickelbick

Berichterstatter: Prof. Rolf Müller

Prof. Rita Bernhardt

Vorsitz: Prof. Uli Kazmaier

Akad. Mitarbeiter: Dr. Maksym Myronovskyi

Die vorliegende Arbeit wurde von Februar 2014 bis März 2019 unter Anleitung von Herrn Prof. Dr. Rolf Müller am Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) angefertigt.



## Acknowledgments

First of all, I would like to express my deepest gratitude to my supervisor Prof. Rolf Müller for the opportunity he granted me to complete my PhD thesis in his research group at the HIPS but also for his insightful tutoring and for the fruitful scientific discussions we could have. It has been a pleasure to work in the incredible research environment we had in the MINS group.

I would also like to thank Prof. Rita Bernhardt who kindly accepted to be my second supervisor and reviewed my dissertation.

I want to express my gratitude to all of the former and current MINS group members for both the scientific insights they could share with me and the friendly work environment, especially Dr. Sascha Baumann for his tutoring during the beginning of my thesis, Dr. Chengzhang Fu for reviewing the manuscript and Dr. Jesko Köhnke for the insights and critical point of view he was always able to provide. I would also like to thank the analytical team for their help with mass spectrometry, especially Dr. David Auerbach, Patrick Haack and Daniel Sauer. I would like to thank Dr. Antoine Abou Fayad and Dr. Emilia Oueis for their patient guidance in synthetic chemistry. Thank you especially to Domen Pogorevc, Asfandyar Sikandar, Alexander Popoff and Katarina Cirnski who have become more than colleagues during my time at HIPS for being great companions.

I would also like to take this opportunity to express my gratitude to all the teachers and professors that have awakened and nurtured my scientific curiosity during my scholarship and led me to make a PhD.

Finally I would like to thank my friends and my family, especially my parents and my sister for their encouragements and support, and my deepest thanks go to Marie Martin de Lassalle for her unconditional support during these five years.

## Summary

This thesis focused on the investigation of unusual biosynthetic reactions in the Cystobactamide pathway. Cystobactamides present an unusual hexapeptidic backbone mostly constituted from modified *para*-amino benzoic building blocks and a unique methoxy-isoasparaginylyl moiety. Furthermore, the extensive *in trans* tailoring makes this pathway particularly interesting from an enzymological point of view. The present work details the *in trans* biosynthesis of the isopropoxyl decorations carried out on the two last *para*-amino benzoate units through *in vitro* reconstitution of these biochemical reactions including the generation of fully functional non-ribosomal peptide synthetase modules *in vitro*. Furthermore an in-depth biochemical investigation of the unprecedented bifunctional non-ribosomal peptide synthetase domain leading to the isomerization or dehydration of asparagine was performed in parallel with the characterization of the *in trans* hydroxylation and of the shuttling of this moiety. The self-resistance mechanisms of the producer strain *Cystobacter velatus* 34 were also investigated in comparison to the self-resistance mechanism for the related antibiotic Albicidin. Finally, attempts at complete *in vitro* reconstitution of the pathways were performed using the unique heterologous expression platform for non-ribosomal peptide synthetase modules developed during this thesis.

## Zusammenfassung

Diese Dissertation beschäftigt sich mit der Untersuchung ungewöhnlicher biosynthetischer Reaktionen im Cystobactamid-Biosyntheseweg. Cystobactamide haben eine hexapeptidische Grundstruktur, die hauptsächlich aus modifizierten *para*-Aminobenzoessäure Bausteinen und einer einzigartigen Methoxyisoasparaginyeinheit besteht. Außerdem macht die extensive Verwendung von *in trans* enzymatischen Modifikationen diesen Biosyntheseweg aus enzymologischer Sichtweise besonders interessant. Die vorliegende Arbeit untersucht durch *in vitro* Rekonstitution dieser biochemischen Reaktionen, die *in trans* Biosynthese der Isopropoxyl Dekorationen, die auf den beiden letzten *para*-Aminobenzoateinheiten entsteht, einschließlich der Erzeugung voll funktionsfähiger nicht ribosomaler Peptidsynthetase-Module *in vitro*. Außerdem wurde eine umfassende biochemische Analyse der neuen bifunktionellen nicht ribosomalen Peptidsynthetase Domäne durchgeführt, die zur Isomerisierung oder Dehydratisierung von Asparagin führte, parallel zur Charakterisierung der *in trans* Hydroxylierung und des Transfer dieses Moleküls. Die Eigenresistenzmechanismen des Produzenten *Cystobacter velatus* 34 wurden im Vergleich zum Selbstresistenzmechanismus des verwandten Antibiotikums Albicidin untersucht. Letztendlich wurden Versuche einer kompletten *in vitro* Rekonstitution des Biosyntheseweges durchgeführt, mit Hilfe der in dieser Dissertation neuentwickelten heterologen Expressionsplattform für nicht ribosomale Peptidsynthetase-Module.

## Vorveröffentlichungen der Dissertation

Teile dieser Arbeit wurden vorab mit Genehmigung der Naturwissenschaftlich-Technischen Fakultät III, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen veröffentlicht oder sind derzeit in Vorbereitung zur Veröffentlichung:

### Publications

Wang, Y., **Schnell, B.**, Baumann, S., Müller, R., and Begley, T. P. (2017) Biosynthesis of Branched Alkoxy Groups: Iterative Methyl Group Alkylation by a Cobalamin-Dependent Radical SAM Enzyme, *Journal of the American Chemical Society* 139, 1742–1745.

**Schnell, B.**, Baumann, S. and Müller, R. CysL: Investigation of a secondary metabolism related para-aminobenzoate CoA ligase and its feedback inhibition by Cystobactamides, to be published

**Schnell, B.\***, Sikandar, A.\*, Baumann, S., Köhnke, J. and Müller R. Biochemical and Structural investigation of CysF, an *in trans* acting NRPS methyl transferase, to be published

**Schnell, B.**, Auerbach, D., Baumann S. and Müller R. Biosynthesis of the unusual hydroxy iso-asparagine moiety in Cystobactamides by a new NRPS aminomutase domain and its shuttling through an aminoacyl transferase, CysB, to be published

(\*: Authors contributed equally to the work)

### Conference Contributions (Posters and Oral Presentations)

**Schnell, B.**, Baumann, S., Wang, Y., Bradley, T.P. and Müller R. (2016) Cystobactamides: Insights in the biosynthesis of novel antibacterial topoisomerase inhibitors. PhD day of the Faculty of Natural Sciences and Technology, Saarbrücken, Germany.

# Table of Contents

Chapter 1: Introduction .....	1
1 Natural products .....	1
1.1 Secondary metabolism .....	1
1.2 Historical records .....	3
1.3 Challenges in antibiotics drug discovery .....	5
1.4 Myxobacteria.....	6
2 Biosynthesis of natural products .....	10
2.1 Non-ribosomal peptide synthetase .....	11
2.1.1 Core domains .....	12
2.1.1.1 Thiolation domain – Peptidyl carrier protein.....	13
2.1.1.2 Adenylation domain.....	14
2.1.1.3 Condensation domain.....	16
2.1.1.4 Thioesterase domain .....	17
2.1.2 Auxiliary Tailoring domains .....	18
2.1.2.1 Epimerization domain.....	18
2.1.2.2 Oxidation and Reduction domains.....	18
2.1.2.3 N-methyl transferase domain.....	19
2.1.3 Accessory proteins.....	19
2.1.3.1 Phosphopantetheine transferase .....	19
2.1.3.2 Adenylation Domain Activators .....	20
2.1.3.3 Type II thioesterase.....	21
2.2 Polyketide synthases .....	21
2.2.1 Modular type I PKS .....	23
2.2.2 Iterative Type I PKS .....	24
2.2.3 Type II PKS .....	24
2.2.4 Type III PKS.....	25
3 Outline of this work.....	26
3.1 Cystobactamides.....	26
3.2 Cystobactamide biosynthetic gene cluster .....	27
3.3 Biosynthesis .....	28
3.4 PABA Biosynthesis.....	31

4	References .....	33
Chapter 2:	CysL - Investigation of a secondary metabolism related para-aminobenzoate CoA ligase and its feedback inhibition by Cystobactamides .....	45
1	Introduction .....	45
2	Material and methods .....	46
2.1	Cloning, overexpression and purification .....	46
2.2	Coadenylation assays .....	46
2.3	PNP-MESG assay .....	47
2.4	Reaction scale-up .....	48
2.5	Inhibition assay .....	48
3	Results and discussions .....	48
3.1	Activity on pABA and derivatives .....	48
3.2	Reaction kinetics .....	49
3.3	Reaction scale up and purification of CoA derivatives .....	52
3.4	Inhibition effect of Cystobactamides .....	52
3.5	Module separation .....	55
3.6	CysL in the biosynthesis .....	55
4	Conclusion.....	56
5	References .....	57
Chapter 3:	CysF and CysS – Biosynthesis of an uncommon isopropoxyl decoration by iterative methylation.....	59
Introduction	.....	59
Chapter 3.1:	CysF - Biochemical and Structural investigation of an <i>in trans</i> acting NRPS methyl transferase.....	62
1	Introduction .....	62
2	Materials and Methods .....	63
2.1	Cloning, overexpression and purification .....	63
2.2	<sup>13</sup> C methionine feeding.....	63
2.3	<i>In vitro</i> activity assays on CoA .....	64
2.4	CysG cloning, overexpression and purification .....	64
2.5	<i>In trans</i> activity assays on CysG .....	66
2.6	Crystallization and Structure determination.....	66
3	Results and Discussions .....	66
3.1	<i>In vitro</i> activity assays.....	67

3.2	Structural Analysis .....	70
4	Conclusion.....	72
Chapter 3.2:	CysS - Biosynthesis of Branched Alkoxy Groups: Iterative Methyl Group Alkylation by a Cobalamin-Dependent Radical SAM Enzyme .....	73
1	Introduction .....	73
2	Material and Methods.....	74
2.1	Cloning, overexpression and purification .....	74
2.2	In vitro activity assays.....	75
2.3	B12 feeding .....	76
3	Results and Discussions .....	76
3.1	<i>In silico</i> analysis and protein expression.....	76
3.2	Iron sulfur cluster regeneration .....	77
3.3	In vitro activity reconstitution .....	78
3.4	B12 feeding .....	79
4	Conclusion.....	80
	References .....	81
Chapter 4:	From CysH to CysK - The biosynthesis of hydroxy-isoasparagine and its incorporation into the peptide chain.....	85
	Introduction .....	85
Chapter 4.1:	CysH and CysJ - <i>In cis</i> and <i>in trans</i> tailoring working hand in hand for hydroxy-isoasparagine biosynthesis.....	88
1	Introduction .....	88
2	Material and Methods.....	91
2.1	Cloning, overexpression and purification .....	91
2.2	Adenylation assay .....	92
2.3	CysJ Activity assays.....	92
2.4	Unloading assays.....	92
2.5	Synthesis of the di(ethylcarbonyl)asparaginyldicysteamine references .....	93
2.6	Protein MS analysis.....	93
2.7	Asparagine feeding.....	94
3	Results and Discussions .....	94
3.1	CysH - Adenylation and thiolation domain activity.....	94
3.2	CysJ - <i>in silico</i> analysis .....	97
3.3	Unloading assays.....	97

3.4	Protein MS.....	99
3.5	CysH - <i>In silico</i> analysis.....	101
3.6	Labeled Asparagine feeding.....	103
3.7	Dehydration of asparagine .....	104
3.8	CysH - Substrate screening .....	105
3.9	Mode of Action .....	106
3.9.1	Aminomutases .....	106
3.9.2	Dehydratases.....	108
3.9.3	Mechanism proposal.....	108
3.9.4	Stereochemistry of the isomerisation .....	109
4	Conclusion.....	110
Chapter 4.2: CysB - A new amino acyl transferase shuttles modified asparagine residues between a standalone NRPS module and the assembly line .....		111
1	Introduction .....	111
2	Material and Methods.....	112
2.1	Cloning, overexpression and purification .....	112
2.2	Activity assays.....	113
2.2.1	Ping - loading from H to B .....	113
2.2.2	Pong - loading from B to K3 .....	114
2.3	Protein MS.....	114
2.4	Unloading .....	114
3	Results and Discussions .....	115
3.1	Ping - loading from CysH to CysB .....	115
3.2	Pong – loading from CysB to CysK3.....	117
3.2.1	CysK3 overexpression.....	117
3.2.2	Protein MS CysK3.....	117
4	Conclusion.....	118
References .....		119
Chapter 5: CysP - Cystobactamide resistance protein .....		125
1	Introduction .....	125
2	Materials and methods .....	126
2.1	Cloning, overexpression and purification .....	126
2.2	MIC measurements .....	127

2.3	Binding studies - SEC .....	127
2.4	Binding studies - Fluorescence .....	127
3	Results and discussions .....	128
3.1	<i>In silico</i> analysis .....	128
3.2	MIC comparisons .....	129
3.3	Binding with tripeptides .....	131
4	Conclusion.....	133
5	References .....	134
Chapter 6: CysK and CysG - Study of the assembly line and attempts at <i>in vitro</i> reconstitution		137
1	Introduction .....	137
2	Material and methods .....	138
2.1	Cloning, overexpression and purification .....	138
2.2	Loading and condensation assays .....	140
2.3	Protein MS.....	140
2.4	Synthesis of the SNAC esters.....	140
3	Results and discussions .....	141
3.1	Linker modelling .....	141
3.2	CysA.....	143
3.3	CysE .....	144
3.4	CysK sequencing.....	145
3.5	Loading Assays .....	146
3.6	Condensation reactions .....	156
4	Conclusion.....	158
5	References .....	159
Chapter 7: Discussions.....		161
1	General summary of this work .....	161
2	Cystobactamide biosynthesis: a canonical NRPS assembly line? .....	162
3	General discussions .....	167
3.1	Challenges in the <i>in vitro</i> study of NRPS machinery.....	167
3.2	Limits of NRPS engineering and <i>de novo</i> NRPS biosynthesis .....	169
4	Conclusion and Outlook.....	175
5	References .....	177



# Chapter 1: Introduction

## 1 Natural products

The term “Natural products” is commonly used to designate a large group of chemicals of low molecular weight (<3000 Da) which are produced by a living organism. These compounds of natural origin exhibit a huge variety in structure and biological activity. A commonly accepted definition of natural products was formulated by Williams and al.(1) as a “substance appearing to have no explicit role in the internal economy of the organism that produces it”; this definition directly relates natural products to secondary metabolism.

### 1.1 Secondary metabolism

On a biochemical level it is common to make a distinction between primary and secondary metabolism. The primary metabolism encompasses the biochemical pathways that are essential for the survival of the cell and such as the Krebs cycle, metabolism of fatty acids and carbohydrates, protein synthesis, nucleic acid metabolism, etc. These biochemical processes are conserved throughout Evolution and are present in all living organisms. On the other hand secondary metabolites are not essential for the survival of the organism that produces them but they often contribute to its fitness.(2) They are produced by specific pathways which are more or less characteristic for certain groups of organisms and can be used for taxonomic purposes. The distinction between primary and secondary metabolism was first discussed by Kossel in 1891 in the founding “Über die chemische Zusammensetzung der Zelle” (3):

*„Ebenso wie die mikroskopische Forschung dahin gelangt ist, daß sie die Zelle alles unwesentlichen Beiwerkes entkleidet hat, daß sie das Gehäuse und die in ihr aufgespeicherten Reservestoffe von den eigentlichen Trägern des Lebens zu trennen weiß, so muß auch die Chemie versuchen, diejenigen Bestandteile heraus zu sondern, welche in dem entwicklungsfähigen Protoplasma ohne Ausnahme vorhanden sind, und die zufälligen oder für das Leben nicht unbedingt nötigen Zellstoffe als solche zu erkennen. Die Aufsuchung und Beschreibung derjenigen Atomkomplexe, an welche das Leben geknüpft ist, bildet die wichtigste Grundlage für die Erforschung der Lebensprozesse. Ich schlage vor, diese wesentlichen Bestandteile der Zelle als primäre zu bezeichnen, hingegen diejenigen, welche nicht in jeder entwicklungsfähigen Zelle gefunden werden, als sekundäre.“*

*Just as microscopy has succeeded in stripping the cell of its nonessential accessories and in separating its casing and stored reserves from the actual life carriers, chemistry must attempt to separate those compounds which always are present in the protoplasm which is capable of developing and to recognize the substances which are either incidental or not absolutely necessary for life. Finding and describing those atom complexes to which life is bound is the most important basis for the investigation of life processes. I propose calling these essential components of the cell primary components, and those which are not found in every cell capable of developing, secondary components.*

During the century that followed it appeared that the line between primary and secondary metabolism was more blurred than it looked in 1891. Hence certain polysaccharides or fatty acids are secondary metabolites produced by primary metabolism pathways and the same goes for Ribosomally synthesized and post-translationally modified peptides (RIPPs), which are peptidic natural products produced by the ribosome. Furthermore the primary metabolism is usually the provider of the building blocks used by secondary metabolism pathways to synthesize natural products. However, a certain number of features can be used to define a secondary metabolite such as a restricted distribution among species and a biosynthesis performed by specific enzymes encoded separately from the primary metabolism genes.(4) The expression and the activity of these enzymes are often under strict control from environmental factors.

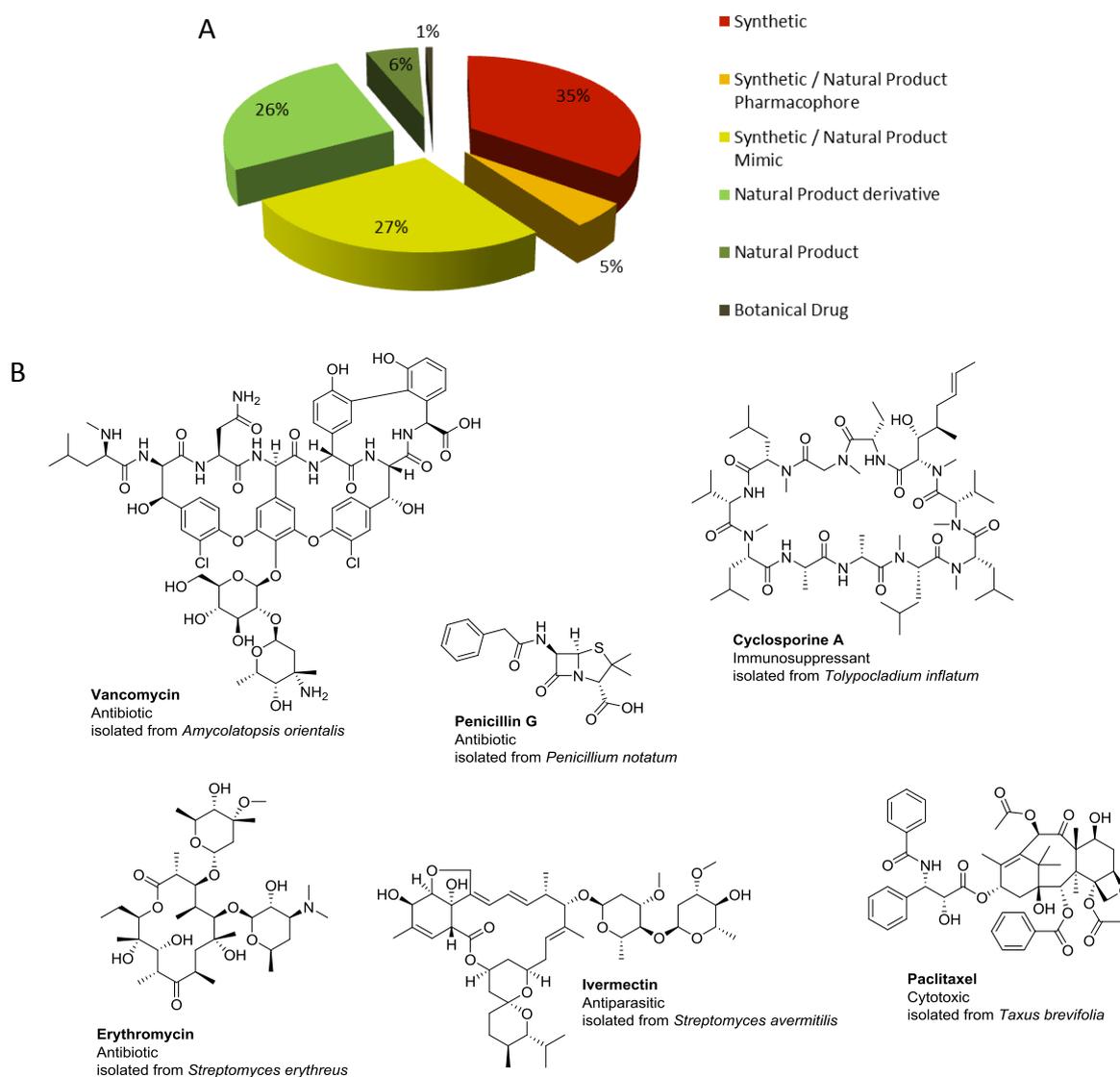


Figure 1.1: A-Distribution of newly approved drugs between 1981 and 2014; B-Selected examples of natural products used in clinic

From an evolutionary point of view the development of secondary metabolism is still an open question, indeed the complexity of the biosynthesis of most natural products is dazzling and such pathways are often very energy consuming for the organisms that have them. These features support the idea of a strong selective pressure on organism to develop such pathways and would imply an important fitness gain for them. However, most side products and some secondary metabolites are devoid of apparent activity and would thus not provide any fitness gain. Following basic evolutionary laws such pathways should very fast be silenced to prevent fitness loss through the unnecessary energy cost, stopping *de facto* their evolution in new compounds. A number of authors have discussed this matter and proposed hypothesis on to why natural products are made and how secondary metabolism pathways would arise.(1,2,5,6)

Natural products exhibit an almost unlimited chemical diversity which makes them a crucial class of chemicals that is used for many applications ranging from drugs and cosmetics to agrochemicals or biofuels. However, secondary metabolites are often produced in low yield incompatible with an industrial use of the compounds. Secondary metabolites are the outcome of millions of years of evolution to adapt to specific targets, this gives a high probability of finding pharmacophores targeted at a specific biological process. This key feature makes natural products the best targets to find new bioactive chemicals (7), while combinatorial organic synthesis plays an important a role of “accelerated evolution” in order to explore the chemical space around a newly found natural product. This interdisciplinary approach combining natural product discovery and medicinal chemistry to enhance the activity and pharmacological properties of a new found compound is still a key aspect of drug research since the middle of the twentieth century.(8) Hence nowadays 50 % of all the small molecule drugs on the market are natural products or synthetic derivatives of a natural products pharmacophore.(9) (Figure 1.1)

## 1.2 Historical records

Throughout the ages, people have relied on natural sources to prevent and cure diseases. Evidence for the use of plant material date back to the Middle Paleolithic age some 60 000 years ago.(10) However the first historical records are clay tablets from Sumeria and Mesopotamia, written in cuneiform. These texts, which date back to 3000BC and 2600BC, describe the use of over 200 plants among which oils from *Cedrus* species (cedar) and *Cupressus sempervirens* (cypress), *Glycyrrhiza glabra* (licorice), *Commiphora* species (myrrh), and *Papaver somniferum* (poppy juice) (11) that are still in use today for the treatment of coughs and colds, parasitic infections, inflammation or pain. The first known extensive text is the Ebers Papyrus that dates back to 1500BC and gives a precise inventory of ancient Egyptians medicinal knowledge;(12) the papyrus describes a pharmacopoeia of 811 complex prescriptions used in contexts ranging from the migraine to parasitic or bacterial infections. Extensive historical records exist for traditional Chinese medicine for which early evidences date back to 1100BC with the *WuShi’Er BingFang* (13) and for the Ayurveda with the *Sushruta Samhita* (14) for which the earliest contributions are estimated around 1000BC.

In the western world the ancient Greek physician and philosopher Hippocrates (460-377BC) is commonly considered the “father of medicine”. He left a corpus describing about 380 remedies mainly of plant origin classified by physiological action.(15) This medical knowledge was later extended by botanical descriptions of Theophrastus (371-287BC) in *Peri Phytōn Historia*.(16) A few centuries later, Celsus (25BC-50AD) with *De Medicina* (17) and Dioscorides (40-90AD) with *De Materia Medica* (18) wrote two historical texts describing more than thousand medicinal products as well as anatomical descriptions, medical treatments and hygiene rules that were used for more than a millennium. The ancient Greek and Roman medical knowledge could be conserved in monasteries across Europe but the Arabic and Persian world extended this knowledge during the medieval period. The Persian philosopher Avicenna (980-1037AD) compiled *al-Qānūn fī at-Tibb* a medical encyclopedia considered by many as “the final codification of all Greco-Roman medicine” which was used until the end of the sixteenth century.(19)

While the pharmacological knowledge in the seventeenth and eighteenth centuries profited a lot from the exploration of the new world and the addition of tropical plants to the pharmacopoeia it is at the beginning of the nineteenth century that the modern era of Natural product chemistry started. The development of chemistry allowed the isolation and identification of active compounds of medicinal plants that were used for centuries. Thus, Morphine was isolated in 1805 by Setürner and identified as the active compound of opium poppy, *Papaver somniferum*, which was used for thousands of years.(20) A few years later in 1820 Pelletier and Caventou isolated quinine from *Cinchona officinalis* tree bark (21) which was used as a powerful antimalarial drug since the early seventeenth century in Europe and probably since many centuries in South America. As early as the end of the nineteenth century the idea of chemical structure was accepted and allowed for further development of natural compounds. Thus salicylic acid, the active compound from the White willow tree bark, *Salix alba*, which was already mentioned in Sumerian texts and isolated in 1828 by Buchner (22) could be further improved though acetylation to reduce its adverse effects in 1853 by Gerhardt.(23) The drug was brought to the market by Bayer in 1899 under the name Aspirin and is still in use today in the same form.

While most of the natural products used throughout the ages were of plant origin the development of microbiology in the late nineteenth century by Pasteur, Koch and others allowed access to a new and unexpected source of drugs: microorganisms. The microbial drug era was started in 1928 when Alexander Flemming discovered in a mold contaminated petri dish seeded with *Staphylococcus aureus* that a compound produced by the mold was killing the bacteria.(24) Investigation of the mold identified as *Penicilium notatum* lead to one of the most important drugs of the twentieth century: penicillin. Chemical enhancement of the  $\beta$ -lactam structure and discovery of natural homologues lead to the development of a number of generations of Penicillins, and Amoxicillin is still the most prescribed antibiotic nowadays. Later other antibiotics such as Streptomycin in 1943 (25) and Chloramphenicol in 1947 (26) were discovered from microorganisms leading the way for the “golden age of antibiotics” in the nineteen fifties and sixties when half the drugs commonly used today were discovered.

From the late nineteen seventies on, rediscovery rate increased and market approval for new antibiotics dropped despite continuous effort to explore new sources.(27)

### 1.3 Challenges in antibiotics drug discovery

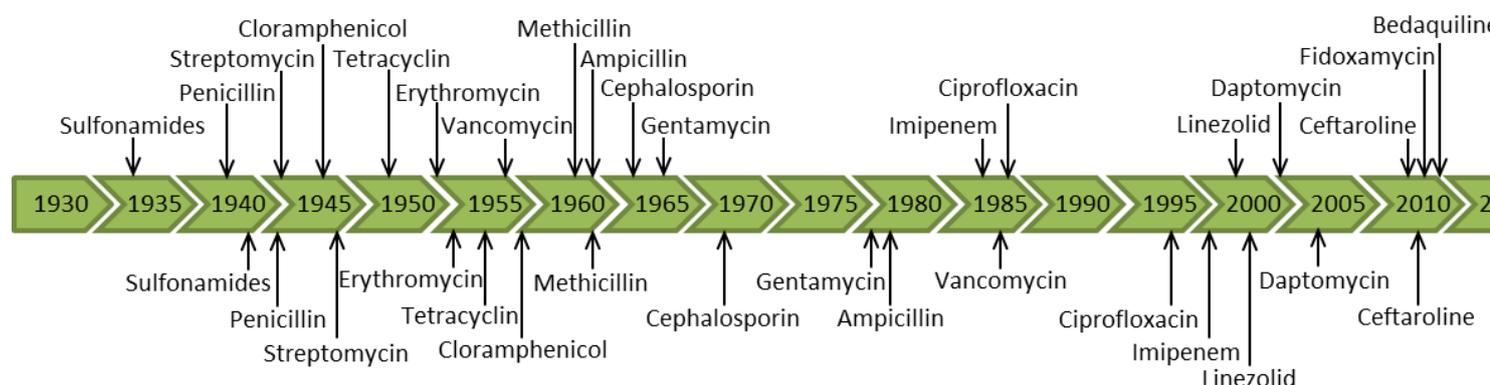
Antimicrobial resistance is one of the major challenges faced today in the clinic. Since their introduction on the market millions of tons of antibiotics have been used in human and veterinary medicine (28), this created a great evolutionary pressure on bacteria to evolve resistance mechanisms against antibiotics. Since the appearance of Penicillin resistant *Staphylococcus* in 1944 (29), only a few years after its introduction to the market, the delay between the introduction of a new antibiotic and the appearance of resistance genes has been incredibly low, even if some antibiotics lasted longer than others, mostly depending on their mode of action.(30) Additionally, horizontal gene transfer can allow bacteria to acquire resistance from other organisms spreading these genes very fast. In 2017 the WHO declared that “*Antimicrobial resistance is a global health emergency that will seriously jeopardize progress in modern medicine*”. These resistances are particularly concerning in the case of nosocomial infections, where the main strains involved from both gram positive and negative origin were dubbed ESKAPE (31) (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter sp*). In the US these organisms are estimated to cause 2 million infections a year.(32)

In parallel to the rise of antibiotic resistances (Figure 1.2), the rate of market introduction of new classes of antibacterial compounds has decreased (33) and newly approved antibiotics are nowadays kept as reserve drugs to use only in last resort. This decline of novelty is mainly due to a failure of industrial high throughput screening strategies and the lack of profitability of antibiotics against “quality of life” drugs, this lead to a gradual and global disinterest of the pharmaceutical industry for antibiotic drugs. On the other hand academic research faces the same issue in the form a high rediscovery rate amongst natural products. This is mainly due to the lack of originality in the organisms that are chosen to be investigated.(34) Indeed most of the research on antibiotics is still done on the three classes of organisms that yielded the most compounds: *Streptomyces* and rare *Actinomycetes*, and microscopic fungi such as *Aspergillus* and *Penicillium*. It is however known that other classes of bacteria have a high biosynthetic potential to produce natural products. These new sources present, however, a major challenge regarding cultivation methods and the vast majority, about 90%, of microorganisms have yet to be cultured in laboratory conditions thus preventing a thorough investigation of their biosynthetic capabilities.(35)

A renewal in the field of natural product research came with the increasing availability of genomic data. Indeed since the discovery of the DNA double helix in 1953 (36) advances in sequencing technology such as Illumina or PacBio sequencing allowed a cost effective access to large amount of genomic data. The amount of data gathered in only a few years is colossal and keeps increasing following Moore’s law.(37) Furthermore, it is noteworthy that secondary metabolism in microorganisms is clustered, meaning that the set of genes responsible for the biosynthesis of a single natural product are gathered in a single

genomic locus simplifying genomic based investigations for new natural products. The first exploration of genomic data revealed that even organisms that were thoroughly studied for many decades such as *Streptomyces coelicolor* which was known to produce five different secondary metabolites at that time still harbored 18 other orphan biosynthetic gene clusters.(38) It soon became clear that the biosynthetic capabilities of many organisms were gravely underestimated. The approach combining classical natural product isolation techniques and molecular biology techniques allowed the investigation of new sources of natural products that are difficult or impossible to cultivate.(39)

Antibiotic market introduction:



Antibiotic resistance development:

Figure 1.2: Timeline of the market introduction of major antibiotics (top) versus the development of resistance in the clinic (bottom)

## 1.4 Myxobacteria

Amongst the understudied bacterial classes, myxobacteria stood out as a potential source for natural products due to their relatively large genomes, often around 9-13 Mb and sometimes more, harboring a lot of secondary metabolites biosynthesis clusters.(40,41) Myxobacteria were named after the Greek word *myxo* which means mucus. They are Gram-negative bacteria forming the order of Myxococcales in the class of  $\delta$ -proteobacteria, and present a high GC content. Myxobacteria display unique multicellular behavior with a cooperative gliding motility forming swarming colonies and a complex lifecycle involving sporulation under starvation. (42) Despite these fascinating properties myxobacteria were relatively understudied and close to nothing was known about them outside the type strains *Myxococcus xanthus* and *Stigmatella aurantiaca* before their ability to produce novel and diverse secondary metabolites was discovered.

Myxobacteria are known for their characteristic and unique “social” behavior which is exceptional amongst prokaryotes. They are ubiquitous environmental organisms that can be found in soil samples, dung of herbivores, barks of trees, rotting woods, freshwater and marine sediments.(43) The vast majority of myxobacteria are mesophilic but they have also been isolated in extreme environments such as the Antarctic ice sheet (44), hot springs (45) or

alkaline peat bogs.(46) Myxobacteria are Predatory organisms and their coordinated gliding motility allows them to form swarms that can overgrow prey organisms that they lyse through a cocktail of antimicrobial secondary metabolites and lytic enzymes. Under nutrient limitation, myxobacteria can undergo complex cooperative cellular morphogenesis to form dormant myxospores within macroscopic multicellular fruiting bodies. The structure and size of these fruiting bodies is specie specific going from simple mounds (e.g. *Myxococcus xanthus*) to complex tree like structures (e.g. *Chondromyces spp.*).(47)

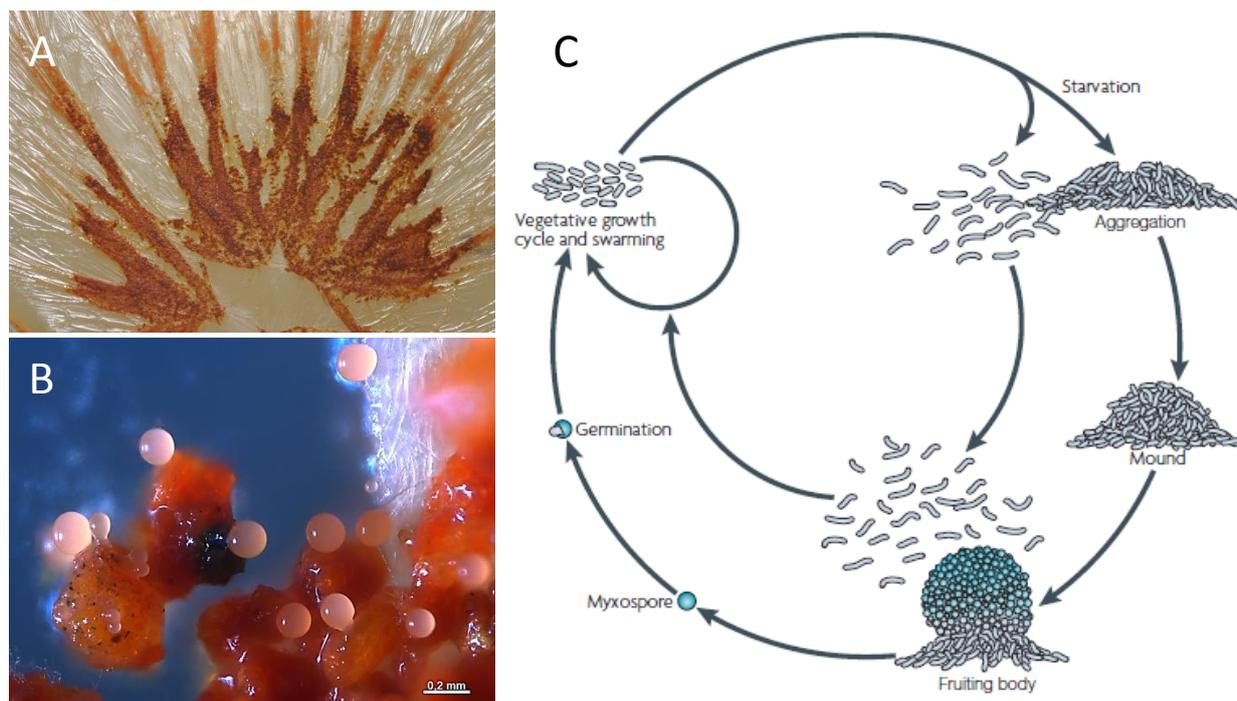


Figure 1.3: A-Cystobacter velatus swarming colony; B- Myxococcus sp. Fruiting bodies; C- *Myxococcus xanthus* lifecycle as described by Zusman et al. (48) Under nutrient limitation, cells will aggregate into macroscopic mounds and fruiting bodies in which the rod shaped cells will undergo morphogenesis to form spherical myxospores.

The taxonomy of myxobacteria is mainly based on the morphological characteristics of vegetative cells, of myxospores, of the swarm colony and especially the shape, size, color and structure of the fruiting bodies. The order of myxococcales currently includes three suborders, 10 families, 29 genera and 61 species.(49)

Myxobacteria exhibit tremendous biosynthesis capabilities, it is common to observe more than 20 clusters in a myxobacteria genome, but this is to be put in perspective with the large size of such genomes. The largest prokaryotic genome known to date belongs to *Sorangium cellulosum* So ce0157-2 with its 14,7 Mb and is estimated to contain 38 secondary metabolite biosynthesis clusters.(50) Myxobacterial products represent about 5% of the known compounds from microorganisms (51), more than 100 compound classes with over 1000 derivatives are known from myxobacteria (52), these natural products often exhibit novel targets and different binding site when compared to more classical *Streptomyces* or

fungal compounds.(53) More than 40% of the myxobacterial secondary metabolites are completely novel structures and most of the limited redundancy is with marine natural products from tunicates, sponges and molluscs raising the question of the origin of these compounds.(54) Myxobacterial compounds are mostly polyketides, non-ribosomal peptides and very often hybrids of both. They often make use of uncommon building blocks such as  $\beta$ -amino acids and other non proteinogenic or highly modified amino acids, on the other hand they rarely show glycosylation which is common in actinomycetes.(55) Myxobacteria also produce metabolites that are unusual in prokaryotes such as steroids or cerebrosides, as well as volatile compounds like Geosmin responsible for their earthy smell.(56,57)

Secondary metabolites from myxobacteria exhibit a very wide range of bioactivities which are thought to be mainly related to predation and competition in their environment like antifungal and antimicrobial compounds but structures with cytotoxic, antiviral or antimalarial activity have also been isolated. Other molecules such as pigments are thought to be related to the intercellular signaling which is essential to myxobacterial social lifestyle. One of the most prominent myxobacterial compounds is Epothilone which is a cytotoxic microtubule stabilizer originally isolated from *Sorangium cellulosum* So ce90. Despite the lack of structural similarities to Paclitaxel they share the same activity and binding site.(58) Chlorotonil is another macrolactone isolated from So ce1525 (59), original data suggested an antibiotic activity against Gram positive bacteria but recent results showed that it was also a very good antimalarial with nanomolar activities against chloroquine resistant *Plasmodium falciparum*.(60) Another antibiotic isolated from *Sorangium cellulosum* So ce2 is the macrolide polyether Sorangicin which is active on the bacterial RNA polymerase.(61) A further classical example of myxobacterial polyketides is Ambruticin; this antifungal cyclopropyl-polyene is one of the first myxobacterial compounds isolated.(62) Myxobacteria are also proficient producers of non-ribosomal peptides such as the antifungal cyclic peptides Vioprolides isolated from *Cystobacter violaceus* 35 (63) and the immunomodulatory Argyrin.(64) As already mentioned, myxobacteria often hybridize polyketide and non-ribosomal biosynthesis pathways as seen in the depsipeptide Aetheramide which is a potent antiviral isolated from *Aetherobacter rufus*.(65) Other examples include Myxothiazol (66) or the polyenes DK xanthene (67) and Disorazol (68). Leupyrrins are a good example of the hybridization capabilities of myxobacteria, as these compounds possess a polyketide moiety and an isoprenoid moiety associated on a highly modified peptidic backbone (69).

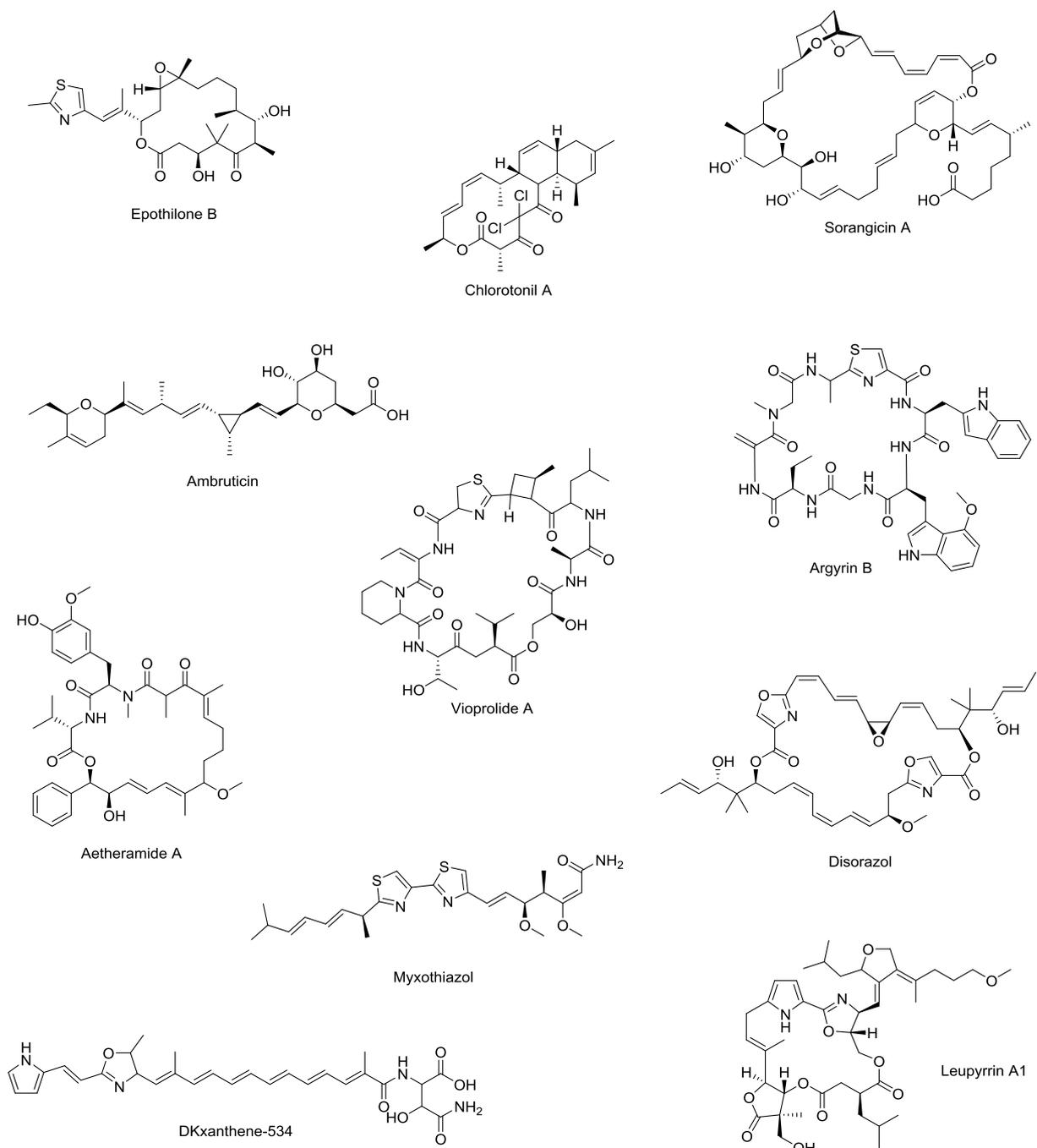


Figure 1.4: Structure of a few examples of representative myxobacterial natural products

## 2 Biosynthesis of natural products

The immense diversity of natural products originates from various biosynthesis pathways that mostly utilize building blocks derived from the primary metabolism. Surprisingly a very limited number of building blocks are used in secondary metabolism, but the evolutionary inventiveness of nature allows for the production of a virtually unlimited chemical diversity through their association and modification.<sup>(70)</sup> The main secondary metabolism pathways are classified and named after the type of building block that they use, which are the acetate involved in the polyketides (PKs) biosynthesis, the amino acids responsible for the formation of the non-ribosomal peptides (NRPs) and alkaloids, and finally the isoprenoids which serve for the terpenes and other isoprenoids. It is noteworthy that these pathways are more or less preferred depending on the organism's taxonomy, thus when plants preferably produce terpenes and amino acid derivatives, microorganisms usually favor PKs and NRPs.

The microbial natural products are primarily produced via two major pathways: the polyketide synthase (PKS) and the non-ribosomal peptide synthetase (NRPS).<sup>(71)</sup> These two systems are quite similar in regard of their general architecture which is based on large multi-modular protein complexes called megasynthases. They are following an assembly line logic where successive domains perform a sequential set of reactions leading to a final product. In this setup each biosynthesis protein is composed of a set of modules each responsible for the integration of a single building block, each module is further subdivided in independent catalytic domains each responsible for the catalysis of a single reaction.<sup>(72)</sup> PKS and NRPS differ, however, in the building blocks that they use and therefore in the chemistry they use to condense and modify these building blocks. Both systems use carrier domains to covalently tether specific building blocks through a thioester linkage on a phosphopantetheine arm. The first cycle begins with the starter unit on module 1 which is condensed onto the second unit present on module 2, the  $n^{\text{th}}$  unit of the growing chain is further condensed in the same fashion onto the  $(n+1)^{\text{th}}$  unit until the product reaches the desired length on the last module.<sup>(73)</sup> There, the intermediate is finally released from the assembly line by a terminal domain that either simply cleaves off the thioester bond or may perform a further reduction, or even a cyclisation of the product on an alcohol or amine group to yield the corresponding lactone or lactam. (Figure 1.5) Further modifications of the product can be performed by so called "tailoring enzymes" which often accompany these multi-modular protein complexes, they can make a wide set of modification either on the final product, on an intermediate during the assembly or on the building blocks themselves to increase the diversity of the metabolites produced by these pathways.<sup>(74)</sup> Classical tailoring reactions include methylation, oxidation or reduction, halogenation, oxidative crosslinking, glycosylation, lipidation, amongst others.

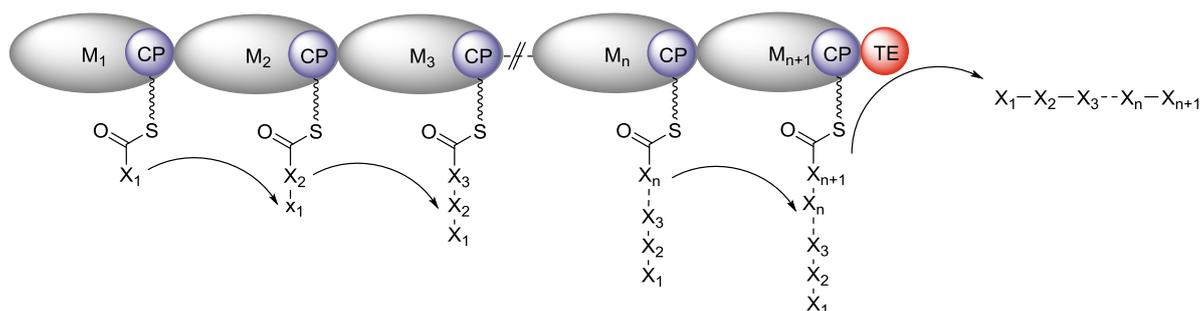


Figure 1.5: General mechanism of megasynthase assembly lines

On a genetic level secondary metabolism in microorganisms is clustered, meaning that the set of enzymes responsible for the biosynthesis of a single natural product are gathered in a single genomic locus where the genes follow one another. This allows the organism to control the production of secondary metabolites through a limited number of promoters often activated depending on environmental factors.<sup>(75)</sup> It also allows the bioinformatic identification of these gene clusters pretty easily due to the characteristic architecture of the NRPS and PKS which can be observed even on a genomic level. Despite the growing number of exceptions, especially in PKS biosynthesis, these megasynthases follow the colinearity rule dictating that the order of domains in the consecutive modules is identical to the order of enzymatic steps during the biosynthesis.<sup>(76)</sup> According to this rule, identification and characterization of the various domains composing an NRPS and PKS driven biosynthesis makes it predictable to a certain extent, allowing for “retro-biosynthetic” analysis of clusters to find the genes responsible for the production of a natural product. The progressing knowledge gathered in the last 20 years about these systems allow nowadays to predict some of the outcome of a biosynthesis encoded in a specific cluster and starts to allow biosynthetic re-engineering of those clusters to modify their product.<sup>(77)</sup> This has tremendous biotechnological implications since it could allow *in fine* to produce various compounds *de novo* through these systems.

## 2.1 Non-ribosomal peptide synthetase

Non-ribosomal peptides have a huge variability in structure and activity, they include a number of clinically important compounds such as antifungals (Bacillomycin), antibacterials (Daptomycin), antivirals (Luzopeptin), antitumorals (Actinomycin D), siderophores (Enterobactin), and immunosuppressants (Cyclosporin). NRPs make use of a multifunctional type of building blocks, amino acids, which present both an amine and an acid function and can thus be assembled in chains using amide bonds formation as condensation reaction. As their name suggests NRPs do not utilize ribosomes to form the peptidic bonds and they do not directly require any nucleic acid based information support such as RNA. Thus they are not limited to the 20 proteinogenic amino acids and can incorporate a huge range of amino acids, current estimates are over 530 different amino acids derivatives.<sup>(78)</sup> Using peptides indeed allows for great opportunities regarding tailoring, and

NRPS make extensive use of  $\beta$ amino acids, D-amino acids, or heterocyclisations, etc. increasing the chemical diversity of this compound class.

A set of conserved domains are characteristic for the minimal NRPS module. The thiolation domain (T) also called peptidyl carrier protein (PCP) carries the phosphopantetheine cofactor on which the growing peptidic chain is linked. The adenylation domain (A) is responsible for the selection of the aminoacyl residue to incorporate, which it activates as AMP adenylate prior to loading on the PCP. Finally the condensation domain (C) makes the peptide bond between the thioester linked carboxyl of the donor aminoacyl and the amine of the acceptor. A number of auxiliary domains can perform *in cis* modifications on PCP bound aminoacyls prior to the condensation, such as epimerization (E), oxidation (Ox), methyl transfer (NMT or CMT) or halogenation. The classical module architecture is C/Cyc-(E)-A-(Ox/Red/NMT)-T.(79,80)

NRPS usually follow the colinearity principle but as every rule presents exceptions some assembly lines present an iterative behavior or even work in a nonlinear fashion. Thus, some authors tried to classify NRPS in 3 subtypes.(81) The type A represents the majority of NRPS and consists of linear assembly lines following the collinearity principle. It is however notable that module skipping, or iteration events can occur on single modules of otherwise canonical assembly lines such as the Myxochromide (82) biosynthesis. The type B corresponds to iterative NRPS assembly line such as the Enniatin (83) or the Enterobactin (84) biosyntheses; in this case the iteration concerns a set of modules which is used to create a cyclic peptide with repetitive sequences. The assembly line usually features an additional T-C module or TE domain responsible for the condensation of the repetitive peptides.(85) The last category is the type C NRPS which regroups nonlinear assembly lines such as the Vibriobactin (86) biosynthesis where the collinearity principle is not respected, modules are often stand alone and even domains can be separate.

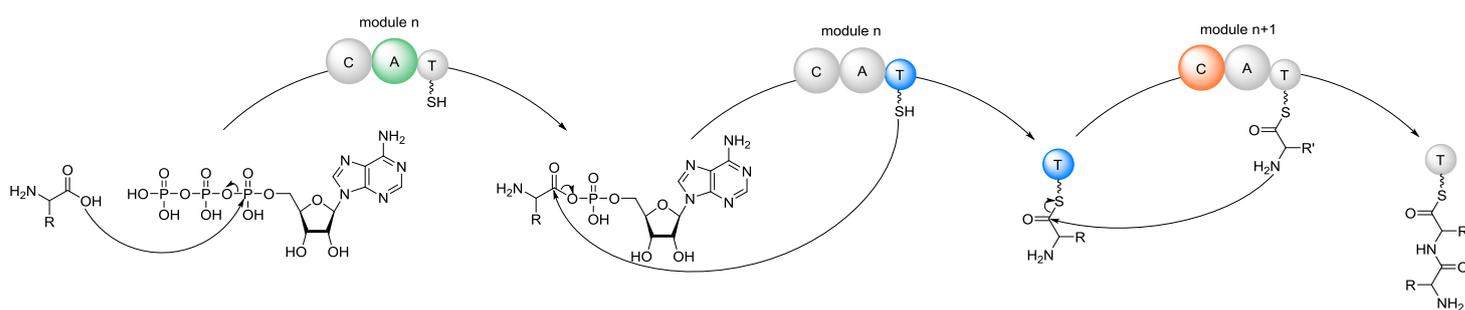


Figure 1.6: NRPS biosynthetic logic

### 2.1.1 Core domains

As already mentioned, the minimal NRPS module is composed of the three core domains C, A and T which are responsible for activation, loading and condensation of aminoacyl building blocks. It is however noteworthy that the first module of an assembly line

is usually lacking the condensation domain. Its presence in the first module is often related to N-acylation of the first amino acid.(80) The last domain for its part is usually featuring a thioesterase domain (TE) that can release the peptide from the megasynthase and possibly cyclize it as lactone or lactam. The consistent conservation of the TE domain in the vast majority of assembly lines and its almost essential nature designates it as a fourth core domain. It is however noteworthy that some assembly lines replace the TE domain by a C domain usually displaying a macrocyclisation activity (87) or by a reductase domain performing a reductive release leading *in fine* to a C-terminal alcohol rather than an acid.(88)

### 2.1.1.1 Thiolation domain – Peptidyl carrier protein

Every enzymatic activity in the NRPS assembly line is centered on the aminoacyl which has to travel from an active site to the other in sequence during the global catalytic cycle. This makes the T domain carrying the aminoacyl the centerpiece of the NRPS assembly line despite its small size of 80 to 100 aa. The PCP has to be activated by a phosphopantetheine transfer from CoA to a conserved serine residue in the conserved motif GX(D/H)S(L/I)(D/K) by a Ppant transferase to yield an active holo-PCP.(89) As carrier protein the T domain is non catalytic but its function implies major structural rearrangement possibilities and large spatial mobility. Indeed the 20 Å of the prosthetic arm are not sufficient to reach the active sites of all the domains since there is already an estimated 47 Å between the active sites of the A and C domains.(90) Furthermore the PCP has to accommodate for at least five distinct positions corresponding to the interaction with the A domain, the C domain of both the  $n^{th}$  module and the  $(n+1)^{th}$  module, a potential auxiliary domain or tailoring enzymes and finally the TE domain for chain release. (Figure 1.7)

Structurally the T domain was originally reported to be composed of a small bundle of 4  $\alpha$  helices,(91) however it was later made clear that the apo and holo forms of the carrier protein had two different conformations; respectively the A-state and the H-state. Additionally a third state common to both apo and holo forms was observed, the A/H state, which turned out to resemble the canonical four-helix carrier protein structure.(92) Furthermore the 20 aa linker between the adenylation domain and the thiolation domain allow for a large 61 Å shift with a 75° rotation around the A domain.(93)

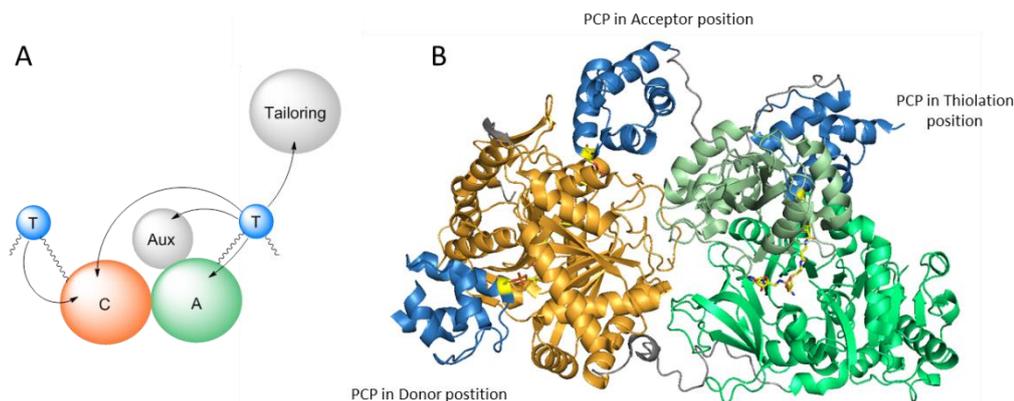


Figure 1.7: A- Schematic representation of the various interactions the T Domain must perform; B- Superimposition of three crystal structures showing the PCP (blue) in three different positions (94,95)

It is noteworthy that in contrast with PKS, it is not clear whether NRPS have a conserved quaternary structure. Dimers of NRPS modules have already been observed, especially through mutational studies on VibF from the type C Vibriobactin biosynthesis.(96) But recent cryo-EM analysis of a dimodular NRPS DhbF from the type B Bacillibactin biosynthesis revealed a relaxed monomeric conformation.(97) These data could explain how the mobility of the PCP can be sufficient to reach sequentially both the C domains in the  $n$ th module and in the  $(n+1)$ th module. But further structural studies especially on type A linear assembly lines would be required to answer the question.

### 2.1.1.2 Adenylation domain

The Adenylation domain is responsible for the initial level of substrate specificity of the assembly line since it performs the activation of a specific amino acid and its subsequent loading on the phosphopantetheine from the carrier protein. Extensive structural studies and mutational analysis to shift substrate specificity have made the A domain one of the best characterized domains in the NRPS system.(98–102) The catalytic cycle takes place in two steps with first the adenylation reaction and subsequently the thiolation. The Adenylation starts with binding of the ATP,  $Mg^{2+}$  and the aminoacyl which is activated as aminoacyl-adenylate via the action of a conserved catalytic lysine. A rearrangement of the protein then allows the Ppant arm of the PCP to come in contact with the aminoacyl-adenylate to catalyze the transfer of the acyl group from the adenylate onto the terminal thiol of phosphopantetheine arm of the PCP domain.(103)

The first crystal structure of an A domain was of the Phenylalanine activating A domain from Gramicidin S synthetase. Extensive work on this structure, additionally to phylogenetic and biochemical investigations allowed for the determination of the 10 key amino acids that form the binding pocket for the aminoacyl to be activated. The so called “Stachelhaus code” (98) was defined which is responsible for the substrate specificity of adenylation domains. It has tremendous implication since most NPRS follow the collinearity

principle, determination of the specificity of the A domains allows a reasonable guess at the final structure from only genomic data. Additionally 10 phylogenetically conserved core motives were determined which allow for the easy identification of adenylation domains on genomic level. Structurally, the A domain consists of a large N-terminal core domain of about 450 aa which is responsible for substrate and cofactor binding, and a small C-terminal sub domain of about 100 aa bearing the catalytic lysine residue in core motif 10. Finally a hinge centered on a conserved aspartate in the core motif 8 allows for a large structural rearrangement of the sub domain over the core domain with a rotation of about  $180^\circ$  leading to a  $21 \text{ \AA}$  shift of the subdomain. This movement is allowing the PCP to come in contact with the A core domain and present its Ppant arm for thiolation.(93)

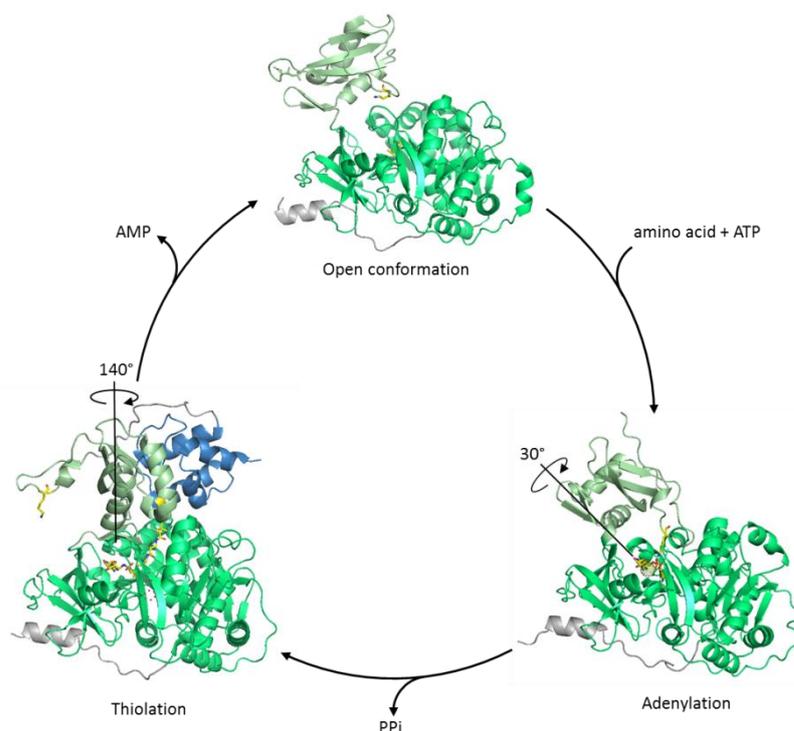


Figure 1.8: Catalytic cycle of the adenylation domain reconstructed from the structures of SrfA (Open) (104), PheA (Adenylation) (98) and EntF (Thiolation) (94) - An “open” conformation of the Asub (light green) allows amino acid, ATP and  $Mg^{2+}$  binding, after which the Asub closes with a  $\approx 30^\circ$  rotation to position the catalytic lysine (yellow) for the adenylation reaction, producing the activated aminoacyl-adenylate. Finally, the Asub rotates  $\approx 140^\circ$  out of the way allowing the PCP domain to bind and the thiol of the Ppant arm to accept the activated amino acyl

Analysis of the adenylation domain substrate recognition initially lead to the idea that it would act as a gate keeper for the NRPS assembly line and that the modification of the specificity of the A domain would allow to engineer NRPS assembly lines on this basis. However, attempts at doing so proved that the specificity of the A domain can indeed be changed but the final yield in modified product usually drops considerably leading to the hypothesis that a secondary gate keeper might be involved, probably through the specificity of the condensation domain.(105,106)

### 2.1.1.3 Condensation domain

The condensation domain is a large monomeric domain of about 450 aa present at the N-terminus of every elongation module. It catalyzes the peptidic condensation between the donor aminoacyl of the  $(n-1)^{th}$  module and the acceptor aminoacyl of the  $n^{th}$  module by bringing both substrates in contact in the active site where the thioester of the donor acyl undergoes nucleophilic attack from the amine of the acceptor aminoacyl.(87) Biochemical investigations on several C domains from the tyrocidine synthetase revealed strict substrate selectivity on the acceptor side and a lower selectivity on the donor side.(107) This finding makes the C domain an important gate keeper in the NRPS machinery. Even if C domains are mostly known to catalyze amide bond formation, several examples of ester bond forming C domains have been reported.(108,109)

Structurally The C domain is composed of a pseudo dimer with two subdomains both homologous to the chloramphenicol acyltransferase arranged in a V shape.(110) Two channels lead to the putative active site harboring a conserved HHXXXDG motif (89) between the two subdomains. The second H was soon designated as the catalytic amino acid through mutational analysis and proposed to be responsible for the deprotonation of the acceptor amine to allow the nucleophilic attack.(111) However further studies and theoretical pKa calculations disproved this hypothesis and lead to the idea that electrostatic interactions were responsible for the catalysis rather than a general acid/base mechanism.(95) Despite its central importance the C domain biochemistry remains elusive mainly because of the absence of structural data of C domains with bound substrates. A recent breakthrough came with the use of small probes covalently tethered to the acceptor site channel which confirmed the second Histidine as catalytic center acting through substrate positioning rather than general base.(112)

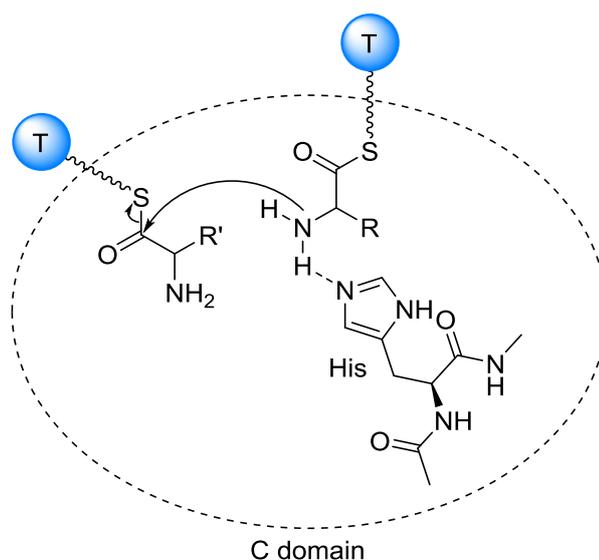


Figure 1.9: peptide condensation mediated by the second histidine of the HHXXXDG motif



## 2.1.2 Auxiliary Tailoring domains

One of the striking characteristics of non-ribosomal peptides is the presence of a huge variety of amino acids. These are generally generated from primary metabolism amino acids which are modified after their loading on the phosphopantetheine arm from the PCP. Even if part of these modifications can be carried out *in trans* by independent tailoring enzymes, a few classical modifications are performed *in cis* such as epimerization,(117) oxidation(118) or methylation.(119)

### 2.1.2.1 Epimerization domain

A hallmark of NRPS is the frequent presence of D configured amino acids, which increase tremendously the stereochemical diversity of NRPs compared to ribosomal peptides. D-peptides can indeed adopt conformations inaccessible to L-peptides that can be crucial for their functions,(120) moreover D-peptides are more resistant to peptidases increasing the stability of NRPs.(121) Even if A domains have been seen to specifically activate D-amino acids previously formed from L-amino acids by an independent racemase, this classical modification is in the vast majority of cases carried out by an epimerization domain (E). This 450 amino acids domain is usually situated at the C-terminus of the module with a C-A-T-E architecture. It performs the base driven deprotonation of the C $\alpha$  leading to racemization and is usually only active *in cis* on PCP bound intermediates. Biochemical investigation on the GrsA-E1 domain leads to a 2:1 ratio of D- to L-phenylalanine with rapid kinetic in both directions.(117) Hence this domain is actually a racemase rather than an epimerase, as the specificity of the following C domain acting as a gate keeper ensures that only D configured amino acids are further processed.

Structurally the epimerization domain is very closely related to the condensation domain with only subtle variations in cross linking regions between the two lobes leading to a different interaction surface for the donor PCP. It however shares the same HHXXXDG motif, but with more variability on last G. Although this hypothesis was disproven in the case of the C domains, Chen and al. (122) state that the second H might after all act as general base in this setup, with help of a nearby conserved glutamic acid as hydrogen donor promoting the D configuration.

### 2.1.2.2 Oxidation and Reduction domains

Oxidation domains (Ox) are found in conjunction with the heterocyclisation domain Cy which performs the ring closure of cysteine, serine or threonine into thiazoline or oxazoline. These heterocycles can be oxidized into a thiazol/oxazol ring such as in the Epothilone or the Bleomycin biosynthesis. The oxidase domain is usually embedded in the adenylation domain just after the core 8 hinge region leading to a C-A<sub>core</sub>-Ox-A<sub>sub</sub>-T architecture. This mechanism is known to be FMN dependent but no structural data exists on such domains.(118) It is noteworthy that a thiazoline/oxazoline heterocycles can also be

reduced into a thiazolidine/oxazolidine ring seen for instance in Pyochelin by a reductase domain.(123)

But reductase domains are usually referred to as domains found at the end of the assembly line in place of the thioesterase domain where they perform a reductive release of the peptidic or polyketidic chains as C-terminal aldehydes.(124) This domain shows sequence homology with short chain dehydrogenase/reductase and uses NADPH as cofactor for the reaction. The aldehyde is very often further reduced to the corresponding alcohol. A number of reductase domains are able to perform a 4 electron reduction to release the thioester as alcohol directly as seen in the Myxochelin A biosynthesis.(88)

### 2.1.2.3 N-methyl transferase domain

Methylation is a common modification for natural products, in the case of non-ribosomal peptides N-methylation of the amide bond has a tremendous importance since it confers resistance to proteases. Even if this type of modification can often be performed *in trans* by independent tailoring enzymes, on precursors or on the product of the assembly line it is also sometimes performed by a dedicated N-methyl transferase domain (NMT). Similarly to the Ox domain it is embedded in the A domain and leads to the same C-A<sub>core</sub>-NMT-A<sub>sub</sub>-T architecture, it is however structurally distinct and closely related to SAM dependent type I methyl transferases.(119)

### 2.1.3 Accessory proteins

A few proteins are often found alongside the NRPS or PKS assembly lines which are important for the smooth functioning of such huge protein complexes. They are designated as accessory proteins and are responsible for the modification of the carrier proteins from *apo* to *holo* form,(125) for the stabilization and activation of the adenylation domains(126) and for the “cleaning” of the assembly line from misloaded acyls or amino acyls.(127)

#### 2.1.3.1 Phosphopantetheine transferase

The first of these accessory proteins is the Ppant transferase which is responsible for post translational modification of carrier proteins. They catalyze the transfer of the phosphopantetheine moiety originating in CoA in an Mg<sup>2+</sup> dependent reaction, converting the inactive apo-PCP to the active holo-PCP.(125) Three types of Ppant transferases are characterized by their substrate specificity and their size. The first is the AcpS type from *E.coli* which is rather small (120 aa) and specific for primary metabolism carrier protein (mainly the fatty acid synthase). The second type is the Sfp Ppant transferase class which is larger with about 240 aa and has a broader substrate specificity since it is able to modify both primary and secondary metabolism carrier proteins; the myxobacterial MtaA belongs to this class. The last type of Ppant transferases is limited to eukaryotic primary metabolism, it takes the form of a domain which is part of the FAS.(128)

Ppant transferases are usually found in the vicinity of secondary metabolism clusters, but a single Sfp type Ppant transferase is often responsible for the activation of all the PKS and NRPS assembly lines of an organism.(128) Thus it cannot be considered to be clustered since it has to be under the control of a different promoter than the gene cluster it is nearest to. The Sfp type Ppant transferase exhibits an  $\alpha/\beta$  fold is divided in two similar halves with the active site in between. A conserved glutamate is responsible for the deprotonation of the carrier protein serine attachment site which can attack the  $\beta$  phosphate of the  $Mg^{2+}$  complexed CoA, releasing AMP while the phosphopantetheine stays covalently tethered to the carrier protein.(129)

### 2.1.3.2 Adenylation Domain Activators

Adenylation domain activator (ADA) proteins are a family of small proteins known to bind to adenylation domains and to promote their activity. ADAs are often referred to as MbtH like proteins named after the first characterized member of the family. These small proteins of about 70 aa length are not systematically encoded in NRPS gene clusters making them accessory, and some adenylation domains have been successfully overexpressed and shown to be active *in vitro* in the absence of an ADA. But when they are present in the cluster they are usually required for a successful expression of the adenylation domains, and thus of the NRPS genes.(130)

*In vivo* ADAs have been shown to be essential for the production of certain metabolites, for instance the deletion of the MbtH type ADA from the Chlorobiocin cluster in *S. coelicolor* decreased the production and further deletion of two closely related homologues from the genome completely abolished it.(131) This hints at the fact that ADAs are not specific for a particular cluster and can interact with a non-cognate assembly line to a certain extent.(132) But they still retain a degree of specificity and have preferred interaction partners even between the different modules of an assembly line. For instance, MbtH itself has been shown to have variable interaction strength with different modules from the Mycobactin assembly line.(133) It is noteworthy that ADAs are sometimes discrete and expressed separately from the clusters they impact but are still necessary for correct expression.(134)

The precise activity and mode of action of these proteins is still quite unclear, but co-crystal structures of the quaternary complex between MbtH type proteins and Adenylation domains revealed a tight complex where the small ADA binds on the underside of the large  $A_{core}$  subunit close to its N terminus.(94) (Figure 1.12) Besides their primary effect, which is to increase the solubility of NRPS modules (126), they have been shown to impact the adenylation domain affinity for its substrate.(94) This is quite puzzling since no structural rearrangement of A domain could be made out through X-ray crystallography and ADAs do not interact with the binding pocket residues. ADAs could thus be considered to be modulators of the activity of assembly lines independent of genetic expression regulation.

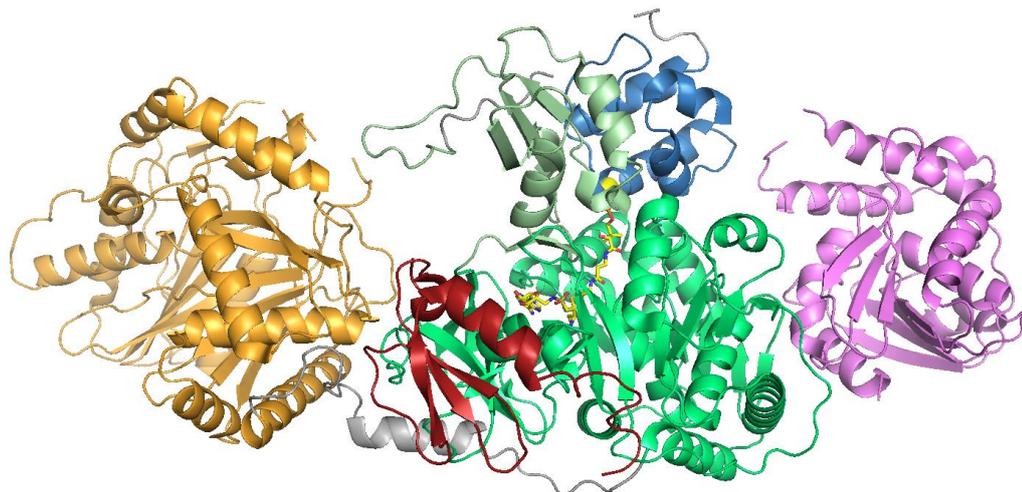


Figure 1.12: The ADA YbdZ (red) binding on the A domain of EntF (green) (94)

It is noteworthy that MbtH type ADAs have been observed as domains integrated to the assembly line in front of the A domain.(135) Recently, a new type of ADA, SibB, was characterized in the Sibiromycin biosynthesis with the same biochemical activity as MbtH but no sequence similarity.(136) Furthermore truncated C domains can also play the role of ADA and be necessary for adenylation activity despite the lack of their own activity.(137)

### 2.1.3.3 Type II thioesterase

In contrast with the type I TE domain which is responsible for terminal peptide chain release, the type II thioesterase is a standalone protein often associated with PKS and NRPS clusters. Its role is to “clean” the assembly line in order to maintain maximal efficiency by removing missprimed acyls from carrier proteins and non-cognate acyls that would have reacted with the free thiol of the Ppant arm.(127) Since 50-75% of the CoA pool available in bacteria is acetylated (138) and the Sfp Ppant transferases are known to not discriminate acylated CoA from free CoA probably leading to a rather high overall misspriming, it is suggested their main function is to offload missprimed CPs. Indeed, the disruption of the TEII in PKS-NRPS clusters usually leads to dramatic yield decreases.

Type II thioesterases are quite original for their inverted substrate specificity since they have to cleave all the incorrect acyl-CPs but avoid removal of the cognate substrates. Structurally they exhibit a rather classical  $\alpha/\beta$  hydrolase fold with a variable lid region composed of  $\alpha$  helices of varying size responsible for substrate recognition.(139)

## 2.2 Polyketide synthases

Just as NRPs, polyketides have a huge variability in activity and structure and even if their biosynthesis is related and based on the same assembly line type megasynthases, PKs make use of a different type of building blocks, acetate, and thus they use a different

chemistry for the condensation reaction. They include various classes of compounds such as macrolides, polyethers, polyphenols, polyenes and enediynes which are all biosynthesized by polyketide synthases. In the context of antibiotic research, polyketides are an important class of compounds with products such as Erythromycin or Doxycycline. Despite their structural diversity, the biosynthesis of all PKs involves the same construction mechanisms based on the formation of C-C bonds by thioester Claisen condensation of malonyl units, or analogues, resulting after decarboxylative condensations in a poly- $\beta$ -keto acyl based on condensed acetate building blocks. (Figure 1.13) The resulting  $\beta$ -ketones can then go through a series of reduction steps leading to alcohol, alkene or alkane. The complete process can be summarized up in a few steps: chain initiation, chain elongation, chain termination and tailoring.<sup>(140)</sup> The polyketide synthase (PKS) derives from the fatty acyl synthase (FAS) and features the same basic components but in a different tertiary and quaternary architecture.

Depending on their architecture and their operating mode PKS are generally classified into three classes. The type I PKS include both the large multimodular complexes responsible for the canonical PKS biosynthesis where each module is usually used only once, and the iterative PKS mainly originating in fungi. Type II PKS are a group where the enzyme complex is dissociable, modules can therefore be used more than once. The type III PKS refers mainly to the chalcone synthase type from plants, it is a group of iterative PKS where the ACP domain is missing therefore the growing acyl chain is not bound to the enzymatic complex.

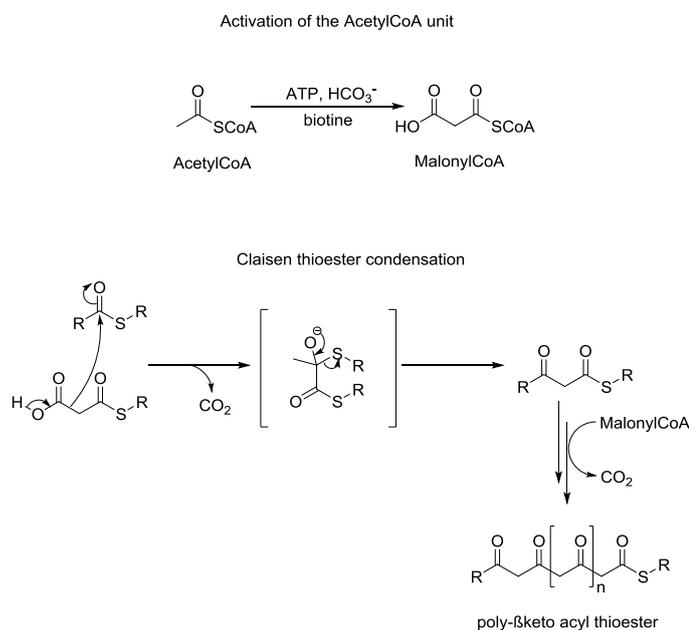


Figure 1.13: Chemical mechanism underlying the PKS biosynthetic logic

### 2.2.1 Modular type I PKS

These biosynthesis complexes represent the archetypical multimodular enzyme complexes that follow the colinearity rule. They are well studied through the type system of the Erythromycin biosynthesis by the 6-deoxy-Erythronolide (6-dEB) synthase (DEBS).<sup>(141)</sup> (Figure 1.14) On a modular PKS the KS/AT/ACP domains responsible for chain elongation and attachment of the synthesized product to the PKS are essential and conserved on each module, whereas the KR/DH/ER  $\beta$ -keto processing domains responsible for the reduction of the currently processed acetate unit are optional.

The Acyltransferase (AT) domain is in charge of the selection of the extender units, typically malonate or methyl malonate which come in a preactivated form as CoA thioesters. Through a ping-pong mechanism the building block is transferred from the CoA on the AT domain and finally on the phosphopantetheine arm carried by the ACP domain. AT domains in the first module often process more exotic acyl or aryl moieties. The keto synthase (KS) domain is responsible for the decarboxylative Claisen thioester condensation itself. Prior to the catalysis of this key reaction the ACP bound intermediate of the  $(n-1)^{th}$  module is transferred on a conserved cysteine of the  $n^{th}$  KS domain via trans thioesterification reaction. The Claisen condensation between the C $\alpha$  of the malonate extender loaded on the ACP of the same module ( $n^{th}$ ) and the thioester of the KS bound intermediate produces a  $\beta$ -keto-acyl thioester on the ACP. The  $\beta$ -keto reductive loop is performed by optional reduction domains: the ketoreductase (KR) domain which is responsible for the reduction of the  $\beta$ -keto function into an alcohol function, the dehydratase (DH) domain reduces the alcohol function and results in a double bond and the enoylreductase (ER) domain which reduces the double bond into a single bond. This PKS is closely related to the fatty acid synthase (FAS) but the latter functions in an iterative fashion and carries out a complete reductive cycle at each step.<sup>(142)</sup>

A special subtype of type I modular PKS lack the AT domain in all modules, they are called trans-AT PKS as opposed to the canonical cis-AT PKS. They feature a discrete AT domain which primes all modules with malonyl-CoA.<sup>(143)</sup>

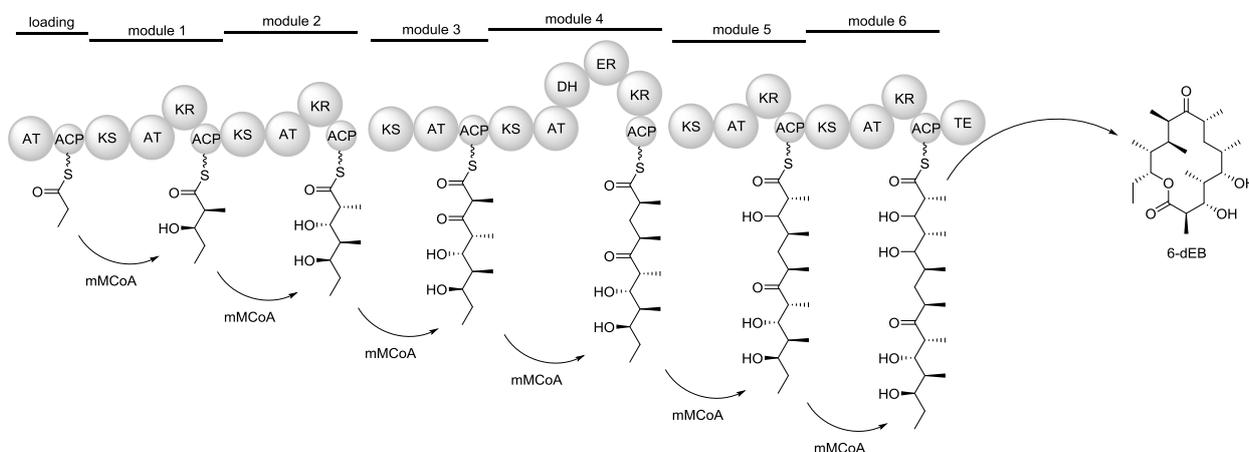


Figure 1.14: 6-dEB biosynthesis by the Erythromycin synthase

## 2.2.2 Iterative Type I PKS

Iterative type I PKS systems are classical for fungal metabolites where the archetype for this system is the lovastatin synthase.<sup>(144)</sup> (Figure 1.15) However they have also been observed in bacteria in which they synthesize small aromatic compounds or polyenes.<sup>(82)</sup> They use the same modules as the modular PKS but possess only one set of domains which can include or not the  $\beta$ -keto reductive loop, depending on this they will be further subdivided into non-reducing, partially reducing or fully reducing PKS. Even if the enzyme complex acts in iterative mode the presence of the reduction domain does not imply that they will perform their activity on each cycle. Therefore the reduction degree of each unit can vary. The factors determining the chain length and the reduction degree remain to be elucidated. In addition, it has been observed that in some modular type I PKS single modules can be used iteratively (e.g. Stimatellin, Borrellidin).

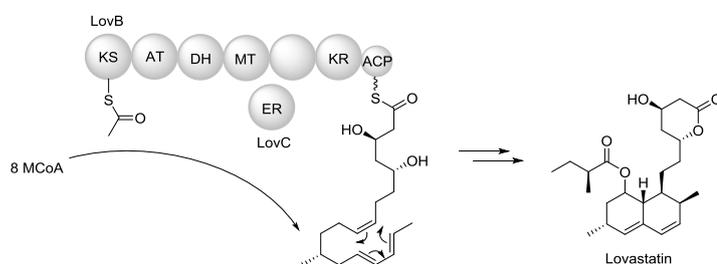


Figure 1.15: Lovastatin biosynthesis

## 2.2.3 Type II PKS

Iterative type II PKS are found mainly in actinomycetes where they are much more common than iterative type I PKS. They are mostly known for the biosynthesis of polyphenols which include compounds such as Tetracyclines or Doxorubicin.<sup>(145)</sup> (Figure 1.16) In this case the minimal PKS is consisting of an ACP to hold the growing acyl chain and two KS domains;  $KS\alpha$  is performing the condensation whereas  $KS\beta$  is known as the chain length factor (CLF) and dictates how many cycles the system will perform. Accessory domains such as a ketoreductase (KR), cyclase (CYC) or aromatase (ARO) can be part of the machinery and are determinant for the folding pattern of the poly- $\beta$ -keto acyl chain.<sup>(146)</sup>

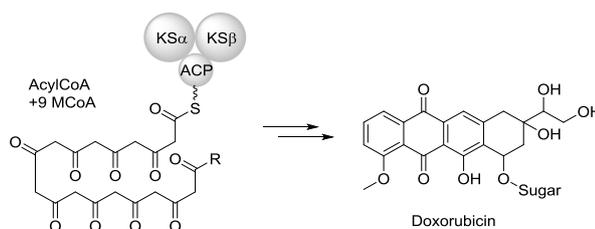


Figure 1.16: Doxorubicin biosynthesis

### 2.2.4 Type III PKS

This system is principally encountered in plants where they are known as the Chalcone synthase family (Figure 1.17). They are mechanistically divergent from the other systems since a single multifunctional KS analogue is selecting a specific starter unit CoA and using malonyl CoA extender units to form a poly- $\beta$ keto acyl chain that internally cyclizes depending on the size of the cavity within the enzyme. No ACP, AT or reducing domains are involved and the whole biosynthesis is performed on CoA bound intermediates.<sup>(147)</sup> Type III PKS were originally thought to be exclusive for plant metabolism but they were later also found in bacteria.<sup>(148)</sup>



Figure 1.17: Naringenin chalcone biosynthesis

### 3 Outline of this work

The primary objective of this work is to unravel the complex biosynthetic mechanisms that underlie the production of the recently isolated myxobacterial compound family Cystobactamides. We chose to clone and overexpress the biosynthesis enzymes in order to characterize their biochemistry individually with the ultimate goal of an *in vitro* biosynthesis reconstitution.

#### 3.1 Cystobactamides

Cystobactamides are a family of linear hexapeptides produced by different myxobacteria of the Cystobacterinae suborder.<sup>(149)</sup> They were first isolated in *Cystobacter velatus* 34 and *Cystobacter ferrugineus* 23 but a few other producers could be identified (*Corallolococci* and *Myxococci*). Cystobactamides 919-1, 919-2 and 507 are the major compounds in Cb v34 extracts, the general structures feature five para-aminobenzoic acids and a highly unusual isoasparagine moiety (pABA-pABA-IsoAsn-pABA-pABA-pABA). In addition the N-terminus is oxygenated into a nitro group, the isoasparagine is usually methoxylated on the  $\alpha$  position, the pABA unit in position 4 is 2,3-hydroxylated while the one in position 5 is only 3-hydroxylated. The 2-hydroxy groups of pABA 4 and 5 are usually ipsopropoxylated except in Cystobactamide 864-2 where they are methoxylated. Cystobactamides 920-1 and 920-2 have been observed in smaller amounts, they feature isoaspartate and aspartate linkers instead of isoasparagine and asparagine.

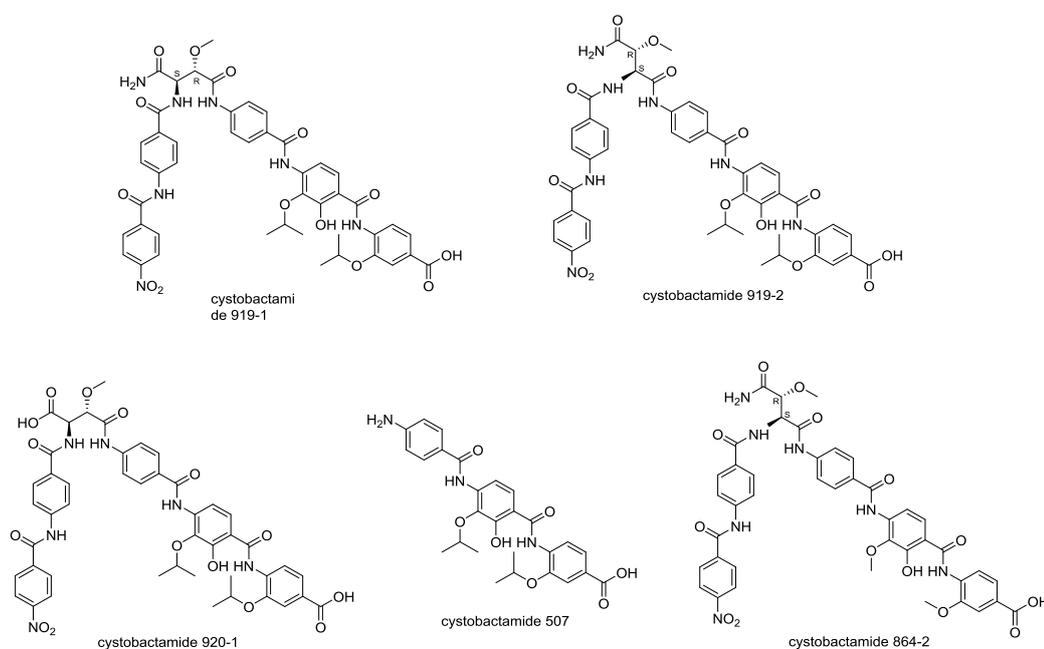


Figure 1.18: Structure of the main Cystobactamide derivatives

Cystobactamides are antibiotics targeting the bacterial gyrase which is already a classical mode of action for antibiotics, with clinically used compounds such as the fluoroquinolones, however no cross resistance has been observed with these implying that they use different binding sites.(149) Biologic activity assays showed a strong antibacterial activity of the Cystobactamides primarily against gram-negative species but also against a few gram-positive organisms. Furthermore, some of the compounds also showed a remarkable activity on *Pseudomonas aeruginosa*. This makes this class of compounds to potent broad spectrum antibiotics with a potential to reach the market someday.

The main structural determinant in the activity seems to be the linker type. The presence of an aspartate type linkers rather than asparagine lead to abolishment of the activity,(150) since these derivatives are minor in the production profile there is a chance that they might be purification artefacts originating in spontaneous deamidation. Cystobactamide 919-2 featuring the asparagine linker shows nanomolar activity against a wide spectrum of pathogenic organisms, while 919-1 which features the isoasparagine retains broad spectrum activity it shows significantly decreased activity with about 100 fold higher MIC.(149) Despite this fact it is noteworthy that 919-1 is the major compound observed in natural producers (ratio of about 4:1), furthermore this derivative is more complex than 919-2 and is the product of one further tailoring step. This raises the question as to the evolutionary driving force that permitted to the development of the pathway leading to a seemingly less active compound.

The modification of a central amino acid in the peptide chain from  $\alpha$  to  $\beta$  configuration implies a larger structural rearrangement than modifications of the oxidation or methylation decoration pattern and could lead to significantly different binding properties between both Cystobactamides. It is likely that such variability in their binding could prevent the easy development of resistance against one or the other compound which would be an advantage for the producer. Even if such a case has never been observed for different compounds of the same family, some *Stigmatella aurantiaca* strains are known to sometimes produce a cocktail of electron chain inhibitors presenting different binding sites (54) such as Stigmatellins,(151) Myxothiazol,(152) Aurachins,(153) Aurafurons,(154) and Myxalamides(155) which could be linked to the prevention of resistance development.

Furthermore it is noteworthy that while Cystobactamide are gyrase inhibitors, the presence of a secondary target cannot be ruled out and both Cystobactamide could, in this case, even have different main targets. Lastly it is important to consider the producer organism in its natural habitat, the predator-prey interactions it undergoes mainly involves other environmental organisms rather than the pathogenic bacteria we screened, and it is probable that Cystobactamides present a different activity profile against these organisms.

## 3.2 Cystobactamide biosynthetic gene cluster

Cystobactamides are non-ribosomal peptides, the gene cluster responsible for their biosynthesis was identified by retro-biosynthetic analysis, it spans approximately 55 kb and

features 22 genes spread on 3 operons. The last operon is dedicated to genes coding for pumps and permeases while the two first are carrying the biosynthesis and resistance genes (table 1.1).

In the scope of this work only the biosynthesis genes were investigated in detail, which cover set of fifteen genes. Three NRPS proteins are encoded on the cluster (CysH, CysK and CysG) coding for seven modules, in addition to six tailoring enzymes:

- CysR- para-aminobenzoate N dioxygenase
- CysJ- isoasparagine oxygenase
- CysQ- hydroxy-isoasparagine methyl transferase
- CysC- para-amino benzoateCoA oxidase
- CysF- hydroxy-para-amino benzoate methyl transferase
- CysS- methoxy-para-amino benzoate methyl transferase

Three other enzymes are also biosynthesis related and include a shuttling protein (CysB), an activating CoA ligase (CysL) and an adenylation domain stabilizing protein (CysA). Finally, the last three are responsible for the biosynthesis of the para-amino-benzoate building block (CysN- DAHP synthase, CysD- aminodeoxychorismate synthase and CysI- aminodeoxychorismate lyase), these enzymes are already well described and where not investigated further.

### 3.3 Biosynthesis

Preliminary annotation of the genes led to the hypothesis that the Cystobactamide backbone peptide is synthesized by 7 NRPS modules encoded on three genes. CysK harbors modules 1 to 4, the adenylation domain of the third module is inactive while the Shtachelhaus code of the three other modules is predicted to be specific for aromatic amino acids which fits with para-amino benzoic acid. CysG codes for modules 5 and 6 and is thought to incorporate 3-hydroxy para-amino benzoic acid. CysH codes for module 3', the adenylation domain is specific for asparagine and it is lacking a condensation domain. In the absence of an active A domain in module 3 the module 3' bound isoasparagine would be transferred onto CysK3 by an unusual shuttling protein: CysB. A large set of tailoring enzymes is responsible for oxidations and methylations

Table 1.1: Annotation of the Cystabactamide biosynthetic gene cluster in Cb v34

Name	size	Function	Annotations
CysS	1929 bp / 73 kDa	3-Methoxy-pABA C-methyltransferase	Radical SAM, B12 binding domain C-methylation of PCP bound 3-Methoxy-pABA
CysR	1002 bp / 37,8 kDa	pABA N-dioxygenase	35% identity / 57% similarity to AurF Adrenodoxine/Adrenodoxine dependent reductase
CysQ	795 bp / 29,2 kDa	2-hydroxy-isoasparagine O-methyltransferase	SAM dependent OMT
CysP	702 bp / 25 kDa	Resistance protein / Prodrug Mechanism	Bleomycin resistance protein family
CysO	612 bp / 22,6 kDa	Resistance protein	Pentapeptide repeat protein Folds to right-handed quadrilateral $\beta$ -helix that binds to DNA Gyrase
CysA	213 bp / 8 kDa	A domains activator	MbtH superfamily
CysB	954 bp / 34,9 kDa	Isoasparaginyl-transferase	Aminoacyl transferase, $\alpha/\beta$ hydrolase superfamily 20% identity / 35% similarity to SyrC 51% identity / 67% similarity to AlbXI
CysC	1380 bp/ 52,6 kDa	pABA-CoA oxygenase	BenzoateCoA oxidase component B 60% identity / 72% similarity
CysD	2199 bp / 80,6 kDa	Amino-deoxychorismate Synthase	pABA biosynthesis
CysE	732 bp / 26,9 kDa	Thioesterase	Type II TE
CysF	1038 bp/ 37,7 kDa	3-hydroxy-pABA-CoA O-methyltransferase	SAM dependant, related to type12 methyltransferases
CysG	5979 bp/ 219,6 kDa	2 modules NRPS	6 domains: A-T / C-A-T-TE $A_5$ and $A_6$ are specific for 3OH-pABA
CysH	2928 bp / 108,2 kDa	1 module NRPS	3 domains: A-X-T A is specific for asparagine X is an unprecedented bifunctional asparaginyl dehydratase / $\beta$ -hydroxy-asparaginyl mutase domain
CysI	819 bp / 30 kDa	4-amino-4-deoxy chorismate lyase	pABA biosynthesis
CysJ	984 bp / 36,9 kDa	Asparagine hydroxylase	$\beta$ -hydroxylation of PCP bound asparagine
CysK	13638 bp / 506,9 kDa	4 modules NRPS	13 domains: C-A-T / C-A-T / C-A-T / C-A-T / C $C_1$ and $A_3$ are inactive $A_1$ , $A_2$ and $A_4$ are specific for pABA
CysL	3072 bp / 110,9 kDa	pABA-CoA ligase	Analogue to benzoate-CoA and anthranilate- CoA ligases
CysN	1074 bp / 39,3 kDa	DAHP synthase	pABA biosynthesis

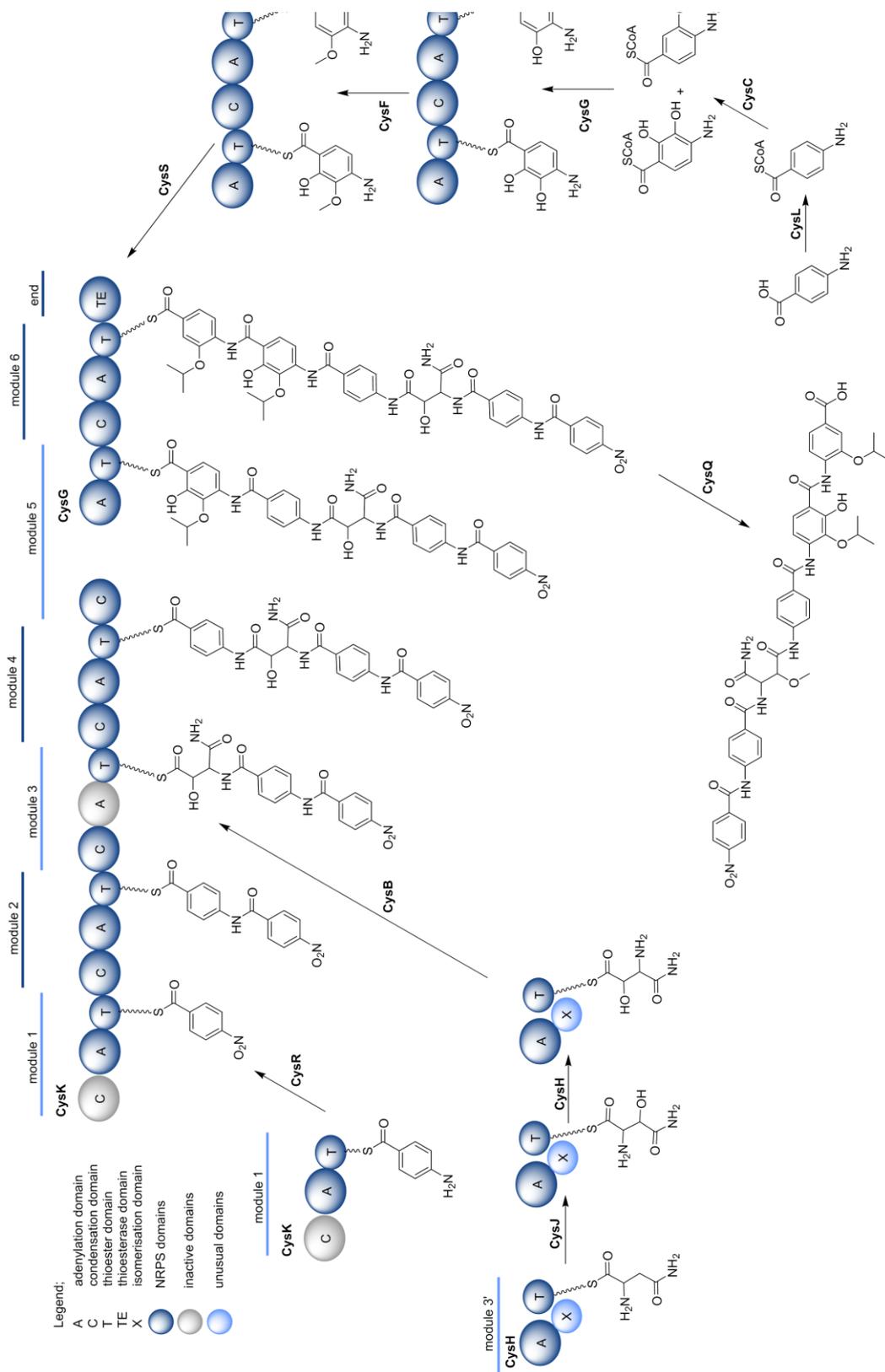


Figure 1.19: General scheme of the Cystobactamides biosynthesis

While, for most of them, it was impossible to determine from *in silico* data alone whether tailoring enzymes perform their activity on building block precursors, *in trans* on PCP bound intermediates or after release of the assembled peptide backbone, they could be split into two groups responsible for the biosynthesis of the unusual moieties: methoxy-isoasparagine and isopropoxy-para-amino benzoate. The biosynthesis of the unusual moieties (methoxy-isoasparagine and isopropoxy-para-amino benzoate) is thought to be performed by two more or less independent systems of enzymes. The first one features CysH, CysJ, CysB and CysK3 and is responsible for the biosynthesis of the hydroxy-isoasparagine. And the second system features CysL, CysC, CysF, CysS and ends on CysG, it biosynthesizes the two 3-isopropoxy-para-amino benzoates.

CysQ and CysR are the remaining independent tailoring enzymes. Since CysF is related to aromatic methyltransferases CysQ was thought to be responsible for the methylation of the hydroxy-isoasparagine moiety. CysR is a pABA N-oxygenase is closely related to AurF, a thoroughly characterized di-iron pABA dioxygenase, CysR was thus hypothesized to be responsible for the formation of the N-terminal nitro group. It is noteworthy that while AurF is active on free para-amino benzoate rather than on the PCP bound intermediate the first adenylation domain is almost identical to the second and fourth (97% identity), it is thus highly probable that they all activate the same substrate pNBA, and CysR would thus have to perform its activity *in trans* rather than on a precursor.(156,157)

### 3.4 PABA Biosynthesis

Para-amino benzoate is a central building block for the Cystobactamide biosynthesis, it is used by five out of the six modules from the assembly line. As for most secondary metabolism building blocks, pABA originates in primary metabolism. pABA is synthesized through the shikimate pathway and is a key intermediate in the folate biosynthesis which is central for the one carbon metabolism essential for nucleic acid biosynthesis. The Cystobactamide gene cluster features three key genes for the biosynthesis of pABA.

CysN is a DAHP synthase which is the first enzyme of the Shikimate pathway responsible for the condensation of phosphoenol pyruvate with erythrose phosphate to yield 3-Deoxy-D-arabinoheptulosonate-7-phosphate (DAHP).(158) As the first enzyme of the pathway it occupies a key position in controlling the amount of substrate fed into the pathway and thus the amount of shikimate which is produced. An additional copy of this gene in the Cystobactamide cluster will mechanically stimulate the production of shikimate and ultimately the production of pABA.

The two other genes are CysD and CysI which are encoding respectively an aminodeoxychorismate synthase and an aminodeoxychorismate lyase. These two enzymes are responsible for deriving the chorismate from the phenylalanine biosynthesis and lead to the production of pABA (159).

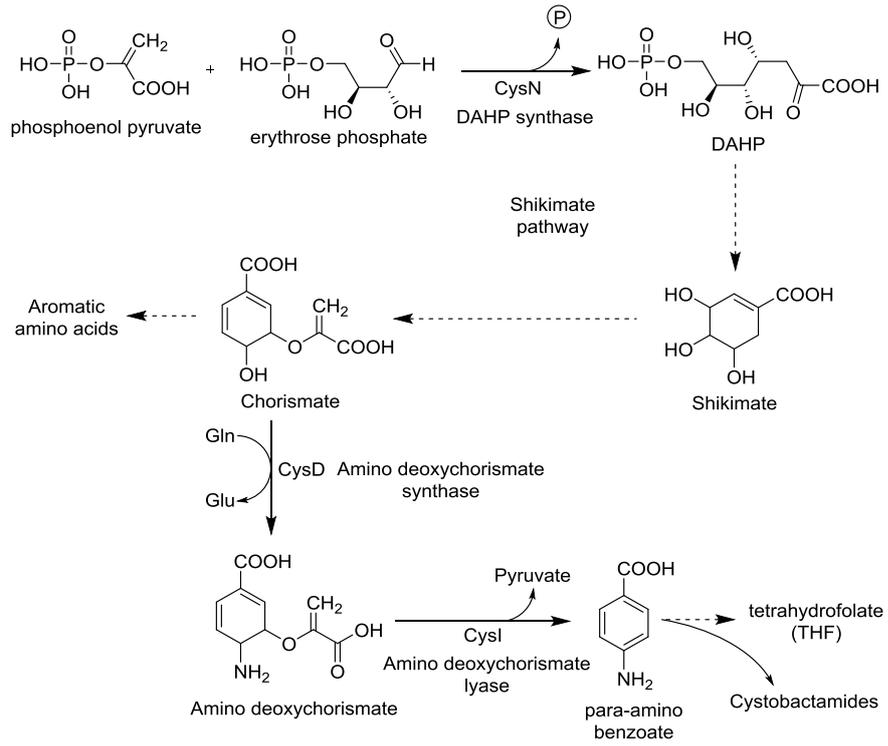


Figure 1.20: CysN, D and I in the pABA biosynthesis

## 4 References

1. Williams, D. H., Stone, M. J., Hauck, P. R., and Rahman, S. K. (1989) Why Are Secondary Metabolites (Natural Products) Biosynthesized?, *J. Nat. Prod.* 52, 1189–1208.
2. Stone, M. J., and Williams, D. H. (1992) On the evolution of functional secondary metabolites (natural products), *Molecular microbiology* 6, 29–34.
3. Kossel Albrecht (1891) Über die chemische Zusammensetzung der Zelle. In *Archiv für Physiologie Jahrgang 1891* (Emil du Bois-Raymond, Ed.), pp 181–186, Berlin [etc.] de Gruyter [etc.].
4. Luckner, M. (1984) Secondary Metabolism in Microorganisms, Plants and Animals, Springer Berlin Heidelberg, Berlin, Heidelberg.
5. Fischbach, M. A., Walsh, C. T., and Clardy, J. (2008) The evolution of gene collectives: How natural selection drives chemical innovation, *Proceedings of the National Academy of Sciences of the United States of America* 105, 4601–4608.
6. Traxler, M. F., and Kolter, R. (2015) Natural products in soil microbe interactions and evolution, *Natural product reports* 32, 956–970.
7. Newman, D. Screening and identification of novel biologically active natural compounds.
8. Newman, D. J. (2008) Natural products as leads to potential drugs: an old process or the new hope for drug discovery?, *J. Med. Chem.* 51, 2589–2599.
9. Newman, D. J., and Cragg, G. M. (2016) Natural Products as Sources of New Drugs from 1981 to 2014, *Journal of natural products* 79, 629–661.
10. Lietava, J. (1992) Medicinal plants in a Middle Paleolithic grave Shanidar IV?, *Journal of Ethnopharmacology* 35, 263–266.
11. Scurlock, J. A. (2014) Sourcebook for ancient Mesopotamian medicine, Society of Biblical Literature, Atlanta Georgia.
12. Bryan, C. P., and Joachim, H. (1930) The Papyrus Ebers, Ares publishers inc., London.
13. Harper, D. J. (1998) Early Chinese medical literature. The Mawangdui medical manuscripts / translation and study by Donald J. Harper, Kegan Paul International, London.
14. Tipton, C. M. (2008) Susruta of India, an unrecognized contributor to the history of exercise physiology, *Journal of applied physiology (Bethesda, Md. : 1985)* 104, 1553–1556.
15. Hippocrates, and Adams, F. (1849) The genuine works of Hippocrates ; translated from the Greek with a preliminary discourse and annotations, Sydenham Society, London.
16. Theophrastus, and Hort, A. (1916-1926) Enquiry into plants, and minor works on odours and weather signs, W. Heinemann; Cambridge, Mass., Harvard University Press, London.
17. Celsus, A. C., and Spencer, W. G. (1935) De medicina, W. Heinemann ltd.; Cambridge, Mass. : Harvard university press, London.
18. Dioscorides Pedanius, Osbaldeston, T. A., and Wood, R. P. A. (2000) Dioscorides de materia medica, IBIDIS, Johannesburg.
19. Avicenna, and Gruner, O. C. (1970) A treatise on the Canon of medicine of Avicenna, Kelley, New York.

20. Sertürner, F. (1805) Untitled letter to the editor. In *Journal der Pharmacie für Ärzte, Apotheker und Chemisten* (F.C.W. Vogel, Ed.), pp 229–243.
21. Pelletier and Caventou (1820) Suite des recherches chimiques sur les quinquinas. In *Annales de Chimie et de Physique* (Crochard, Ed.), pp 337–365.
22. Buchner, A. (1828) Repertorium für die pharmacie. Ueber das Rigatellische Fiebermittel und über eine in der Weidenrinde entdeckte alcaloidische Substanz pp 405-420, J. L. Schrag, Nürnberg.
23. Gerhardt C. (1953) Recherches sur les acides organiques anhydres, *Ann. de Chim. et Phys.* 3, 285–342.
24. Fleming, A. (1929) On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae., *the British Journal of Experimental Pathology* 10, 226–236.
25. Schatz, A., Bugle, E., and Waksman, S. A. (1944) Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.\*, *Experimental Biology and Medicine* 55, 66–69.
26. Ehrlich, J., Bartz, Q. R., Smith, R. M., Joslyn, D. A., and Burkholder, P. R. (1947) Chloromycetin, a New Antibiotic From a Soil Actinomycete, *Science* 106, 417.
27. Demain, A. L., and Sanchez, S. (2009) Microbial drug discovery: 80 years of progress, *The Journal of antibiotics* 62, 5–16.
28. van Boeckel, T. P., Brower, C., Gilbert, M., Grenfell, B. T., Levin, S. A., Robinson, T. P., Teillant, A., and Laxminarayan, R. (2015) Global trends in antimicrobial use in food animals, *Proc. Natl. Acad. Sci. U.S.A.* 112, 5649–5654.
29. Kirby, W. M. (1944) Extraction of a highly potent penicillin inactivator from penicillin resistant staphylococci, *Science (New York, N.Y.)* 99, 452–453.
30. Ventola, C. L. (2015) The antibiotic resistance crisis: part 1: causes and threats, *P & T : a peer-reviewed journal for formulary management* 40, 277–283.
31. Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., Scheld, M., Spellberg, B., and Bartlett, J. (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America, *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 48, 1–12.
32. CDC (2013) Antibiotic Resistance Threats in the United States, 2013.
33. Bérdy, J. (2012) Thoughts and facts about antibiotics: where we are now and where we are heading, *The Journal of antibiotics* 65, 385–395.
34. Bérdy, J. (2005) Bioactive microbial metabolites, *The Journal of antibiotics* 58, 1–26.
35. Rappé, M. S., and Giovannoni, S. J. (2003) The uncultured microbial majority, *Annual review of microbiology* 57, 369–394.
36. Watson, J. D., and Crick, F. H. C. (1953) Molecular Structure of Nucleic Acids. A Structure for Deoxyribose Nucleic Acid, *Nature* 171, 737–738.

- 
37. Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., and Ussery, D. W. (2015) Insights from 20 years of bacterial genome sequencing, *Functional & integrative genomics* 15, 141–161.
38. Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabinowitsch, E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J., and Hopwood, D. A. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), *Nature* 417, 141–147.
39. Challis, G. L. (2008) Genome mining for novel natural product discovery, *Journal of medicinal chemistry* 51, 2618–2628.
40. Wenzel, S. C., and Müller, R. (2009) The impact of genomics on the exploitation of the myxobacterial secondary metabolome, *Nat. Prod. Rep.* 26, 1385–1407.
41. Weissman, K. J., and Müller, R. (2009) A brief tour of myxobacterial secondary metabolism, *Bioorganic & Medicinal Chemistry* 17, 2121–2136.
42. Reichenbach, H. (2005) Order VIII. Myxococcales. Tchan, Pochon and Prévot 1948, 398AL. In *Bergey's manual of systematic bacteriology* (Brenner, D.J., Krieg, N.R., and Staley, J.T., Eds.), 2nd ed., pp 1059–1144, Springer.
43. Reichenbach, H. (1999) The ecology of the myxobacteria, *Environ. Microbiol.* 1, 15–21.
44. Dawid, W., Gallikowski, C. A., and Hirsch, P. (1988) Psychrophilic Myxobacteria from Antarctic Soils, *Polarforschung* 58, 271–278.
45. Iizuka, T., Tokura, M., Jojima, Y., Hiraishi, A., Yamanaka, S., and Fudou, R. (2006) Enrichment and Phylogenetic Analysis of Moderately Thermophilic Myxobacteria from Hot Springs in Japan, *Microbes Environ.* 21, 189–199.
46. Hook, L. A., Larkin, J. M., and Brockman, E. R. (1980) Isolation, characterization, and emendation of description of *Angiococcus disciformis* (Thaxter 1904) Jahn 1924 and proposal of a neotype strain, *Int J Syst Bacteriol* 30, 135–142.
47. Shimkets, L. J., Dworkin, M., and Reichenbach, H. (2006) The Myxobacteria. In *The Prokaryotes* (Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E., Eds.), pp 31–115, Springer New York, New York, NY.
48. Zusman, D. R., Scott, A. E., Yang, Z., and Kirby, J. R. (2007) Chemosensory pathways, motility and development in *Myxococcus xanthus*, *Nature reviews. Microbiology* 5, 862–872.
49. Shimkets, L., and Woese, C. R. (1992) A phylogenetic analysis of the myxobacteria: basis for their classification, *Proc. Natl. Acad. Sci. USA* 89, 9459–9463.
50. Han, K., Li, Z.-f., Peng, R., Zhu, L.-p., Zhou, T., Wang, L.-g., Li, S.-g., Zhang, X.-b., Hu, W., Wu, Z.-h., Qin, N., and Li, Y.-z. (2013) Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu, *Scientific reports* 3, 2101.
51. Gerth, K., Pradella, S., Perlova, O., Beyer, S., and Müller, R. (2003) Myxobacteria: proficient producers of novel natural products with various biological activities—past and

future biotechnological aspects with the focus on the genus *Sorangium*, *J. Biotechnol.* 106, 233–253.

52. Herrmann, J., Fayad, A. A., and Müller, R. (2017) Natural products from myxobacteria: novel metabolites and bioactivities, *Nat. Prod. Rep.* 34, 135–160.

53. Weissman, K. J., and Müller, R. (2010) Myxobacterial secondary metabolites: bioactivities and modes-of-action, *Nat. Prod. Rep.* 27, 1276–1295.

54. Reichenbach, H. (2001) Myxobacteria, producers of novel bioactive substances, *J. Ind. Microbiol. Biotechnol.* 27, 149–156.

55. Bode, H. B., and Müller, R. (2007) Secondary metabolism in myxobacteria. In *Myxobacteria: Multicellularity and differentiation* (Whitworth, D., Ed.), pp 259–282, ASM Press, Chicago.

56. Bode, H. B., Zeggel, B., Silakowski, B., Wenzel, S. C., Reichenbach, H., and Müller, R. (2003) Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(S)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*, *Mol. Microbiol.* 47, 471–481.

57. Reichenbach, H., and Höfle, G. (1993) Biologically active secondary metabolites from microorganism, *biotechnology advances* 11, 219–277.

58. Höfle, G., Bedorf, N., Steinmetz, H., Schomburg, D., Gerth, K., and Reichenbach, H. (1996) Epothilone A and B: Novel 16-membered macrolides with cytotoxic activity: Isolation, crystal structure, and conformation in solution, *Angew. Chem. Int. Ed. Engl.* 35, 1567–1569.

59. Gerth, K., Steinmetz, H., Höfle, G., and Jansen, R. (2008) Chlorotonil A, ein Macrolid mit einzigartiger gem-Dichlor-1,3-dionfunktion aus *Sorangium cellulosum*, *So ce1525*, *Angew. Chem.* 120, 610–613.

60. Held, J., Gebu, T., Kalesse, M., Jansen, R., Gerth, K., Müller, R., and Mordmüller, B. (2014) Antimalarial activity of the myxobacterial macrolide chlorotonil A, *Antimicrob. Agents Chemother.* 58, 6378–6384.

61. Irschik, H., Jansen, R., Gerth, K., Höfle, G., and Reichenbach, H. (1987) The sorangicins, novel and powerful inhibitors of eubacterial RNA polymerase isolated from myxobacteria, *J. Antibiot.* 40, 7–13.

62. Ringel, S. M., Greenough, R. C., Roemer, S., Connor, D., Gutt, A. L., Blair, B., Kanter, G., and von Strandtmann (1977) Ambruticin (W7783), a new antifungal antibiotic, *The Journal of antibiotics* 30, 371–375.

63. Schummer, D., Forche, E., Wray, V., Domke, T., Reichenbach, H., and Höfle, G. (1996) Antibiotics from gliding bacteria .76. Vioprolides: New antifungal and cytotoxic peptolides from *Cystobacter violaceus*, *Liebigs Ann-Recl Liebigs Ann-Recl*, 971–978.

64. Sasse, F., Steinmetz, H., Schupp, T., Petersen, F., Memmert, K., Hofmann, H., Heusser, C., Brinkmann, V., Matt, P. von, Höfle, G., and Reichenbach, H. (2002) Argyrins, immunosuppressive cyclic peptides from myxobacteria. I. Production, isolation, physico-chemical and biological properties, *J. Antibiot.* 55, 543–551.

65. Plaza, A., Garcia, R., Bifulco, G., Martinez, J. P., Hüttel, S., Sasse, F., Meyerhans, A., Stadler, M., and Müller, R. (2012) Aetheramides A and B, potent HIV-inhibitory depsipeptides from a myxobacterium of the new genus "Aetherobacter", *Org. Lett.* *14*, 2854–2857.
66. Gerth, K., Irschik, H., Reichenbach, H., and Trowitzsch, W. (1980) Myxothiazol, an antibiotic from *Myxococcus fulvus* (myxobacterales). I. Cultivation, isolation, physico-chemical and biological properties, *J. Antibiot.* *33*, 1474–1479.
67. Meiser, P., Bode, H. B., and Müller, R. (2006) The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation, *Proc. Natl. Acad. Sci. U.S.A.* *103*, 19128–19133.
68. Jansen, R., Irschik, H., Reichenbach, H., Wray, V., and Höfle, G. (1994) Disorazoles, highly cytotoxic metabolites from the Sorangicin-producing bacterium *Sorangium cellulosum*, strain So ce12, *Liebigs Ann. Chem.* *1994*, 759–773.
69. Kopp, M., Irschik, H., Gemperlein, K., Buntin, K., Meiser, P., Weissman, K. J., Bode, H. B., and Müller, R. (2011) Insights into the complex biosynthesis of the leupyrrins in *Sorangium cellulosum* So ce690, *Mol. Biosyst.* *7*, 1549–1563.
70. Dewick, P. M. (2009) Medicinal natural products. A biosynthetic approach / Paul M. Dewick. 3rd ed., Wiley-Blackwell, Oxford.
71. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014) Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes, *Proceedings of the National Academy of Sciences of the United States of America* *111*, 9259–9264.
72. Fischbach, M. A., and Walsh, C. T. (2006) Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms, *Chemical reviews* *106*, 3468–3496.
73. Cane, D. E., Walsh, C. T., and Khosla, C. (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations, *Science (New York, N.Y.)* *282*, 63–68.
74. Walsh, C. T., Chen, H., Keating, T. A., Hubbard, B. K., Losey, H. C., Luo, L., Marshall, C. G., Miller, D. A., and Patel, H. M. (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines, *Current opinion in chemical biology* *5*, 525–534.
75. Scherlach, K., and Hertweck, C. (2009) Triggering cryptic natural product biosynthesis in microorganisms, *Organic & biomolecular chemistry* *7*, 1753–1760.
76. Marahiel, M. A. (1997) Protein templates for the biosynthesis of peptide antibiotics, *Chemistry & biology* *4*, 561–567.
77. Winn, M., Fyans, J. K., Zhuo, Y., and Micklefield, J. (2016) Recent advances in engineering nonribosomal peptide assembly lines, *Natural product reports* *33*, 317–347.
78. Caboche, S., Pupin, M., Leclère, V., Fontaine, A., Jacques, P., and Kucherov, G. (2008) NORINE: a database of nonribosomal peptides, *Nucleic acids research* *36*, 31.
79. Finking, R., and Marahiel, M. A. (2004) Biosynthesis of nonribosomal peptides<sup>1</sup>, *Annual review of microbiology* *58*, 453–488.

80. Hur, G. H., Vickery, C. R., and Burkart, M. D. (2012) Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology, *Natural product reports* 29, 1074–1098.
81. Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases, *Chembiochem : a European journal of chemical biology* 3, 490–504.
82. Wenzel, S. C., Kunze, B., Höfle, G., Silakowski, B., Scharfe, M., Blöcker, H., and Müller, R. (2005) Structure and biosynthesis of myxochromides S1-3 in *Stigmatella aurantiaca*: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis, *Chembiochem : a European journal of chemical biology* 6, 375–385.
83. Sy-Cordero, A. A., Pearce, C. J., and Oberlies, N. H. (2012) Revisiting the enniatins: a review of their isolation, biosynthesis, structure determination and biological activities, *The Journal of antibiotics* 65, 541–549.
84. Gehring, A. M., Mori, I., and Walsh, C. T. (1998) Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF, *Biochemistry* 37, 2648–2659.
85. Yu, D., Xu, F., Zhang, S., and Zhan, J. (2017) Decoding and reprogramming fungal iterative nonribosomal peptide synthetases, *Nature communications* 8, 15349.
86. Keating, T. A., Marshall, C. G., and Walsh, C. T. (2000) Vibriobactin Biosynthesis in *Vibrio cholerae*. VibH Is an Amide Synthase Homologous to Nonribosomal Peptide Synthetase Condensation Domains †, *Biochemistry* 39, 15513–15521.
87. Bloudoff, K., and Schmeing, T. M. (2017) Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity, *Biochimica et biophysica acta* 1865, 1587–1604.
88. Li, Y., Weissman, K. J., and Müller, R. (2008) Myxochelin biosynthesis: direct evidence for two- and four-electron reduction of a carrier protein-bound thioester, *Journal of the American Chemical Society* 130, 7554–7555.
89. Crécy-Lagard, V. de, Marlière, P., and Saurin, W. (1995) Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation, *Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie* 318, 927–936.
90. Mitchell, C. A., Shi, C., Aldrich, C. C., and Gulick, A. M. (2012) Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains, *Biochemistry* 51, 3252–3263.
91. Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A., and Holak, T. A. (2000) Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases, *Structure (London, England : 1993)* 8, 407–418.
92. Koglin, A., Mofid, M. R., Löhr, F., Schäfer, B., Rogov, V. V., Blum, M.-M., Mittag, T., Marahiel, M. A., Bernhard, F., and Dötsch, V. (2006) Conformational switches modulate protein interactions in peptide antibiotic synthetases, *Science (New York, N.Y.)* 312, 273–276.

- 
93. Reimer, J. M., Aloise, M. N., Harrison, P. M., and Schmeing, T. M. (2016) Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase, *Nature* 529, 239–242.
94. Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C., and Gulick, A. M. (2016) Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture, *The Journal of biological chemistry* 291, 22559–22571.
95. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A., and Essen, L.-O. (2007) Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase, *Structure (London, England : 1993)* 15, 781–792.
96. Hillson, N. J., and Walsh, C. T. (2003) Dimeric structure of the six-domain VibF subunit of vibriobactin synthetase: mutant domain activity regain and ultracentrifugation studies, *Biochemistry* 42, 766–775.
97. Tarry, M. J., Haque, A. S., Bui, K. H., and Schmeing, T. M. (2017) X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture, *Structure (London, England : 1993)* 25, 783.
98. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases, *Chemistry & biology* 6, 493–505.
99. May, J. J., Kessler, N., Marahiel, M. A., and Stubbs, M. T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases, *Proceedings of the National Academy of Sciences of the United States of America* 99, 12120–12125.
100. Yonus, H., Neumann, P., Zimmermann, S., May, J. J., Marahiel, M. A., and Stubbs, M. T. (2008) Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains, *The Journal of biological chemistry* 283, 32484–32491.
101. Linne, U., and Marahiel, M. A. (2004) Reactions catalyzed by mature and recombinant nonribosomal peptide synthetases, *Methods in enzymology* 388, 293–315.
102. Villiers, B. R. M., and Hollfelder, F. (2009) Mapping the limits of substrate specificity of the adenylation domain of TycA, *Chembiochem : a European journal of chemical biology* 10, 671–682.
103. Strieker, M., Tanović, A., and Marahiel, M. A. (2010) Nonribosomal peptide synthetases: structures and dynamics, *Current opinion in structural biology* 20, 234–240.
104. Tanovic, A., Samel, S. A., Essen, L.-O., and Marahiel, M. A. (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase, *Science (New York, N.Y.)* 321, 659–663.
105. Nguyen, K. T., Ritz, D., Gu, J.-Q., Alexander, D., Chu, M., Miao, V., Brian, P., and Baltz, R. H. (2006) Combinatorial biosynthesis of novel antibiotics related to daptomycin, *Proceedings of the National Academy of Sciences of the United States of America* 103, 17462–17467.

106. Kries, H., Niquille, D. L., and Hilvert, D. (2015) A subdomain swap strategy for reengineering nonribosomal peptides, *Chemistry & biology* 22, 640–648.
107. Belshaw, P. J. (1999) Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis, *Science* 284, 486–489.
108. Lin, S., van Lanen, S. G., and Shen, B. (2009) A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis, *Proceedings of the National Academy of Sciences of the United States of America* 106, 4183–4188.
109. Marxen, S., Stark, T. D., Rüttschle, A., Lücking, G., Frenzel, E., Scherer, S., Ehling-Schulz, M., and Hofmann, T. (2015) Depsipeptide Intermediates Interrogate Proposed Biosynthesis of Cereulide, the Emetic Toxin of *Bacillus cereus*, *Scientific reports* 5, 10637.
110. Keating, T. A., Marshall, C. G., Walsh, C. T., and Keating, A. E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains, *Nature structural biology* 9, 522–526.
111. Stachelhaus, T., Mootz, H. D., Bergendahl, V., and Marahiel, M. A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain, *The Journal of biological chemistry* 273, 22773–22781.
112. Bloudoff, K., Alonzo, D. A., and Schmeing, T. M. (2016) Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases, *Cell chemical biology* 23, 331–339.
113. Marshall, C. G., Burkart, M. D., Keating, T. A., and Walsh, C. T. (2001) Heterocycle Formation in Vibriobactin Biosynthesis. Alternative Substrate Utilization and Identification of a Condensed Intermediate †, *Biochemistry* 40, 10655–10663.
114. Bloudoff, K., Fage, C. D., Marahiel, M. A., and Schmeing, T. M. (2017) Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis, *Proceedings of the National Academy of Sciences of the United States of America* 114, 95–100.
115. Bruner, S. D., Weber, T., Kohli, R. M., Schwarzer, D., Marahiel, M. A., Walsh, C. T., and Stubbs, M. T. (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE, *Structure (London, England : 1993)* 10, 301–310.
116. Samel, S. A., Wagner, B., Marahiel, M. A., and Essen, L.-O. (2006) The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide, *Journal of molecular biology* 359, 876–889.
117. Stachelhaus, T., and Walsh, C. T. (2000) Mutational Analysis of the Epimerization Domain in the Initiation Module PheATE of Gramicidin S Synthetase †, *Biochemistry* 39, 5775–5787.
118. Schneider, T. L., Shen, B., and Walsh, C. T. (2003) Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation, *Biochemistry* 42, 9722–9730.
119. Weckwerth, W., Miyamoto, K., Iinuma, K., Krause, M., Glinski, M., Storm, T., Bonse, G., Kleinkauf, H., and Zocher, R. (2000) Biosynthesis of PF1022A and related cyclooctadepsipeptides, *The Journal of biological chemistry* 275, 17909–17915.

120. Kawai, Y., Ishii, Y., Arakawa, K., Uemura, K., Saitoh, B., Nishimura, J., Kitazawa, H., Yamazaki, Y., Tateno, Y., Itoh, T., and Saito, T. (2004) Structural and Functional Differences in Two Cyclic Bacteriocins with the Same Sequences Produced by Lactobacilli, *Applied and environmental microbiology* 70, 2906–2911.
121. Roberto Bessalle , Aviva Kapitkovsky , Alfred Goreaz, Itamar Shalit and Mati Fridkin (1990) All-D-magainin. Chirality, antimicrobial activity and proteolytic resistance, *FEBS Letters* 274, 151–155.
122. Chen, W.-H., Li, K., Guntaka, N. S., and Bruner, S. D. (2016) Interdomain and Intermodule Organization in Epimerization Domain Containing Nonribosomal Peptide Synthetases, *ACS chemical biology* 11, 2293–2303.
123. Patel, H. M., and Walsh, C. T. (2001) In vitro reconstitution of the *Pseudomonas aeruginosa* nonribosomal peptide synthesis of pyochelin: characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities, *Biochemistry* 40, 9023–9031.
124. Du, L., and Lou, L. (2010) PKS and NRPS release mechanisms, *Natural product reports* 27, 255–278.
125. Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., and Walsh, C. T. (1996) A new enzyme superfamily - the phosphopantetheinyl transferases, *Chemistry & biology* 3, 923–936.
126. Felnagle, E. A., Barkei, J. J., Park, H., Podevels, A. M., McMahon, M. D., Drott, D. W., and Thomas, M. G. (2010) MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases, *Biochemistry* 49, 8815–8817.
127. Schwarzer, D., Mootz, H. D., Linne, U., and Marahiel, M. A. (2002) Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases, *Proceedings of the National Academy of Sciences of the United States of America* 99, 14083–14088.
128. Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P., and Burkart, M. D. (2014) The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life, *Natural product reports* 31, 61–108.
129. Mofid, M. R., Finking, R., Essen, L. O., and Marahiel, M. A. (2004) Structure-based mutational analysis of the 4'-phosphopantetheinyl transferases Sfp from *Bacillus subtilis*: carrier protein recognition and reaction mechanism, *Biochemistry* 43, 4128–4136.
130. Herbst, D. A., Boll, B., Zocher, G., Stehle, T., and Heide, L. (2013) Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes, *The Journal of biological chemistry* 288, 1991–2003.
131. Wolpert, M., Gust, B., Kammerer, B., and Heide, L. (2007) Effects of deletions of mbtH-like genes on clorobiocin biosynthesis in *Streptomyces coelicolor*, *Microbiology (Reading, England)* 153, 1413–1423.
132. Schomer, R. A., and Thomas, M. G. (2017) Characterization of the Functional Variance in MbtH-like Protein Interactions with a Nonribosomal Peptide Synthetase, *Biochemistry* 56, 5380–5390.

133. McMahon, M. D., Rush, J. S., and Thomas, M. G. (2012) Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*, *Journal of bacteriology* *194*, 2809–2818.
134. Esquilín-Lebrón, K. J., Boynton, T. O., Shimkets, L. J., and Thomas, M. G. (2018) An orphan MbtH-like protein interacts with multiple nonribosomal peptide synthetases in *Myxococcus xanthus* DK1622, *Journal of bacteriology* Epub 2018. DOI: 10.1128/JB.00346-18.
135. Boll, B., and Heide, L. (2013) A domain of RubC1 of rubradirin biosynthesis can functionally replace MbtH-like proteins in tyrosine adenylation, *Chembiochem : a European journal of chemical biology* *14*, 43–44.
136. Saha, S., and Rokita, S. E. (2016) An Activator of an Adenylation Domain Revealed by Activity but Not Sequence Homology, *Chembiochem : a European journal of chemical biology* *17*, 1818–1823.
137. Kalb, D., Lackner, G., Rappe, M., and Hoffmeister, D. (2015) Activity of  $\alpha$ -Aminoacidate Reductase Depends on the N-Terminally Extending Domain, *Chembiochem : a European journal of chemical biology* *16*, 1426–1430.
138. Chohnan, S., Furukawa, H., Fujio, T., Nishihara, H., and Takamura, Y. (1997) Changes in the size and composition of intracellular pools of nonesterified coenzyme A and coenzyme A thioesters in aerobic and facultatively anaerobic bacteria, *Applied and environmental microbiology* *63*, 553–560.
139. Koglin, A., Löhr, F., Bernhard, F., Rogov, V. V., Frueh, D. P., Strieter, E. R., Mofid, M. R., Güntert, P., Wagner, G., Walsh, C. T., Marahiel, M. A., and Dötsch, V. (2008) Structural basis for the selectivity of the external thioesterase of the surfactin synthetase, *Nature* *454*, 907–911.
140. Hertweck, C. (2009) The biosynthetic logic of polyketide diversity, *Angewandte Chemie (International ed. in English)* *48*, 4688–4716.
141. Rawlings, B. J. (2001) Type I polyketide biosynthesis in bacteria (Part A--erythromycin biosynthesis), *Natural product reports* *18*, 190–227.
142. Keatinge-Clay, A. T. (2017) The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines, *Chemical reviews* *117*, 5334–5366.
143. Piel, J. (2002) A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles, *Proceedings of the National Academy of Sciences of the United States of America* *99*, 14002–14007.
144. Campbell, C. D., and Vederas, J. C. (2010) Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes, *Biopolymers* *93*, 755–763.
145. Hutchinson, C. R. (1997) Biosynthetic Studies of Daunorubicin and Tetracenomycin C, *Chem. Rev.* *97*, 2525–2536.
146. Das, A., and Khosla, C. (2009) Biosynthesis of aromatic polyketides in bacteria, *Accounts of chemical research* *42*, 631–639.
147. Austin, M. B., and Noel, J. P. (2003) The chalcone synthase superfamily of type III polyketide synthases, *Natural product reports* *20*, 79–110.

148. Hug, J. J., Panter, F., Krug, D., and Müller, R. (2018) Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria, *Journal of industrial microbiology & biotechnology* Epub 2018. DOI: 10.1007/s10295-018-2105-6.
149. Baumann, S., Herrmann, J., Raju, R., Steinmetz, H., Mohr, K. I., Hüttel, S., Harmrolfs, K., Stadler, M., and Müller, R. (2014) Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity, *Angewandte Chemie (International ed. in English)* 53, 14605–14609.
150. Hüttel, S., Testolin, G., Herrmann, J., Planke, T., Gille, F., Moreno, M., Stadler, M., Brönstrup, M., Kirschning, A., and Müller, R. (2017) Discovery and Total Synthesis of Natural Cystobactamid Derivatives with Superior Activity against Gram-Negative Pathogens, *Angewandte Chemie (International ed. in English)* 56, 12760–12764.
151. Thierbach, G., Kunze, B., Reichenbach, H., and Höfle, G. (1984) The mode of action of stigmatellin, a new inhibitor of the cytochrome b-c1 segment of the respiratory chain, *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 765, 227–235.
152. Thierbach, G., and Reichenbach, H. (1981) Myxothiazol, a new inhibitor of the cytochrome b-c1 segment of the respiratory chain, *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 638, 282–289.
153. Kunze, B., Höfle, G., and Reichenbach, H. (1987) The aurachins, new quinoline antibiotics from myxobacteria: production, physico-chemical and biological properties, *J. Antibiot.* 40, 258–265.
154. Kunze, B., Reichenbach, H., Müller, R., and Höfle, G. (2005) Aurafuron A and B, new bioactive polyketides from *Stigmatella aurantiaca* and *Archangium gephyra* (Myxobacteria). Fermentation, isolation, physico-chemical properties, structure and biological activity, *J. Antibiot.* 58, 244–251.
155. Gerth, K., Jansen, R., Reifensahl, G., Höfle, G., Irschik, H., Kunze, B., Reichenbach, H., and Thierbach, G. (1983) The myxalamids, new antibiotics from *Myxococcus xanthus* (Myxobacteriales). I. Production, physico-chemical and biological properties, and mechanism of action, *J. Antibiot.* 36, 1150–1156.
156. Simurdiak, M., Lee, J., and Zhao, H. (2006) A new class of arylamine oxygenases: evidence that p-aminobenzoate N-oxygenase (AurF) is a di-iron enzyme and further mechanistic studies, *Chembiochem : a European journal of chemical biology* 7, 1169–1172.
157. Choi, Y. S., Zhang, H., Brunzelle, J. S., Nair, S. K., and Zhao, H. (2008) In vitro reconstitution and crystal structure of p-aminobenzoate N-oxygenase (AurF) involved in aureothin biosynthesis, *Proceedings of the National Academy of Sciences of the United States of America* 105, 6858–6863.
158. Herrmann, K. M. (1995) The shikimate pathway as an entry to aromatic secondary metabolism, *Plant physiology* 107, 7–12.
159. Parsons, J. F., Jensen, P. Y., Pachikara, A. S., Howard, A. J., Eisenstein, E., and Ladner, J. E. (2002) Structure of *Escherichia coli* Aminodeoxychorismate Synthase. Architectural Conservation and Diversity in Chorismate-Utilizing Enzymes, *Biochemistry* 41, 2198–2208.



# **Chapter 2: CysL - Investigation of a secondary metabolism related para-aminobenzoate CoA ligase and its feedback inhibition by Cystobactamides**

## **1 Introduction**

In primary metabolism aromatic acids coenzyme A (CoA) thioesters intermediates are most of the time related to anaerobic catabolism of aromatic substrates. In absence of oxygen, the activation of the ring is necessary for the reduction leading to a loss of aromaticity.(1,2) These mechanisms have been studied thoroughly in the bacteria *Azocardus evansii* (3) and *Thaurea aromatica*.(4) In secondary metabolism, aromatic CoA esters can sometimes be used as starter unit in PKS biosynthesis pathways such as Enterocin (5) or Soraphen.(6) Amino benzoates, due to their bi-functional nature, can be found as building blocks in NRPS assembly lines but they are usually not activated as CoAs prior to integration.(7) Anthranilic acid and 3-amino benzoic acid are much more frequently encountered than para-amino benzoic acid (pABA) which is only observed in a few compounds such as Candicidine (8), Albicidin (9) and Cystobactamides.(10) pABA is originating in the folate biosynthesis for which three key enzymes can be found in the Cystobactamide cluster. The presence of these three enzymes speaks for an upregulation of this metabolic route which is consistent with the central position of pABA in Cystobactamides since it is used by five out of the six modules of the assembly line.

CoA ligases belong to the class I adenylate forming enzymes along with adenylation domains of NRPS assembly lines. They feature the same large N-terminal core domain responsible for substrate recognition and ATP binding and C-terminal subdomain harboring the catalytic lysine.(11) Furthermore, their core regions can be superimposed but the conserved aspartate responsible for the hinge in adenylation domains is not present in CysL. However the large rearrangement which is necessary to bind the PCP in NRPS is not required in the case of CoA ligation. CoA ligases catalyze the formation of a thioester between an acid and CoA in two steps first of which is the ATP dependent activation of the acid as adenylate and finally the nucleophilic attack of the free thiol on the acid anhydride.(12)

*In silico* analysis of the *cysL* gene reveals that it harbors two domains of about 500 aa each, the C-terminal domain being related to Adenylate forming enzymes. The N-terminal domain, however, is of unknown function and has no sequence homology to any known protein. The lack of apparent catalytic sequence/fold led to the hypothesis that this domain could be involved in the regulation of the activity rather than in a tailoring reaction. CysL can be considered as the first enzyme of the biosynthesis since it activates the pABA building blocks, it would thus have a lot of leverage on the turnover of the assembly line. Albicidins and Cystobactamide share a similar structure and hence a similar biosynthesis outside of a few key aspects and AlbVII is recognizable as CysH homologue from their CoA ligase

domain which share 52/65% identity/similarity. However they share no homology on the N-terminal regulatory domain which is less than half the size in AlbVII with about 220 aa.

## 2 Material and methods

### 2.1 Cloning, overexpression and purification

Standard protocols were used for DNA cloning, transformation, amplification, and purification. Genomic DNA was extracted from *Cystobacter velatus* Cbv34 grown in CTT medium, the DNA fragment encoding CysL was amplified using the forward primer 5'TATCATATGGTGAACGTGCTCGCTAGGCATTC and the reverse primer 5'TATGGATCCTCATGGGGGCCCGCTGGTGA. The amplified DNA fragments were digested with NdeI and BamHI, and cloned into pET-28b with an N-terminal His6 tag. The resulting constructs were sequenced to verify that no mutation had been introduced during PCR.

Recombinant protein expression was carried out in *E.coli* BL21 (DE3) grown in LB medium supplemented with 50 µg/ml of kanamycin. The culture was inoculated 1/10 from a fully grown overnight culture and was initially grown at 37 °C until it reached an OD of 0.6 to 0.9, at which point the temperature was decreased to 16 °C. Induction was initiated after 30 minutes at 16°C with 0,1 mM isopropyl-β,D-thiogalactopyranoside (IPTG). The cells were harvested 16h after induction by centrifugation at 8000 g for 10 minutes at 4°C and resuspended in lysis buffer (25 mM TRIS pH7,5; 150 mM NaCl; 20 mM imidazole). The cells were lysed by passage at 1500 bar at 4°C through a M-110P Microfluidizer (Microfluidics) after which the crude extract was centrifuged at 45000 g for 15 minutes at 4°C, the supernatant was treated as soluble protein fraction.

Affinity chromatography was performed on an Äkta Avant system while size exclusion chromatography was performed on an Äkta Pure system. The soluble fraction of the lysed cells was applied to 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10 CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole. The pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200 pg SEC column equilibrated in running buffer (25 mM TRIS pH7,5; 150 mM NaCl; 2 mM DTT). The pooled fractions were controlled via SDS-PAGE, concentrated and frozen at -80°C after addition of 25% glycerol.

### 2.2 Coadenylation assays

In a total volume of 50 µl, 1mM benzoic acid derivatives were mixed with 1mM ATP and 1 mM CoA in 25 mM TRIS pH7,5; 150 mM NaCl, 10 mM MgCl<sub>2</sub>. The reaction was started by adding 1 µM CysL and incubated at room temperature for 30 minutes. The

enzyme was precipitated by addition of 50  $\mu$ l MeOH and centrifuged at 150000 g for 5 minutes in a tabletop centrifuge. The supernatant was analyzed by LC-MS.

CoA thioesters were analyzed on a Dionex Ultimate 3000 UPLC system (Thermo Scientific) coupled with an Amazon ion trap MS (Bruker). The samples were run on a Synergi fusion 4 $\mu$  Polar-RP 80A 250x3,0 mm C18 column (phenomenex). LC conditions: A-Water + 5 mM Ammonium formiate; B-Methanol + 5 mM Ammonium formiate; 0 min: 98% A / 2% B, 1 min: 98 % A / 2 % B, 16 min: 80 % A / 20 % B, 18 min: 5% A / 95 % B, 19 min: 5 % A / 95 % B, 20 min: 98 % A / 2 % B.

### 2.3 PNP-MESG assay

The biochemical assay based on the Phosphate dependent cleavage of 7-methyl-6-Thioguanosine (MESG,  $\lambda_{\text{max}}= 320$  nm) in 7-methyl-6-thioguanine ( $\lambda_{\text{max}}= 360$  nm) and ribose-phosphate by the purine nucleoside phosphorylase (PNP) was chosen to monitor the kinetics of the reaction.(13)

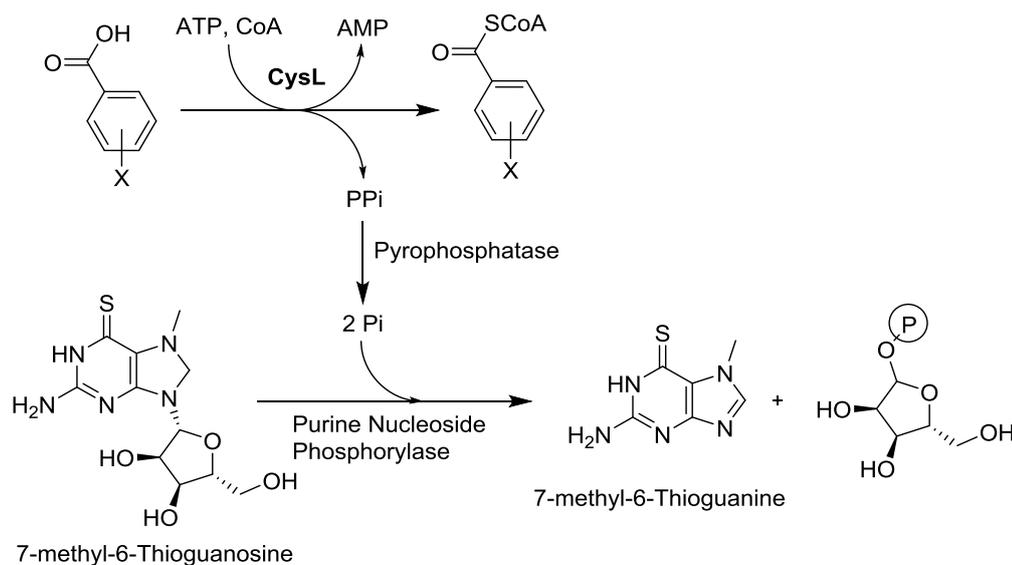


Figure 2.1: PNP-MESG reporter system reaction

For a final concentration of 50 nM CysL, 0,05 UI pyrophosphatase, 0,5 UI PNP, 100  $\mu$ M MESG, 1 mM ATP and 1mM CoA in 25 mM TRIS, 150 mM NaCl, 10 mM  $\text{MgCl}_2$  a master mix was prepared at a 1,05x concentration and incubated 10 minutes at room temperature to get rid of the excess phosphate contained in ATP and CoA. The mix can be kept on ice for at least 3h before incubation at room temperature prior to the assay. 95  $\mu$ l of the mix are added on 5  $\mu$ l of a 20x solution of the substrate.

The assay was done in a 96 well plate but each well was measured separately as small variations in the pipetted volume can have dramatic effects and multichannel pipets lack the

required precision. The absorbance at 360 nm is measured every 500 ms for 3 minutes. pABA concentrations for plot relevance and a good curve quality were usually between 20 nM and 0,2  $\mu$ M.

## 2.4 Reaction scale-up

The scaled up reaction was performed on in a 10 ml volume with 10 mg CoA-SH ( $\approx$ 2 mM), 5 mM pABA or 3-OH-pABA or 3-OMe-pABA, 5mM ATP In 25 mM TRIS, 150 mM NaCl, 10 mM MgCl<sub>2</sub>, pH 7,5 left to stir in a 50 ml falcon tube overnight at room temperature. At this point the remaining ATP was gotten rid of by adding 5 mM glucose and 0,05 UI hexokinase. The purification of the pABA-CoA derivatives was performed on a CHROMABOND SA cation exchange SPE cartridge (Macherey-Nagel), the binding was done at pH 2,5 after addition of formic acid, the column was then washed with 1% formic acid to get rid of adenylates and glucose. The final elution was performed at pH4,5 in ammonium formiate to retrieve pure pABA-CoA esters. The eluate was neutralized with ammonium hydroxide and freeze dried. Purity was assessed through LC-MS.

## 2.5 Inhibition assay

pABA-CoA ligation reactions were realized in 250  $\mu$ l volume as described previously in the presence of increasing concentrations of Cystobactamides 919-1 and 919-2. 1  $\mu$ M CysL was preincubated for 4 h at room temperature with Cystobactamide 919-1 and 919-2 in the reaction buffer before addition of 1 mM pABA, ATP and CoA. 50  $\mu$ l samples were taken at t=0; 5; 10; 20; 60 min and quenched by addition of 50  $\mu$ l MeOH for subsequent LC-MS analysis.

# 3 Results and discussions

Biochemical investigations on CysL revealed relaxed substrate specificity, the enzyme accepting a broad range of benzoate derivatives. The enzyme also showed very fast kinetics for a secondary metabolism enzyme with complete turnover rates in high volumes, making it a good candidate for biotechnological applications. In the scope of biosynthesis the enzyme displays a rather striking feedback inhibition by the final product of the assembly line, a rarely characterized phenomenon in secondary metabolism.(14)

## 3.1 Activity on pABA and derivatives

CysL was expressed in high yields as His<sub>6</sub>-tagged fusion protein in *E.coli* and purified through nickel affinity chromatography and size exclusion chromatography. Contrary to previous reports (15) the protein elutes as a monomer from the column which might be related to the di-domain architecture. The identity of the protein was confirmed by SDS-

PAGE at about 115 kDa. CoA ligation activity on para-amino benzoate and a variety of other benzoate derivatives was analyzed through LC-MS.

Substrate	Activity (%)
4-HydroxyBA	100 ( $\pm$ 3,5)
BA	83 ( $\pm$ 8)
2-Amino-5-HydroxyBA	41 ( $\pm$ 1)
3,4-diHydroxyBA	40 ( $\pm$ 5)
3,4-diAminoBA	35 ( $\pm$ 0,5)
4-AminoBA	26 ( $\pm$ 3)
4-Amino-3-HydroxyBA	19 ( $\pm$ 1,5)
4-Amino-3-MethoxyBA	18 ( $\pm$ 1)
4-ChloroBA	16 ( $\pm$ 2)
3-AminoBA	15 ( $\pm$ 2)
2-AminoBA	12 ( $\pm$ 1)
2-MethylBA	12 ( $\pm$ 1)
4-isoPropylBA	0.5 ( $\pm$ 0,5)
3,4-diChloroBA	0.2 ( $\pm$ 0)
2,3,4-triHydroxyBA	n.d
2,3-diHydroxyBA	n.d
2-Hydroxy-3-MethoxyBA	n.d
2-HydroxyBA	n.d

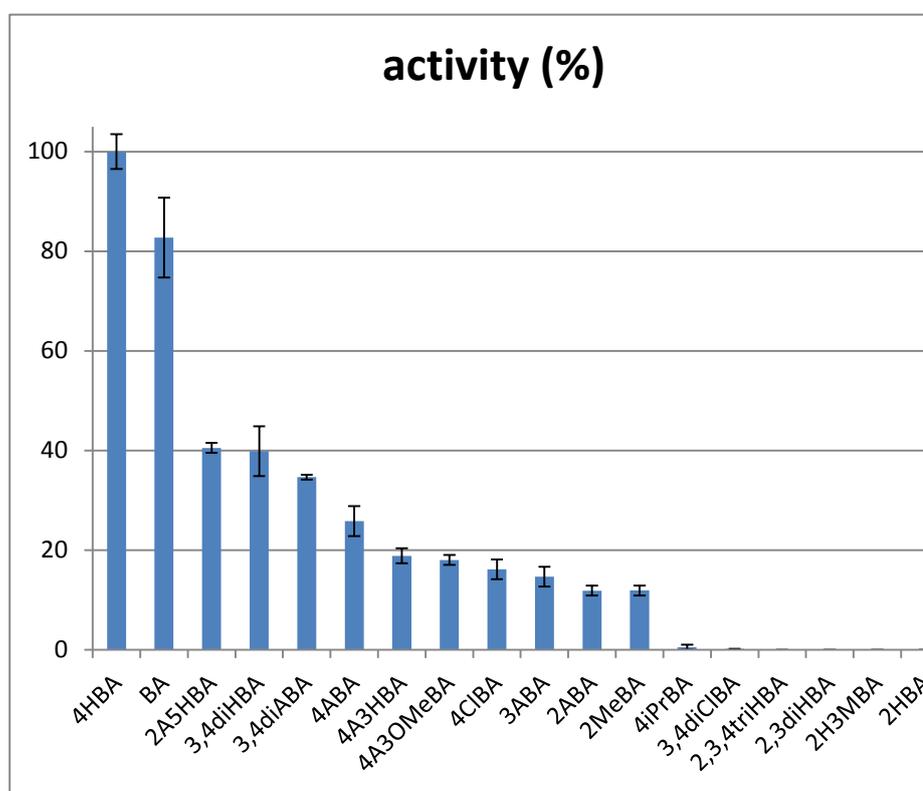


Figure 2.2: Relative activity of CysL on various benzoic acid derivatives (BA: benzoic acid; n.d: not detected)

CysL shows a broad substrate tolerance but, surprisingly, it shows slower kinetics on 4-amino benzoic acid which is the expected substrate in comparison to benzoic acid and 4-hydroxy- or 3,4-dihydroxy-benzoic acid. Most modified benzoates harboring a polar group in positions *ortho*, *meta* or *para* were efficiently processed except derivative showing a hydroxyl in *ortho* which seems to prevent the CoA adenylation, probably due to the pseudo ring structure it forms through hydrogen bonding with the adjacent acid. 4-hydroxy-benzoate is the preferred substrate despite it not being implicated in Cystobactamide biosynthesis. It is noteworthy that the homologous protein AlbVII which can be found in the biosynthesis gene cluster of the related antibiotic Albicidin was postulated to activate 4-hydroxy-benzoate to use as a starter unit for its first PKS module.

### 3.2 Reaction kinetics

The biochemical kinetic constants of CysL were measured thanks to the purine nucleoside phosphorylase / methyl-mercapto-guanosine (PNP/MESG) assay allowing the indirect spectrophotometric titration of phosphate released during the adenylation. This assay

## CysL - Investigation of a secondary metabolism related para-aminobenzoate CoA ligase and its feedback inhibition by Cystobactamides

is based on the phosphate dependent cleavage of 7-methyl-6-thioguanosine (MESG,  $\lambda_{\max} = 320 \text{ nm}$ ) in 7-methyl-6-thioguanine ( $\lambda_{\max} = 360 \text{ nm}$ ) and ribose-phosphate by the purine nucleoside phosphorylase (PNP). The assay conditions had to be slightly modified for the fast paced reaction kinetic from CysL. Initial velocities with varying enzyme concentrations were measured while the limiting substrate concentration was fixed at  $100 \mu\text{M}$  pABA for a large excess of ATP and CoA ( $1 \text{ mM}$ ). For the actual  $K_m$  measurement an enzyme concentration of  $50 \text{ nM}$  was chosen with varying pABA concentrations. The apparent  $K_m$  of the reaction with pABA was determined to be at  $823 \text{ nM}$  with a  $V_m$  of  $55,6 \text{ nM/s}$  which is a hundredfold lower than other reported benzoate CoA.(15,16)

Secondary metabolism enzymes are generally considered to have slower kinetic parameters than primary metabolism enzymes, however CysL probably originates from a primary aromatic acid catabolism pathway which could explain that it retained a high velocity. Additionally, pABA is usually funneled into the tetrahydrofolate biosynthesis which is a major actor of the primary metabolism since it is central for the one carbon metabolism. Thus it seems necessary for the Cystobactamide biosynthesis to immobilize a pool of pABA to be modified and used for secondary metabolism.

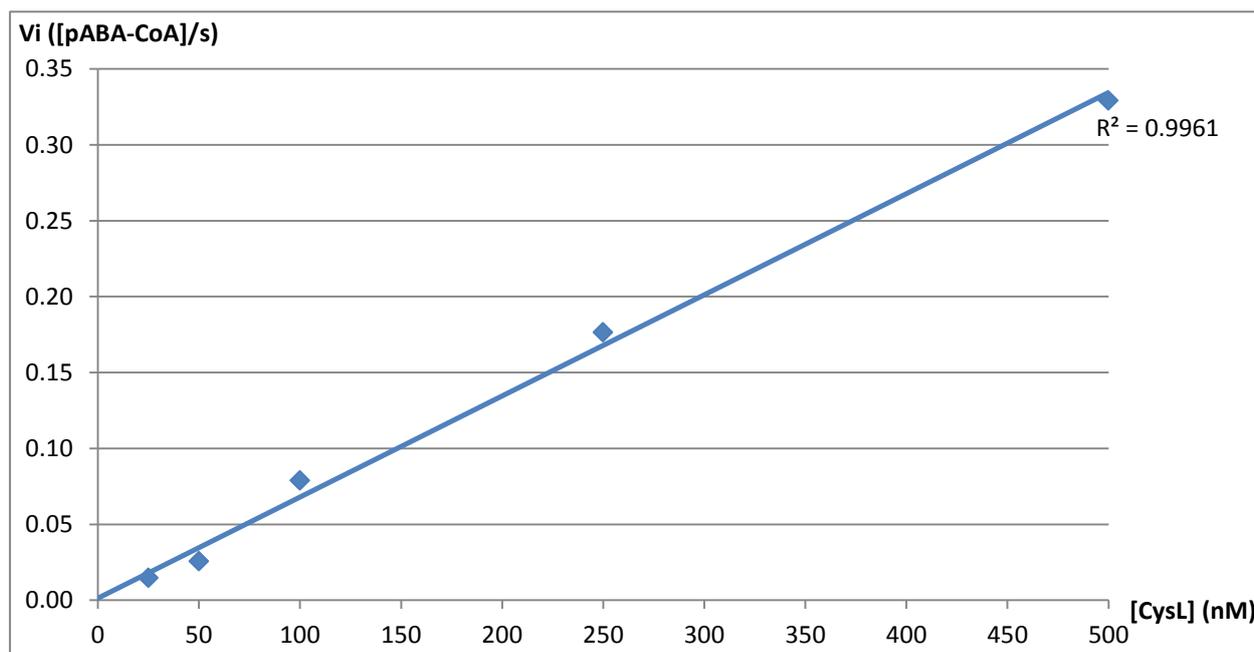


Figure 2.3: velocity against enzyme concentration plot

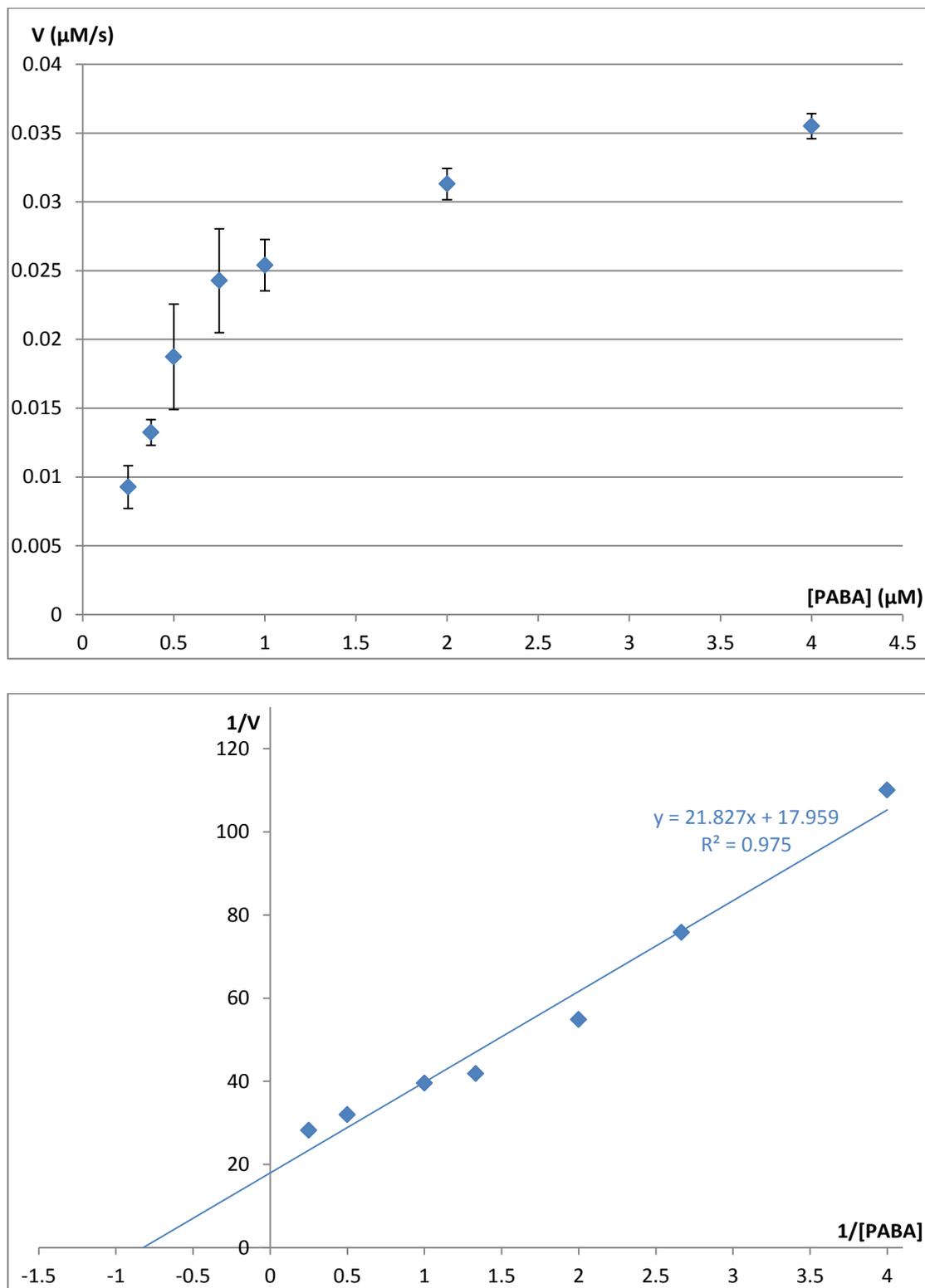


Figure 2.4: Michaelis-Menten kinetic plot of CysL on PABA and Lineweaver-Burk plot used to determine the kinetic parameters

### 3.3 Reaction scale up and purification of CoA derivatives

The production of mg scale amounts of pure benzoate CoA derivatives is relevant for the investigation of subsequent tailoring steps in the Cystobactamides biosynthesis. Since CysL showed to be very active and very stable in solution even in large volumes and important concentrations, it was considered a viable option for biochemical production of these products rather than using organic synthesis.<sup>(17)</sup> Since free CoA was the most problematic substrate to separate from the pABA-CoA products it was set as the limiting factor and the reaction was still driven to completion by a twofold excess of the other substrates (pABA and ATP). ATP also proved to be hard to separate from CoA esters, but since AMP and ADP did not pose any problem, excess ATP was gotten rid of through addition of hexokinase and glucose after completion of the coadenylation. Using this setup the reaction was up scaled for pABA, 3OH-pABA and 3OMe-pABA to a 10 ml volume with 10 mg CoA-SH.

The purification of the pABA-CoA derivatives was performed on a cation exchange SPE cartridge and purity was assessed through LC-MS. Yields ranged from 56% for the 3-OH-pABA to 45% for the pABA which is quite low compared to what could have been obtained though semi preparative LC-MS but was a much faster process.

### 3.4 Inhibition effect of Cystobactamides

Inhibition of the CoA ligation upon addition of Cystobactamides was assayed through LC-MS by comparison of the AUC for the product of the reaction over time with a large excess of substrates. The results show an important decrease in the pABA-CoA production by CysL in the presence of Cystobactamides. The inhibition is concentration dependent and seems more pronounced with Cystobactamide 919-2 than 919-1, the reaction is completely abolished with a 1000 fold excess of the inhibitor over the enzyme. The form of the curves hint at a non-competitive inhibition but a further assay in order to determine a  $K_i$  for the inhibitors would have to be conducted to confirm this.

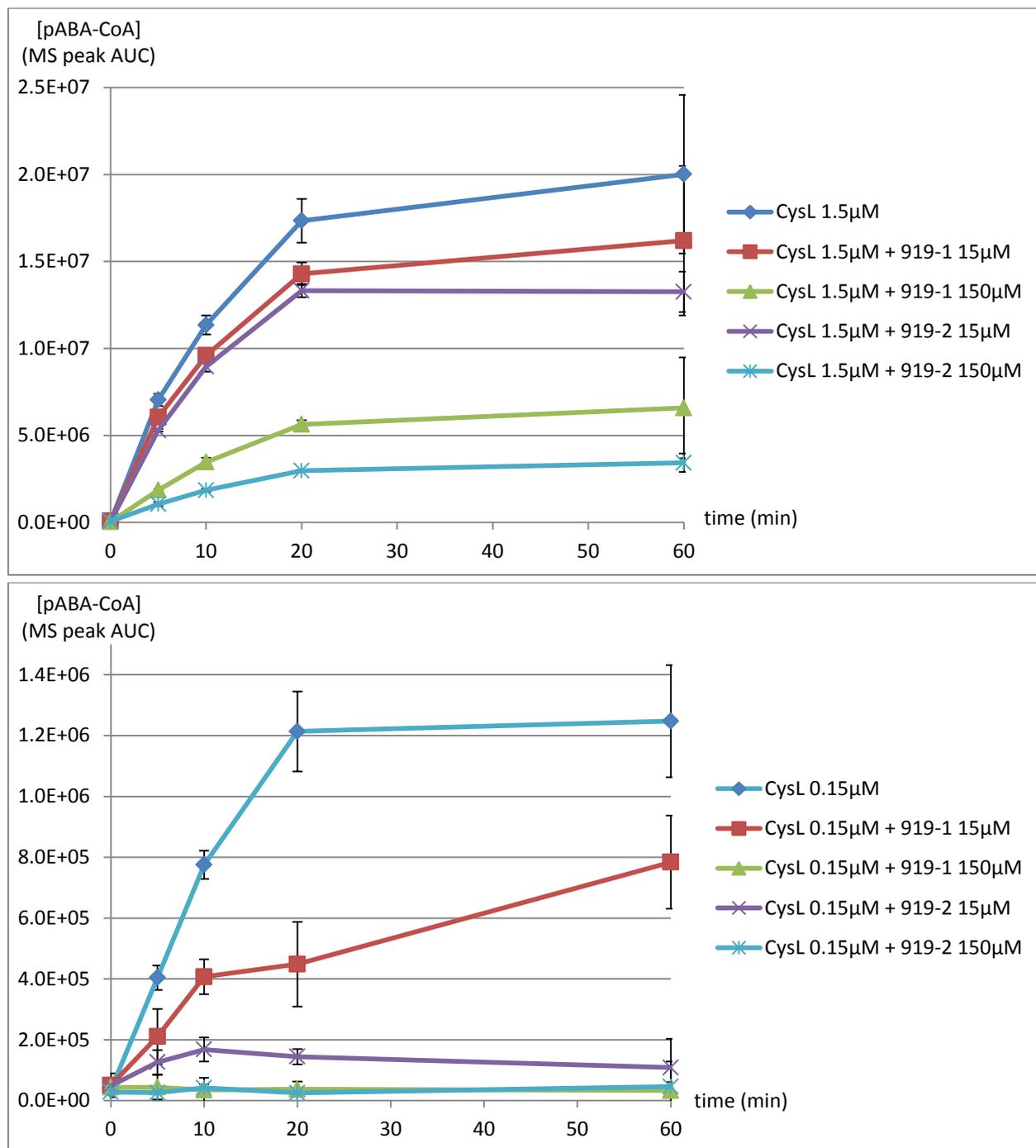


Figure 2.5: graphs measuring the pABA-CoA concentration (MS peak AUC) function of the reaction time (minutes) with different CysL concentrations and different Cystobactamide concentrations

In order to confirm that the inhibition of the CoA ligation is related to a binding of Cystobactamides on CysL an SPR experiment was conducted. CysL was the ligand bound to the chip and Cystobactamide 919-1 was chosen as the analyte. The reference chip was loaded with BSA to rule out unspecific protein interactions. However, the signal did not reach saturation which prevented estimation of a  $K_d$  through this experiment. Since biochemical assays hinted at higher inhibition levels with Cystobactamide 919-2 a new experiment was conducted using this analyte over 919-1, this time an MST experiment was conducted with

## CysL - Investigation of a secondary metabolism related para-aminobenzoate CoA ligase and its feedback inhibition by Cystobactamides

10 nM CysL incubated with increasing concentrations of the ligand. While saturation was not reached in this experiment either a  $K_d$  could be proposed at  $183 \pm 42 \mu\text{M}$ . Although this very high value hints at low affinity it is noteworthy that long incubation times are needed to observe inhibition in the biochemical assay implying that the on-off rate of the interaction might be very slow.

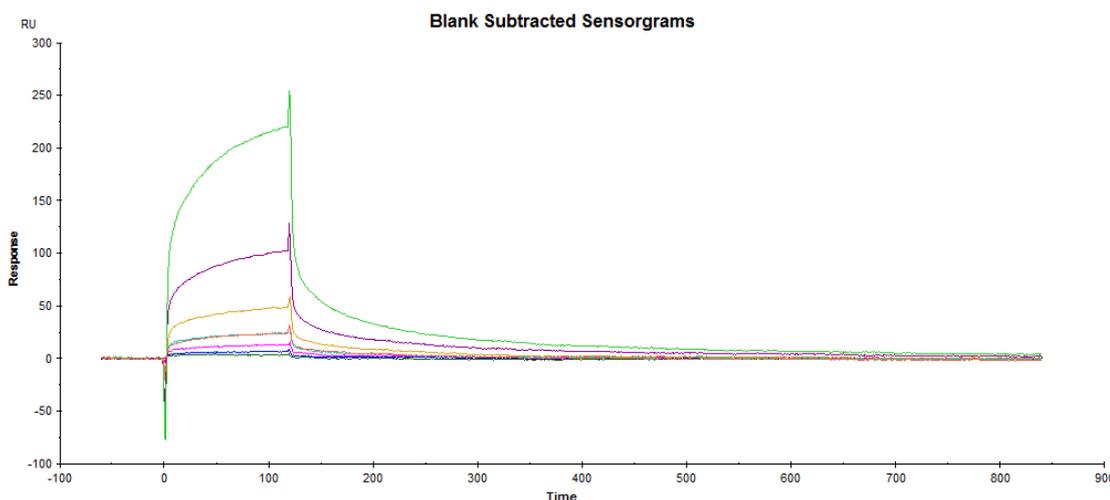


Figure 2.6: Blank subtracted sensorgram, response with increasing concentrations of Cystobactamide 919-1 at 0,5µg/l (dark green) 1 µg/l (blue) 2 µg/l (pink) 4 µg/l (red) 8 µg/l (yellow) 16 µg/l (purple) 32 µg/l (green) and back to 4µg/l (light blue)

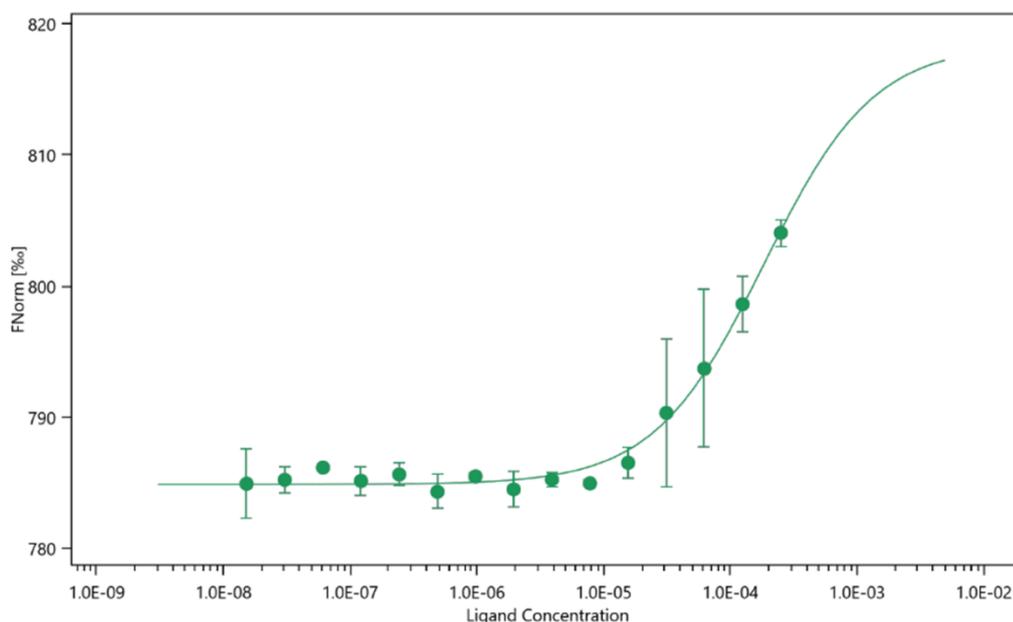


Figure 2.7 : MST measurement of the affinity between 10 nM CysL and increasing concentrations of Cystobactamide 919-2,  $K_d$  was calculated at  $183 \pm 42 \mu\text{M}$

Secondary metabolism regulation is usually transcriptional. In bacteria specific promoters respond to various environmental or internal stimuli to up- or down-regulate the level of gene expression of the cluster. In primary metabolism the feedback inhibition loop is a well-known mechanism which is at play in most pathways to regulate the amount of key

intermediates, however in secondary metabolism this type of inhibition has rarely been observed and biochemically proven before. It is interesting to observe that this feedback inhibition occurs on the enzyme that derails a primary metabolism intermediate, para-amino benzoate, from its original biosynthetic destination, tetrahydrofolate. The low  $K_m$  of the enzyme implies that it needs a secondary regulation to prevent the accumulation of pABA-CoA in the cell.

### 3.5 Module separation

To confirm that the binding does take place on the first domain of CysL they both have to be expressed separately and biochemically characterized, also structural analysis could give insights on the binding mechanism. Unfortunately, completely separate domains could not be over-expressed and thus two new constructs were designed for coexpression of an N terminal tagged CysL1 in a pHisTEV vector and a C terminal tagged CysL2 in a pACYC vector. This strategy allows the free C-terminus of CysL1 to interact with the equally free N terminus of CysL2 and improved the yield of both domains. Finally, differential tag removal allows separating both domains.

The C-terminal module precipitated immediately after separation of the domains and even if the identity of the N-terminal domain could be confirmed through in gel trypsin digest coupled with MALDI-TOF it did not prove stable enough for further investigation.

### 3.6 CysL in the biosynthesis

The presence of a CoA ligase in a pure NRPS pathway is surprising since amino acids are usually processed in a free form by the adenylation domains for a direct loading onto the phosphopantetheine arm of the carrier proteins. pABA is a central intermediate in the folate metabolism leading to THF synthesis. It is possible that CysL is able to immobilize a consistent pool of pABA to be used for the Cystobactamide biosynthesis, the size of this pool being tightly regulated via a feedback inhibition loop. On the other hand an important tailoring step is the oxidation of the two last pABA units as 3-hydroxy-pABA and 2,3-dihydroxy-pABA. This step is carried out by CysC which is homologous to the benzoate oxidase BoxB found in *Azocardus evansii* a dioxygenase which requires a CoA activated substrate.(18)

The acceptance of CoA activated substrates by adenylation domains is still puzzling but since A domains have been shown to act as CoA ligase in the absence of carrier protein (19) it is likely that they could be able to catalyze the reverse reaction leading to the formation of an adenylate from a CoA. This adenylate would then be a viable substrate for the attack by the free thiol of the phosphopantetheine arm.

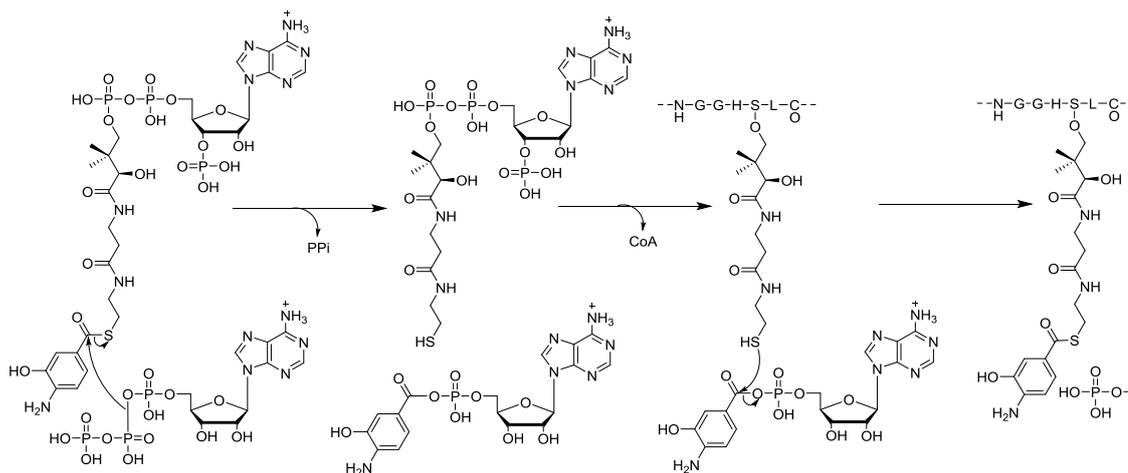


Figure 2.8: Proposed mechanism leading to the loading of a CoA bound intermediate by an adenylation domain

## 4 Conclusion

We were able to express CysL to high yields in stable monomeric form. Investigation on the activity *in vitro* revealed extremely swift kinetic parameters on the main substrate para-amino benzoate. The substrate specificity is however quite low and a number of benzoate derivatives were accepted but no side reactions are observed. These characteristics designated CysL as a viable candidate for biotechnological applications and the reaction could be easily scaled up to preparative scale in order to synthesize para-amino benzoyl CoA and other derivatives in a convenient biochemical setup.

CysL can be considered the first enzyme of the Cystobactamide biosynthesis and could be responsible for immobilizing a pool of pABA from the primary metabolism for subsequent processing by the biosynthetic machinery. An important feedback inhibition loop is regulating the enzymatic activity through binding of the final products, Cystobactamides, to a previously unknown “Cystobactamide binding domain”. Further structural insight would however be needed to fully characterize this new domain.

## 5 References

1. Carmona, M., Zamarro, M. T., Blázquez, B., Durante-Rodríguez, G., Juárez, J. F., Valderrama, J. A., Barragán, M. J. L., García, J. L., and Díaz, E. (2009) Anaerobic catabolism of aromatic compounds: a genetic and genomic view, *Microbiology and molecular biology reviews : MMBR* 73, 71–133.
2. Villemur, R. (1995) Coenzyme A ligases involved in anaerobic biodegradation of aromatic compounds, *Canadian journal of microbiology* 41, 855–861.
3. López Barragán, M. J., Carmona, M., Zamarro, M. T., Thiele, B., Boll, M., Fuchs, G., García, J. L., and Díaz, E. (2004) The bzd gene cluster, coding for anaerobic benzoate catabolism, in *Azoarcus* sp. strain CIB, *Journal of bacteriology* 186, 5762–5774.
4. Schuhle, K., Gescher, J., Feil, U., Paul, M., Jahn, M., Schagger, H., and Fuchs, G. (2003) Benzoate-Coenzyme A Ligase from *Thauera aromatica*. An Enzyme Acting in Anaerobic and Aerobic Pathways, *Journal of bacteriology* 185, 4920–4929.
5. Piel, J., Hertweck, C., Shipley, P. R., Hunt, D. M., Newman, M. S., and Moore, B. S. (2000) Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate 'Streptomyces maritimus': evidence for the derailment of an aromatic polyketide synthase, *Chemistry & biology* 7, 943–955.
6. Ligon, J., Hill, S., Beck, J., Zirkle, R., Molnár, I., Zawodny, J., Money, S., and Schupp, T. (2002) Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from *Sorangium cellulosum* So ce26, *Gene* 285, 257–267.
7. Walsh, C. T., Haynes, S. W., and Ames, B. D. (2012) Aminobenzoates as building blocks for natural product assembly lines, *Natural product reports* 29, 37–59.
8. Gil, J. A., and Campelo-Diez, A. B. (2003) Candicidin biosynthesis in *Streptomyces griseus*, *Applied microbiology and biotechnology* 60, 633–642.
9. Cociancich, S., Pesic, A., Petras, D., Uhlmann, S., Kretz, J., Schubert, V., Vieweg, L., Duplan, S., Marguerettaz, M., Noell, J., Pieretti, I., Hugelland, M., Kemper, S., Mainz, A., Rott, P., Royer, M., and Sussmuth, R. D. (2015) The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine, *Nature chemical biology* 11, 195–197.
10. Baumann, S., Herrmann, J., Raju, R., Steinmetz, H., Mohr, K. I., Hüttel, S., Harmrolfs, K., Stadler, M., and Müller, R. (2014) Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity, *Angewandte Chemie (International ed. in English)* 53, 14605–14609.
11. Schmelz, S., and Naismith, J. H. (2009) Adenylate-forming enzymes, *Current opinion in structural biology* 19, 666–671.
12. Reger, A. S., Carney, J. M., and Gulick, A. M. (2007) Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl-CoA synthetase, *Biochemistry* 46, 6536–6546.
13. Duckworth, B. P., Wilson, D. J., and Aldrich, C. C. (2016) Measurement of Nonribosomal Peptide Synthetase Adenylation Domain Activity Using a Continuous Hydroxylamine Release Assay, *Methods in molecular biology (Clifton, N.J.)* 1401, 53–61.

14. McFarlane, J., Madyastha, K. M., and Coscia, C. J. (1975) Regulation of secondary metabolism in higher plants. Effect of alkaloids on a cytochrome P-450 dependent monooxygenase, *Biochemical and Biophysical Research Communications* 66, 1263–1269.
15. Bains, J., and Boulanger, M. J. (2007) Biochemical and structural characterization of the paralogous benzoate CoA ligases from *Burkholderia xenovorans* LB400: defining the entry point into the novel benzoate oxidation (box) pathway, *Journal of molecular biology* 373, 965–977.
16. Altenschmidt, U., Oswald, B., and Fuchs, G. (1991) Purification and characterization of benzoate-coenzyme A ligase and 2-aminobenzoate-coenzyme A ligases from a denitrifying *Pseudomonas* sp, *J. Bacteriol.* 173, 5494–5501.
17. Beuerle, T., and Pichersky, E. (2002) Enzymatic synthesis and purification of aromatic coenzyme a esters, *Analytical Biochemistry* 302, 305–312.
18. Gescher, J., Eisenreich, W., Wörth, J., Bacher, A., and Fuchs, G. (2005) Aerobic benzoyl-CoA catabolic pathway in *Azoarcus evansii*: studies on the non-oxygenolytic ring cleavage enzyme, *Molecular microbiology* 56, 1586–1600.
19. Linne, U., Schäfer, A., Stubbs, M. T., and Marahiel, M. A. (2007) Aminoacyl-coenzyme A synthesis catalyzed by adenylation domains, *FEBS Letters* 581, 905–910.

# Chapter 3: CysF and CysS – Biosynthesis of an uncommon isopropoxyl decoration by iterative methylation

## Introduction

Methylation is one of the most frequent modifications used by living organisms, it can happen on a huge variety of biologically active substrates including secondary metabolites, hormones, neurotransmitters, lipids, proteins and nucleic acids.(1) This rather simple modification can have drastic effects on the physicochemical properties of the substrate. Methylation can thus be used for diverse biological functions on diverse substrates including biosynthesis, metabolism, detoxification, signal transduction and DNA processing. In the case of antibacterial natural products which are often enzyme inhibitors, the methylation is often a key modification to allow the binding of the product to its target, concurrently a methylation can also prevent resistance proteins from binding to the product.(2) It is interesting to note that nature has devised various methyl transfer mechanisms allowing the methylation of almost any position whether it is reactive or not.(3)

Methyl groups are usually originating in the one carbon metabolism which mostly relies on the methyl-tetrahydrofolate (THF) pool replenished by glycine, serine or formiate. But the methylation cofactor used by most enzymes is S-adenosyl methionine (SAM) which is the second most widely used enzyme substrate after ATP.(4) The reaction leading from SAM to S-adenosyl homocysteine (SAH) and a methyl group via nucleophilic substitution is extremely favorable due to the electron deficient methylsulfonium center. SAM is usually the cofactor used

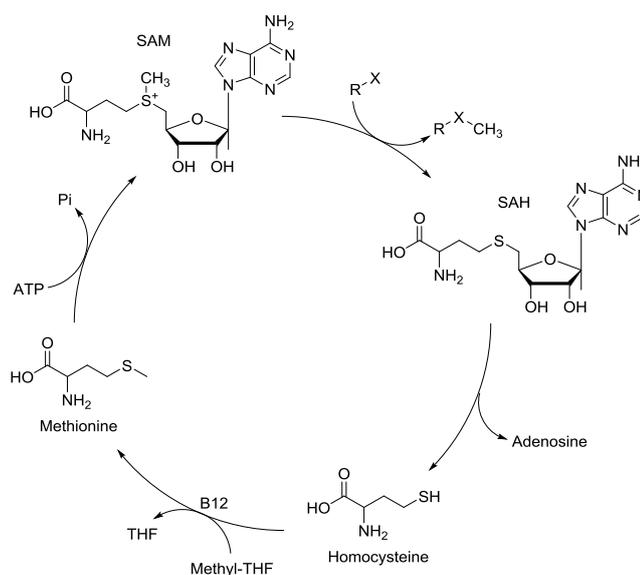


Figure 3.1: SAM regeneration cycle

## CysF and CysS – Biosynthesis of an uncommon isopropoxyl decoration by iterative methylation

for methylation and it is regenerated from SAH through the loss of the adenosine and the subsequent methylation of homocysteine by a B12 dependent methionine synthase using methyl-THF coming from the one carbon folate metabolism.(5) (Figure 3.1)

Depending on their chemical mechanism two families of methyl transferases can be defined. Classical methyl transferases (MT) rely on the nucleophilic attack by heteroatoms or activated carbons whereas radical SAM methyl transferases (RSMT) rely on a radical mechanism through formation of a deoxyadenosyl radical for the methylation of unreactive centers, such as  $sp^3$  carbons.(6) In the Cystobactamide biosynthesis both mechanisms are at play sequentially to produce isopropoxyl groups. (Figure 3.2)

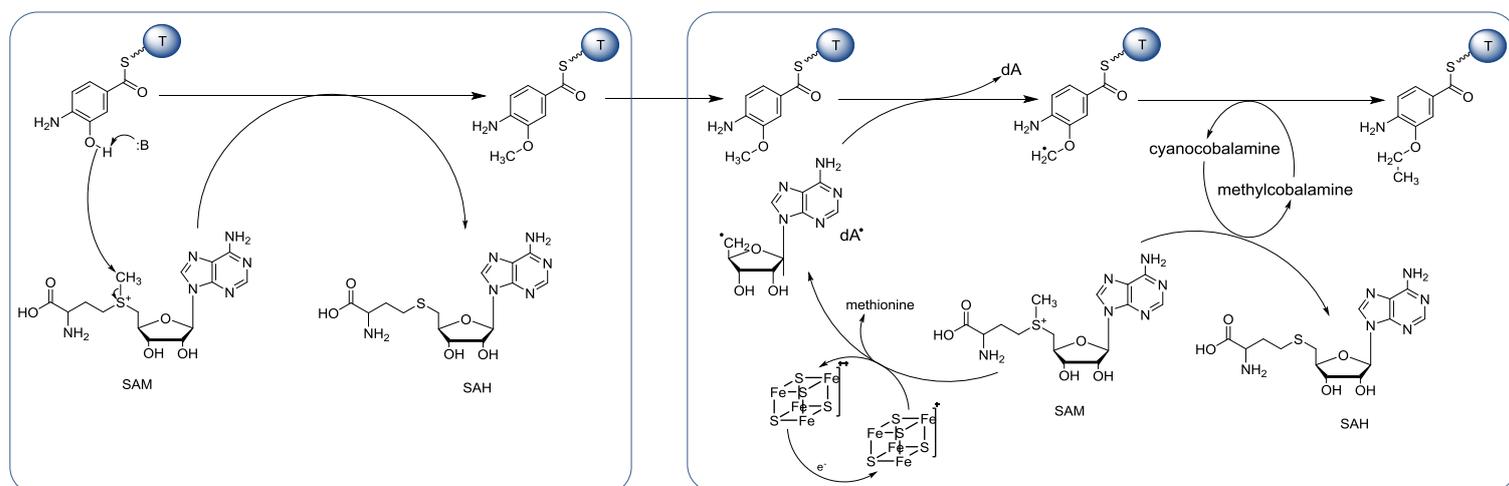


Figure 3.2: CysF and CysS as examples for a Classical MT and a radical SAM MT

Classical methyl transferases are a very wide family of enzymes that display a huge diversity of substrates and sequences, and thus of catalytic folds. They are divided in five major classes but the majority of the enzymes belong to class I, natural product methyl transferases themselves are found only in classes I and III.(2) Biochemically they rely on the reactivity of their substrate by bringing in close proximity a nucleophilic position to be methylated with the positively charged methylsulfonium group of SAM for a classical  $SN_2$  nucleophilic substitution reaction. On a structural level class I MT are based on a C-terminal Core Rossmann fold consisting of alternating  $\beta$  strands ( $\beta_1$ – $\beta_7$ ) and  $\alpha$  helices forming a central seven-stranded  $\beta$  sheet with three helices on each side with the  $\beta$  strand 7 antiparallel to the others ( $6 \uparrow 7 \downarrow 5 \uparrow 4 \uparrow 1 \uparrow 2 \uparrow 3 \uparrow$ ). The Hallmark GxGxG nucleotide binding motif at the end of the first  $\beta$  strand is responsible for SAM binding.(7) The N-terminus on the other hand can vary in size and structure, it is usually related to substrate binding or protein-protein interaction. (Figure 3.3)

Radical SAM enzymes are mechanistically and structurally unrelated to classical methyl transferases and constitute a separate family, they are metalloproteins harboring a [4Fe-4S] cluster for SAM binding and can be easily identified by the CxxxCxxC binding site for the Fe-S cluster.(8) They are subdivided in three classes (A, B and C) depending on

sequence and structure which are most of the time responsible for different biochemical processes such as methylation, oxidation or amine transfer.(6) Natural products radical SAM methyltransferases mostly belong to the Class B which features an additional N-term cobalamin (vitamin B12) binding domain. This family of enzymes rely on the reductive cleavage of SAM bound to  $[4Fe-4S]^+$  cluster to produce methionine and a 5'-deoxyadenosine radical, followed by hydrogen atom abstraction on an unreactive position to be methylated by the B12 methyl cobalamin cofactor.(9) Structurally Radical SAM present diverse N and C termini but possess a central radical SAM domain constituted from a partial  $(\alpha/\beta)_6$  TIM barrel core harboring the  $[4Fe-4S]$  cluster with a B12 binding Rossmann fold on top of it.(10)

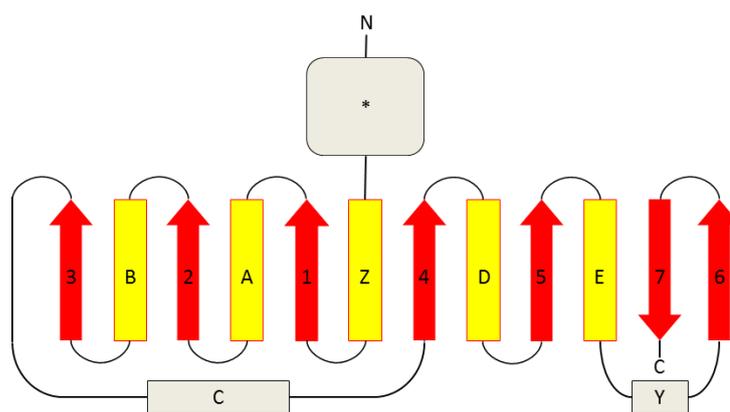


Figure 3.3: Schematic representation of the methyltransferase core Rossmann fold. Helixes are shown as yellow cylinders, strands as red arrows, the variable n terminus is visible in grey and helixes C and Y are shown in grey because they are not always conserved

## Chapter 3.1: CysF - Biochemical and Structural investigation of an *in trans* acting NRPS methyl transferase

### 1 Introduction

As part of our efforts to investigate the structural and mechanistic enzymology of the biosynthetic pathway of Cystobactamides we have studied the enzymes responsible for the formation of the meta-isopropoxyl group on the two last para-amino benzoates (pABA) of the Cystobactamide backbone. This biosynthesis system is thought to proceed from a meta-hydroxy-pABA by successive methylation by a regular O-methyl transferase, CysF, and a radical-SAM C-methyl transferase, CysS. CysF is homologous to a variety of natural products methyl transferases (NPMT) mostly from bacterial origin, indeed *in silico* analysis revealed that it belongs to the Class I Methyl transferases along with the catechol O-methyltransferase (COMT).(2) This was confirmed through structural and biochemical data, furthermore initial activity assays revealed that the enzyme, just as its radical-SAM counterpart CysS, requires activation of the benzoate as thioester intermediate to be able to methylate the meta hydroxyl group.(11)

Methyl transferases are ubiquitous enzymes that process a number of biological molecules (1) rarely requiring an activation of the substrate prior to methylation, this phenomenon was only reported for 5 other characterized methyl transferases.(12–16) *In Silico* analysis of the protein sequences of these 5 enzymes in parallel with CysF revealed that they can be clustered in two groups, despite the similarity of their reaction CysF does not cluster with the plant caffeoyl-CoA O-methyl transferase (CCoAOMT) and closely related DOPA-CoA 4O-methyl transferase (SafC) but rather with the orselinate-CoA 2O-methyl transferase (CalO6) and the 6-methylsalicylyl-CoA 3C-methyl transferase (PokMT1). These two enzymes were proposed to be active *in vivo* on carrier protein bound substrate rather than on CoA bound substrate. The required activation of the aromatic ring can be explained by the poor nucleophilicity of the phenols which would result in slow reaction rates of the SN2 based methylation by SAM dependent methyltransferases, however in the case where a thioester is present on the aromatic ring, the phenolate specie will form readily thus increasing the reaction rate of the enzyme.

To be able to investigate the *in trans* activity of CysF a number of challenges had to be overcome regarding the expression of the multidomain NRPS modules of CysG, we managed to overcome these challenges through the innovative use of two proteins present in the biosynthesis cluster as tools for the protein overexpression. The MbtH type Adenylation activator (17) CysA proved necessary to be coexpressed with the NRPS modules to yield decent amounts of soluble protein, furthermore the thioesterase II (18) CysE was used to remove *in vivo* loaded pABA from the modules in order to generate “clean” unloaded NRPS modules in *holo* form.

## 2 Materials and Methods

### 2.1 Cloning, overexpression and purification

Standard protocols were used for DNA cloning, transformation, amplification, and purification. Genomic DNA was extracted from *Cystobacter velatus* Cbv34 grown in CTT medium, the 1038 bp DNA fragment encoding CysF was amplified using the forward primer 5'TATCATATGACCGCTCAGAACCAAGCCTCC and the reverse primer 5'TACTCGAGTC-AGAGCTGCGCTTTGCCCTTGATCG. The amplified DNA fragments were digested with NdeI and XhoI, and cloned into pET-28b with an N-terminal His6 tag. The resulting construct was sequenced to verify that no mutation had been introduced during PCR.

Recombinant protein expression was carried out in *E.coli* BL21 (DE3) grown in LB medium supplemented with 50 µg/ml of kanamycin. The culture was inoculated 1/10 from a fully grown overnight culture and was initially grown at 37 °C until it reached an OD of 0.6 to 0.9, at which point the temperature was decreased to 16 °C. Induction was initiated after 30 minutes at 16°C with 0,1 mM isopropyl-β,D-thiogalactopyranoside (IPTG). The cells were harvested 16h after induction by centrifugation at 8000 g for 10 minutes at 4°C and resuspended in lysis buffer (25 mM TRIS pH7,5; 150 mM NaCl; 20 mM imidazole). The cells were lysed by passage at 1500 bar at 4°C through a M-110P Microfluidizer (Microfluidics) after which the crude extract was centrifuged at 45000 g for 15 minutes at 4°C, the supernatant was treated as soluble protein fraction.

Affinity chromatography was performed on an Äkta Avant system while size exclusion chromatography was performed on an Äkta Pure system. The soluble fraction of the lysed cells was applied to 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10CV lysis buffer prior to elution with elution buffer (25 mM TRIS pH7,5; 150 mM NaCl; 250 mM imidazole). The pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200 µg SEC column equilibrated in running buffer (25 mM TRIS pH7,5; 150 mM NaCl; 2 mM DTT). The purity of the pooled fractions was checked via SDS-PAGE, and the protein was concentrated and frozen at -80°C after addition of 25% glycerol.

### 2.2 <sup>13</sup>C methionine feeding

25 mL M-medium was inoculated with 1mL of a high density overnight culture of Cbv34. 0,5 mL of XAD-7 was added on day 1 and 12,5 mg of [<sup>13</sup>C-methyl]-L-methionine was fed over 4 days. On day 5 the cultures were centrifuged and the pellet and XAD were extracted with 2x10 mL methanol. The solvent was removed through rotary evaporation and the extract was dissolved in 2 mL methanol.

LC-ESI-MS was performed using a Thermo Scientific Dionex Ultimate 3000 UHPLC system coupled with a Bruker Amazon speed iontrap mass spectrometer using an ESI in positive mode. LC was performed on a Waters Acquity UPLC BEH C18 50 mm 1.7  $\mu$ m column. LC conditions: A-Water + 0.1 % formic acid; B-Acetonitrile + 0.1 % formic acid; 0 min: 95% A / 5% B, 0.5 min: 95 % A / 5 % B, 9.5 min: 5 % A / 95 % B, 10.5 min: 5% A / 95 % B, 10.8 min: 95 % A / 5 % B, 11.8 min: 95 % A / 5 % B.

### 2.3 *In vitro* activity assays on CoA

The 3OH-pABA CoA was generated *in situ* by enzymatic reaction with CysL (Cf. Chap. 2). In a total volume of 50  $\mu$ l, 1mM 3OH-pABA was mixed with 1 mM ATP, 1 mM CoA and 1 mM SAM in 25 mM TRIS pH7,5; 150 mM NaCl, 10 mM MgCl<sub>2</sub>. 1  $\mu$ M CysL was added and the reaction was incubated at room temperature for 10 minutes to allow the formation of 3OH-pABA CoA. The methylation reaction was started by adding 1 $\mu$ M CysF and incubated 20 minutes at room temperature. The enzymes were precipitated by addition of 50  $\mu$ l MeOH and centrifuged at 150000 g for 5 minutes in a tabletop centrifuge. The supernatant was analyzed by LC-MS.

CoA thioesters were analyzed on a Dionex Ultimate 3000 UPLC system (Thermo Scientific) coupled with an Amazon ion trap MS (Bruker) using an ESI in positive mode. The samples were run on a Synergi fusion 4 $\mu$  Polar-RP 80A 250X3,0 mm C18 column (Phenomenex). LC conditions: A-Water + 5 mM Ammonium formiate; B-Methanol + 5 mM Ammonium formiate; 0 min: 98 % A / 2 % B, 1 min: 98 % A / 2 % B, 16 min: 80 % A / 20 % B, 18 min: 5 % A / 95 % B, 19 min: 5 % A / 95 % B, 20 min: 98 % A / 2 % B.

### 2.4 CysG cloning, overexpression and purification

Standard protocols were used for DNA cloning, transformation, amplification, and purification. Genomic DNA was extracted from *Cystobacter velatus* Cbv34 grown in CTT medium.

The 213 bp DNA fragment encoding CysA was amplified using the forward primer 5'TATCATATGAGCATGAACGGGGACG and the reverse primer 5'TATAGATCTTCA-GCAGTTGCTGCGCG. The amplified DNA fragments were digested with NdeI and BglII and cloned into the second MCS of pETduet-1 in native form.

The 1899 bp DNA fragment encoding CysG1 was amplified using the forward primer 5'TATAGATCTGGAAAACCTGTATTTTCAGGGCATGGCCACCAAATTGTCTGACTT C containing the ENLYFQG TEV protease recognition site and the reverse primer 5' TATA-AGCTTTCAGTGGCCGCGGTCCGT, while the 4065 bp fragment encoding CysG2 was amplified using the forward primer 5'GGATCCGGAAAACCTGTATTTTCAGGGCGCGCT-TCCGCTGTGCGC containing the ENLYFQG TEV protease recognition site and the reverse primer 5'TATAAGCTTTCACGAAGCTCGCGTCCTC. The amplified DNA fragments were

digested respectively with BglIII / HindIII and or BamHI / HindIII, and cloned into the first MCS of pETduet-1 digested with BamHI and HindIII to yield N-terminal His6 tag and TEV protease site fusion constructs. The resulting constructs were sequenced to verify that no mutation had been introduced during PCR.

The 732 bp DNA fragment encoding CysE was amplified using the forward primer 5'TACATATGATGATTGCGTTCAACCCGCA and the reverse primer 5'TACTCGAGCTAC-AACACACACTCAAGCA. The amplified DNA fragments were digested with NdeI and XhoI, and cloned into pET-28b with an N-terminal His6 tag. The resulting construct was sequenced to verify that no mutation had been introduced during PCR.

Recombinant protein expression was carried out in *E.coli* BL21 (DE3) grown in LB medium supplemented with 50 µg/ml of kanamycin. The culture was inoculated 1/10 from a fully grown overnight culture and was initially grown at 37 °C until it reached an OD of 0.6 to 0.9, at which point the temperature was decreased to 16 °C. Induction was initiated after 30 minutes at 16°C with 0,1 mM isopropyl-β,D-thiogalactopyranoside (IPTG). The cells were harvested 16h after induction by centrifugation at 8000 g for 10 minutes at 4°C and resuspended in lysis buffer (25 mM TRIS pH7,5; 150 mM NaCl; 20 mM imidazole). The cells were lysed by passage at 1500 bar at 4°C through a M-110P Microfluidizer (Microfluidics) after which the crude extract was centrifuged at 45000 g for 15 minutes at 4°C, the supernatant was treated as soluble protein fraction.

Affinity chromatography was performed on an Äkta Avant system while size exclusion chromatography was performed on an Äkta Pure system. The soluble fraction of the lysed cells was applied to a 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10CV lysis buffer prior to elution with elution buffer (25 mM TRIS pH7,5; 150mM NaCl; 250 mM imidazole).

For CysE the pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200 pg SEC column equilibrated in running buffer (25 mM TRIS pH7,5; 150 mM NaCl; 2 mM DTT). The purity of the pooled fractions was checked via SDS-PAGE, and the protein was concentrated and frozen at -80°C after addition of 25% glycerol.

For CysG the pooled fractions were applied to a HiPrep 26/10 desalting column equilibrated in running buffer (25 mM TRIS pH7,5; 150 mM NaCl; 2mM DTT). In the case of CysG1 and CysG2 the resulting fractions were pooled (≈ 30 ml) and incubated overnight at 4°C with TEV protease (1 mg/20 mg protein) and CysE (1 mM). After ≈ 16 h incubation 20 mM imidazole was added to the solution prior to loading on a 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10 CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole. The pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200 pg SEC column equilibrated in running buffer (25 mM TRIS pH7,5; 150 mM NaCl; 2mM DTT). The pooled fractions were controlled via SDS-PAGE, concentrated and frozen at -80°C after addition of 25% glycerol.

## 2.5 *In trans* activity assays on CysG

In a total volume of 25  $\mu$ l, 5  $\mu$ M CysG1 or CysG2 was incubated with 1mM 3OH-pABA, 1mM ATP and 1 mM SAM in 25mM TRIS pH7,5; 150 mM NaCl; 10mM MgCl<sub>2</sub>, finally 500 nM CysF was added and the reaction was incubated for 1h at room temperature. The solution was analysed by protein LC-MS.

CysG1 and CysG2 were analyzed on a Dionex Ultimate 3000 UPLC system (Thermo Scientific) coupled with an maXis4G Q-TOF MS (Bruker) using an ESI in positive mode. The samples were run on a Aeris Widepore XB-C8, 150 x 2.1 mm, 3.6  $\mu$ m dp column (Phenomenex, USA). LC conditions: A-Water + 0.1% FA; B-Acetonitrile + 0.1 % FA at a flow rate of 300  $\mu$ l/min and 45 °C. 0 min: 98 % A / 2 % B, 0.5 min: 98 % A / 2 % B, 10.5 min: 25 % A / 75 % B, 13.5 min: 25 % A / 75 % B, 14 min: 98 % A / 2 % B. The LC flow was split to 75  $\mu$ l/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany) using the standard Bruker ESI source. In the source region, the temperature was set to 180 °C, the capillary voltage was 4000 V, the dry-gas flow was 6.0 l/min and the nebulizer was set to 1.1 bar. Mass spectra were acquired in positive ionization mode ranging from 150 – 2500 m/z at 2.5 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm (Copyright 1991-2004 Spectrum Square Associates, Inc.).

## 2.6 Crystallization and Structure determination

Crystals of CysF apo, CysF-SAM and CysF-SAH-pABA complexes were obtained at 18 °C in 0.2 - 0.4 M Magnesium sulfate, 22-28 % PEG3350 and 0.1 M Tris-Cl, pH 7.5. The crystals were cryoprotected in mother liquor supplemented with 35 % glycerol and flash cooled in liquid nitrogen. Data was collected at ESRF (Beamline: ID29 and ID23-1) and the structure was solved using PHASER (19) molecular replacement with putative O-methyltransferase from *Nostoc punctiforme* (PDBID: 2R3S) as the search model. The models were manually rebuilt in COOT (20) and refined using PHENIX (21) and Refmac5.(22) The structures were validated using MolProbity (23), and all images presented were created using PyMOL.(24) Interaction diagrams were created using Ligplot.(25)

## 3 Results and Discussions

After *in silico* analysis of the cluster a <sup>13</sup>C methionine feeding was carried out as preliminary experiment to confirm the origin of the isopropoxyl and methyl groups in Cystobactamides. A shift of +7 m/z was observed which confirmed that both isopropyls and the methyl were originating in SAM. (Figure 3.4) The 37.7 kDa CysF was subsequently overexpressed for structural and biochemical characterization as an N-terminal 6His tag fusion protein, the tag was cleaved off via incubation with TEV protease yielding CysF in native form

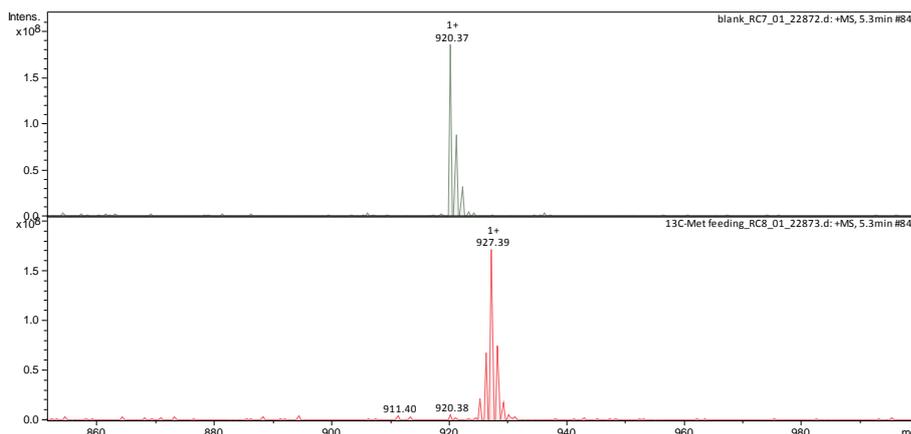


Figure 3.4: : EIC of Cystobactamide 919-1 extracted after  $^{13}\text{C}$ -methionine feeding in red (bottom) against unlabeled methionine in green (top)

### 3.1 *In vitro* activity assays

Activity assays were performed on CoA derivatives, *in trans* on loaded NRPS nodules CysG1 and CysG2 and on free 3-OH-pABA. Initially the activity assays were done on *in situ* generated 3OH-pABA CoA thioesters (Figure 3.5) and the methylation of the hydroxyl group analyzed by LC-MS was confirmed with a synthetic analogue. The kinetics determined for the CoA ligase CysL in the previous chapter showed that the coadenylation process is extremely fast which allows us to use this enzyme to generate the 3-OH-pABA-CoA substrate *in situ* for the methylation. We assume that the methyl transfer would be the rate limiting step in this setup. The presence of trace amounts of 3OH-pABA CoA observed after 30 min incubation implies that this hypothesis was correct and that CysF is indeed slower than CysL. (Figure 3.5)

*In trans* activity on loaded NRPS or PKS modules was never characterized *in vitro* mainly because of the challenges the overexpression of such large multidomain proteins represent. The meta-methoxyl decoration is observed on both pABA units integrated by modules five and six which are encoded on CysG. We decided to overexpress modules CysG1 and CysG2 separately to be able test the activity of tailoring enzymes on both modules independently, thus we started by modelling the linker region between the T5 and C6 on the structure of the GrsA T-E didomain (PDB: 5ISX) with I-TASSER (26) online tool to determine the C-terminus of CysG1 and the N-terminus of CysG2. MbtH type adenylation domain activators have previously been shown to be necessary for the overexpression and activity of NRPS modules. (17) Since initial overexpression attempts of CysG modules yielded almost no soluble protein a coexpression strategy involving the CysA, the MbtH protein found in the Cystobactamide cluster was adopted. The 1,9 kbp CysG1 and 4 kbp CysG2 were subcloned in a dual vector for coexpression with the small 200bp CysA gene. This strategy proved successful and allowed the purification of  $\approx 20$  mg/l CysG1 and  $\approx 8$  mg/l CysG2. A further unusual procedure had to be followed in order to generate substrate for *in trans* activity assays since the NRPS modules were partially loaded *in vivo* with pABA

by *E.coli* during the overexpression. Incubation overnight at 4°C of the Ni-NTA purified modules with CysE the type II thioesterase present in the cluster, in parallel with the TEV digest, allowed to both cleave the Histag and remove the loaded pABA to yield pure unloaded NRPS modules in *holo* form after a second affinity chromatography and a final size exclusion chromatography.

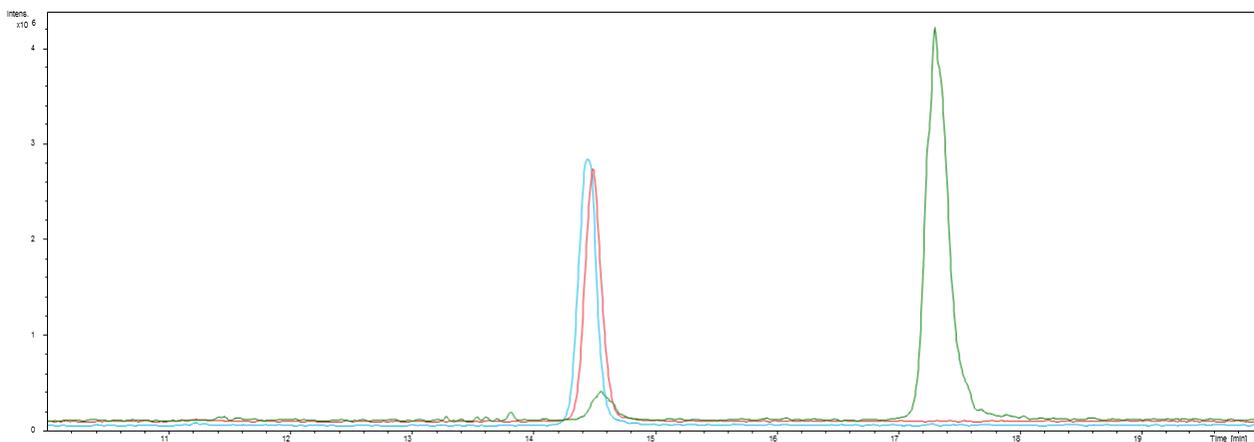


Figure 3.5: LC-MS analysis of the CysF catalyzed methylation of 3OH-pABA CoA thioester, EIC (903.15; 917.17 m/z). Blue trace shows the 3OH-pABA CoA formed upon incubation with CysL alone. The red trace is the same EIC upon addition of CysF alone. The green trace shows the complete turnover from 3OH-pABA CoA to 3OMe-pABA CoA upon

CysG modules were analyzed through protein MS after incubation with various pABA derivatives to investigate their loading specificity and 3-hydroxy-pABA loaded CysG1 and CysG2 was incubated with CysF. The methylation of the hydroxyl was observed in both cases indicating that CysF is indeed able to act *in trans* to process loaded substrate. Despite that, the promiscuity showed by CysG in the adenylation and loading is quite important and it is able to load both 3-OH-pABA and 3-OMe-pABA. (Figure 3.6) Since kinetic analysis on an *in trans* process are difficult to determine and compare to the reaction on CoA derivatives, it is still impossible to determine whether the loading happens prior to this initial methylation or after it.

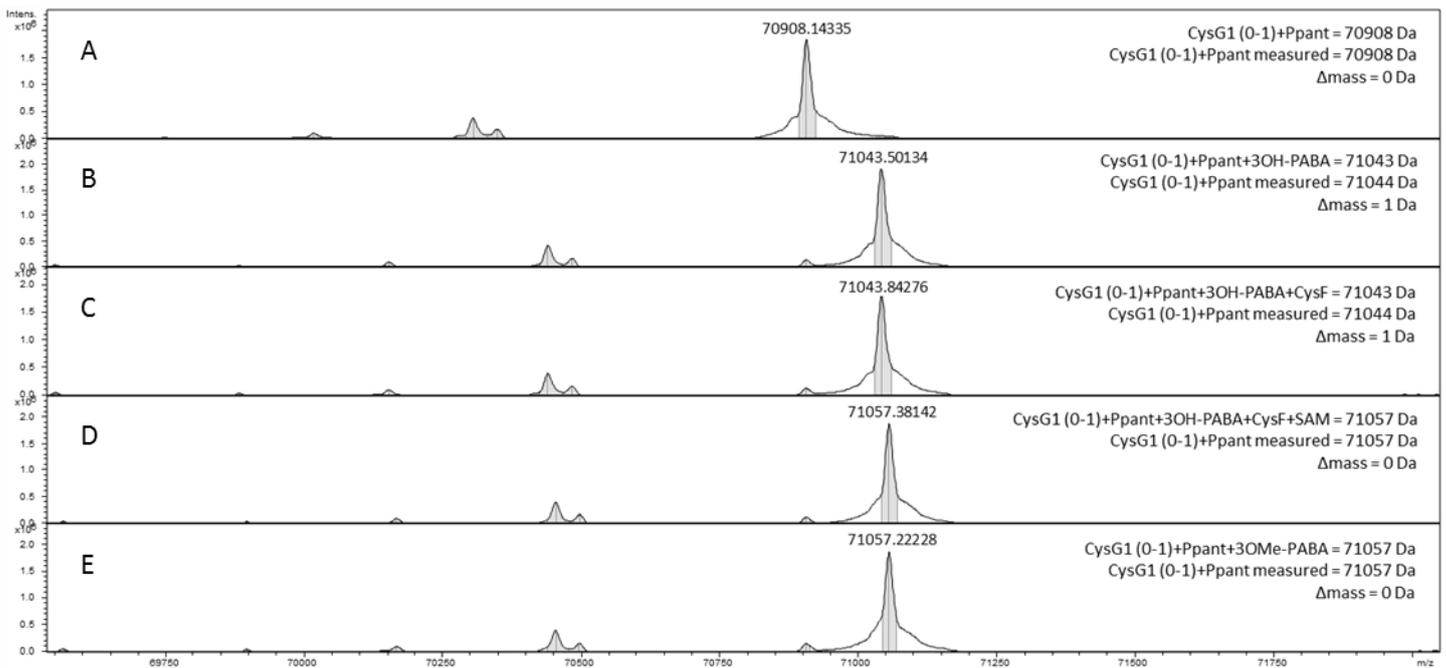


Figure 3.6: Protein MS analysis of CysG1 (A) incubated with 3-OH-pABA (B) and CysF with (D) or without (C) SAM compared to CysG1 incubated with 3-OMe-pABA (E)

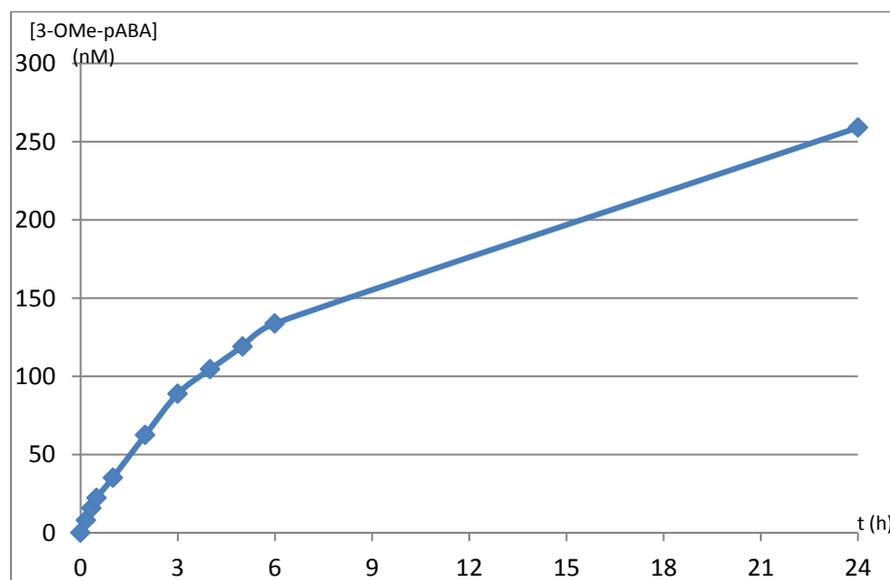


Figure 3.7: Plot showing the concentration (nM) of 3-OMe-pABA product formed in a time course reaction over 24h with 1  $\mu$ M CysF

To confirm the importance of the thioester bond for the reactivity of the substrate, the activity of CysF on free 3-OH-pABA was tested in a time course. It revealed that a residual activity can be observed but less than 0,1% of the substrate is processed after 24h. Since the reaction proceeds to a full turnover after less than 30 minutes in the case where the substrate

is coadenylated, one can safely assume that the thioester activation of the ring is indeed necessary for an efficient catalysis. (Figure 3.7)

### 3.2 Structural Analysis

The structures of apo-CysF, of the binary complex CysF/SAM and of the ternary complexes CysF/3OH-pABA/SAH were solved through molecular replacement using the structure of a putative O-methyltransferase from *Nostoc punctiforme* (PDBID: 2R3S) and were refined to a resolution of respectively 2,40 Å, 2,20 Å and 2,27Å, the apoprotein crystallizes in P 1 2 1 space group while the complex structures on the other hand crystalize in C 1 2 1 space group. In the three cases CysF features a dimeric structure with tightly intertwined N-terminal domains which are folded in a helix-turn-helix conformation ( $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 3$ ,  $\beta 1$ ,  $\beta 2$ ,  $\alpha 4$ ). The dimer interface was calculated using PISA to be 3623.1 Å<sup>2</sup> and focused on hydrophobic interactions between the bottom of a groove formed by the  $\alpha 4$  helix of on monomer and a corresponding ridge formed by  $\alpha 1$  helix of the other monomer. In addition a number of hydrogen bonds are formed between the negatively charged  $\alpha 9$  helix and the positively charged  $\alpha 3$  helix. (Figure 3.8) The long  $\alpha 4$  helix which forms the interface with the C-terminal domain is not completely ordered in the apo structure and is the only part of the N-terminal domain containing amino acids relevant for substrate binding. The C-terminal domain is a classical SAM binding Rossmann fold featuring a 7 stranded  $\beta$  sheet ( $\beta 3$  to  $\beta 9$ ) with  $\beta 9$  antiparallel to the others, surrounded by 6  $\alpha$  helixes and contains the signature GxGxxG SAM binding motif.

The structural analysis confirms *in silico* data suggesting that CysF is a type I methyltransferase, and a DALI search revealed that a number of type I MT share a similar N-terminal domain. They originate mostly from bacteria (e.g the 3-hydroxykynurenine C-methyltransferase (SibL), the Mitomycin 7-O-Methyltransferase (MmcR) or the Daunorubicine 4-O-MT (DnrK)) but can also be found in human (Acetyl serotonin MT (ASMT)) and from plants (Monolignol 4-O-Methyltransferase). Even if they share a very low sequence identity of 22% all these enzymes present the same N-terminal dimerization domain and are active on a wide range of aromatic substrates as O or C methyl transferases. The only similar thioester requiring methyltransferase highlighted in the DALI search is CalO6 (Z = 22.1; rmsd = 4.3 Å) which was already suggested to cluster with CysF on sequence level, additionally secondary structure prediction on PokMT1 suggest that it also shares the same N-terminal architecture as CalO6 and CysF.

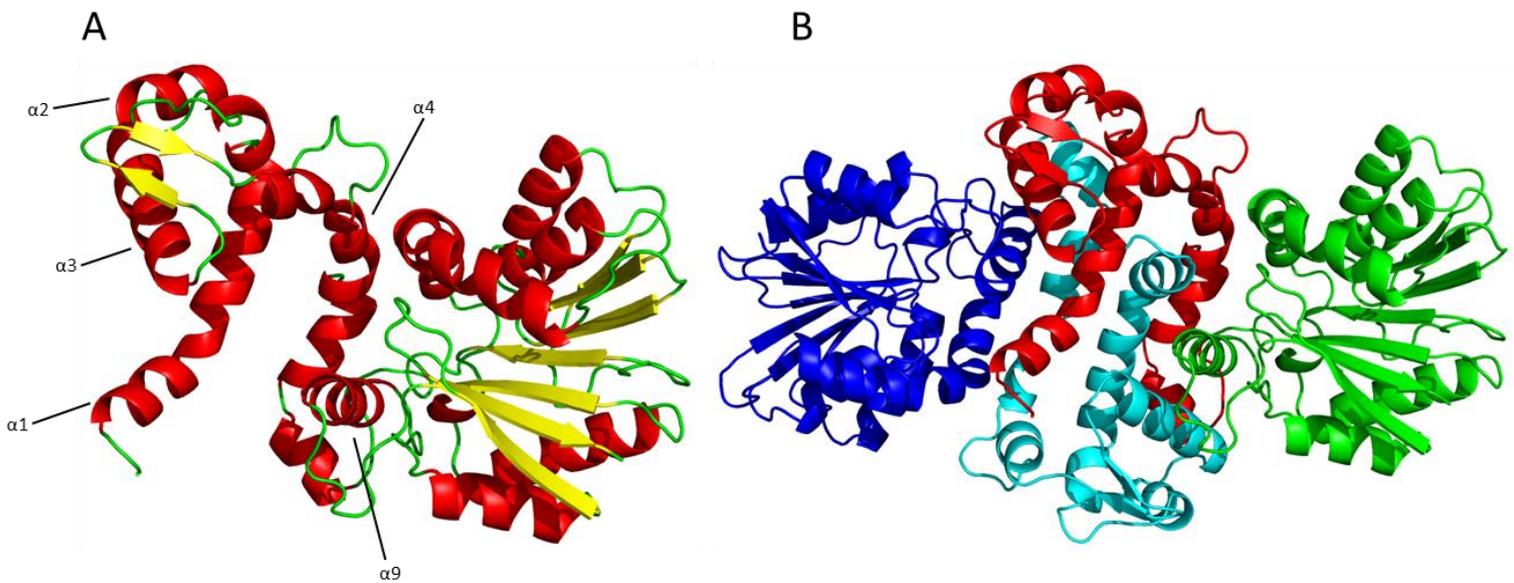


Figure 3.8: A: Cartoon representation of the apo CysF monomer, helices involved in the dimer interface are numbered B: Cartoon representation of the apo CysF dimer the N terminal dimerization domains are shown in red and cyan while the catalytic C terminal domains are shown in green and blue

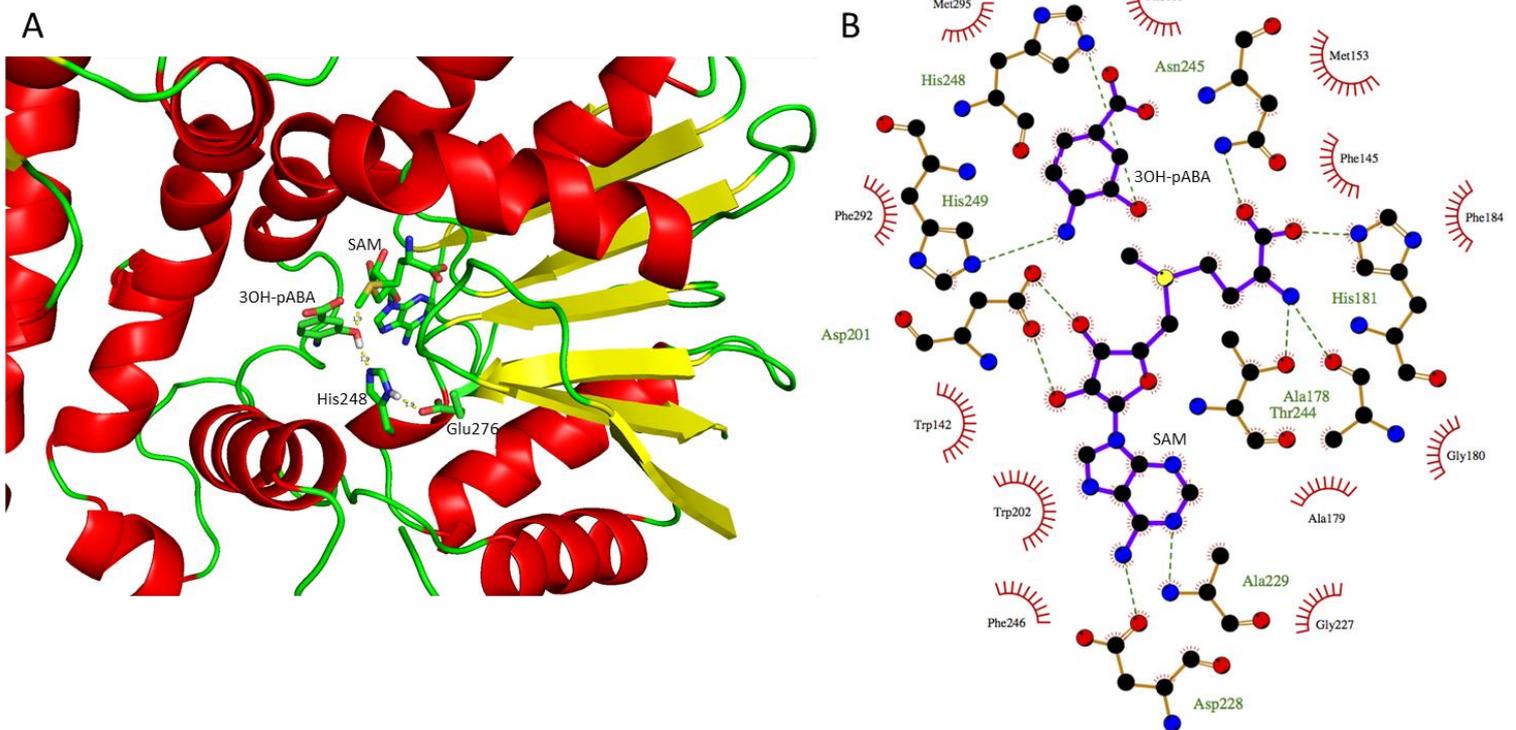


Figure 3.9: A: Composite image of the active site of CysF with bound SAM in binary complex superimposed with 3OH-pABA from the substrate bound complex showing the catalytic Histidine 248 and Helper Glutamate 276; B: Ligplot showing the interactions of 3OH-pABA and Sam in the active site of CysF

In contrast to similar structures there is no movement from the C-term domain towards the N-term domain, the apo form presents a relatively relaxed SAM binding pocket, upon binding of the cofactor the Phe145 flips at 180° and allows the ordering of the adjacent

$\alpha 4$  helix mainly through repositioning of Phe105 which closes the binding site and creates the 3OH-pABA binding pocket. The catalysis is carried out by His248 helped by Glu276, this catalytic duo works as general base to generate the phenolate anion in position to attack the methylsulfonium cation from SAM through nucleophilic substitution. (Figure 3.10) In the active site SAM is stabilized mostly by hydrophobic interactions (Trp142, Ala179 and Phe246),  $\pi$  stacking with Trp202 and hydrogen bonds between Asp228 and the adenosine, Asp201 and the ribose, Thr244 and the carboxylate and finally His181 and the amine. The pABA binding site is constituted by Phe105, Phe145, Phe275, Phe292 and Met295 and a hydrogen bond is formed between the amine and His249. (Figure 3.9)

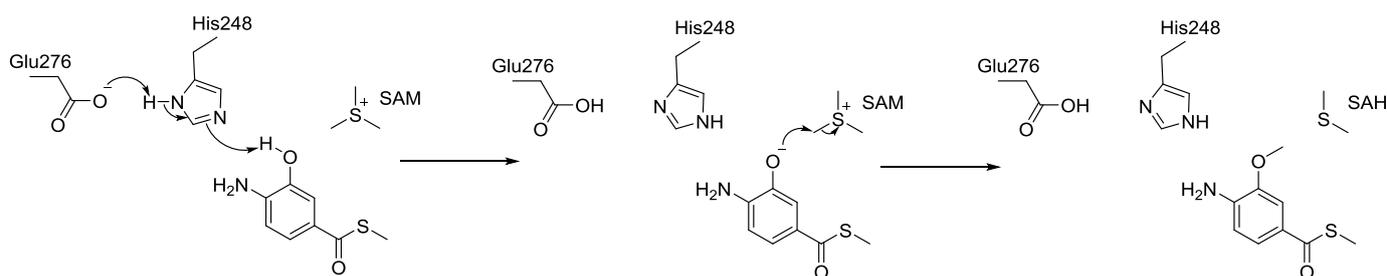


Figure 3.10: Proposed CysF methylation mechanism by His248 and Glu276

Despite the fact that only minimal activity is observed on free 3OH-pABA it is able to bind in the active site, additionally a channel is formed at the interface between both monomers which is likely to accommodate the phosphopantetheine arm and further stabilize the aromatic ring in the right position for an efficient catalysis. However since almost no turnover is observed on substrate featuring a free acid one can assume that the thioester linkage is necessary to activate the ring in order to increase the acidity of the phenol and allow its attack by the His248 to promote the formation of the reactive phenolate.

## 4 Conclusion

CysF is one of the tailoring enzymes of the Cystobactamide biosynthesis; it is responsible for the methylation of the hydroxyl in meta position on the two last hydroxy-pABA of the Cystobactamide backbone. *In vitro* activity assays revealed that the substrate had to be activated as a CoA or Ppant thioester for an efficient catalysis which allows us to classify this enzyme in the growing family of thioester activated aromatic methyltransferases. *In trans* activity had been proposed for other members of the family and has been observed for the first time for CysF through protein MS. Structural characterization of the apoprotein and of the complexes with substrate and cofactor confirmed CysF as a type I natural product methyltransferase and revealed the presence of an N-terminal dimerization domain.

## Chapter 3.2: CysS - Biosynthesis of Branched Alkoxy Groups: Iterative Methyl Group Alkylation by a Cobalamin-Dependent Radical SAM Enzyme

### 1 Introduction

Radical SAM enzymes have gathered tremendous attention in the last ten years due to their uncommon biochemistry involving radical mechanisms that allow them to perform methylations as well as other reactions on unreactive  $sp^3$  carbons.(27,28) *In silico* analysis of large databases show that radical SAM enzymes seem to be much more widespread than initially thought, yet only a limited number of them were investigated *in vitro*. Indeed the reconstitution of their activity *in vitro* is still a challenge due to the inherent oxygen sensibility of the 4Fe/4S cluster underlying their biochemistry, which implies that the overexpression and purification of the enzymes has to be performed under anaerobiosis.(8) While radical SAM can perform various reactions they are mostly known as methyl transferases, in this case the enzyme usually features a secondary B12 binding domain where the B12 cofactor acts as a methyl donor while SAM acts as a radical donor.(6)

Cystobactamide presents a few very distinctive chemical features making it a unique derivative, amongst which the isopropoxyl groups carried by both last para-amino benzoic acid moieties. As common as the use of isopropyl groups might be in organic chemistry, they are a very rare decoration in natural products which are found only in a few metabolites, essentially in terpenes and sugar derivatives isolated from plant extracts.(29,30) Along with Cystobactamides one of the only microbial products exhibiting an isopropoxyl decoration is the enediyne Kedaridine.(31) In this case the isopropoxyl is carried by a naphthoic acid bicycle alongside multiple phenols resulting in a comparable set up as in Cystobactamides. Even if the biosynthesis pathway leading to the isopropoxyl was not confirmed experimentally in Kedaridine, the *in silico* analysis of the cluster reveals the presence of a classical SAM methyltransferase KedN1 and a radical SAM KedN5 which were proposed to be responsible for the sequential methylation of the phenol in a similar fashion as it was described in Chondrochlorins.(32) This myxobacterial polyketide features an ethoxyl group which was proven to be biosynthesized by the sequential activity of a classical SAM methyltransferase PKS domain in CdnE and a radical SAM CdnI.

The sequential methylation of the same position to yield alkyl groups has been postulated to happen in the biosynthesis of a few metabolites but has in fact only been described for one other natural product, the carbapenem antibiotic Thienamycin (33) where the radical SAM ThnK methylates the  $\beta$ -lactam ring on the C-6 and then extends this side chain with a second methylation leading to an ethyl decoration. An iterative methylation process appears to happen twice in the very large RiPP Polytheonamide (34) where both the N terminal threonine is heavily modified into a 5,5-dimethyl-2-oxohexanoic acid and a methionine is dimethylated on the  $\beta$  position, these reactions were proposed to be carried out

## CysS - Biosynthesis of Branched Alkoxy Groups: Iterative Methyl Group Alkylation by a Cobalamin-Dependent Radical SAM Enzyme

by the radical SAM PoyB and PoyC. A last account can be found in the literature in the biosynthesis of SW-163F (35), the methyl transferase Swe8 and radical SAM Swe9 are proposed to be responsible for the biosynthesis of a tertbutyl thioacetal.

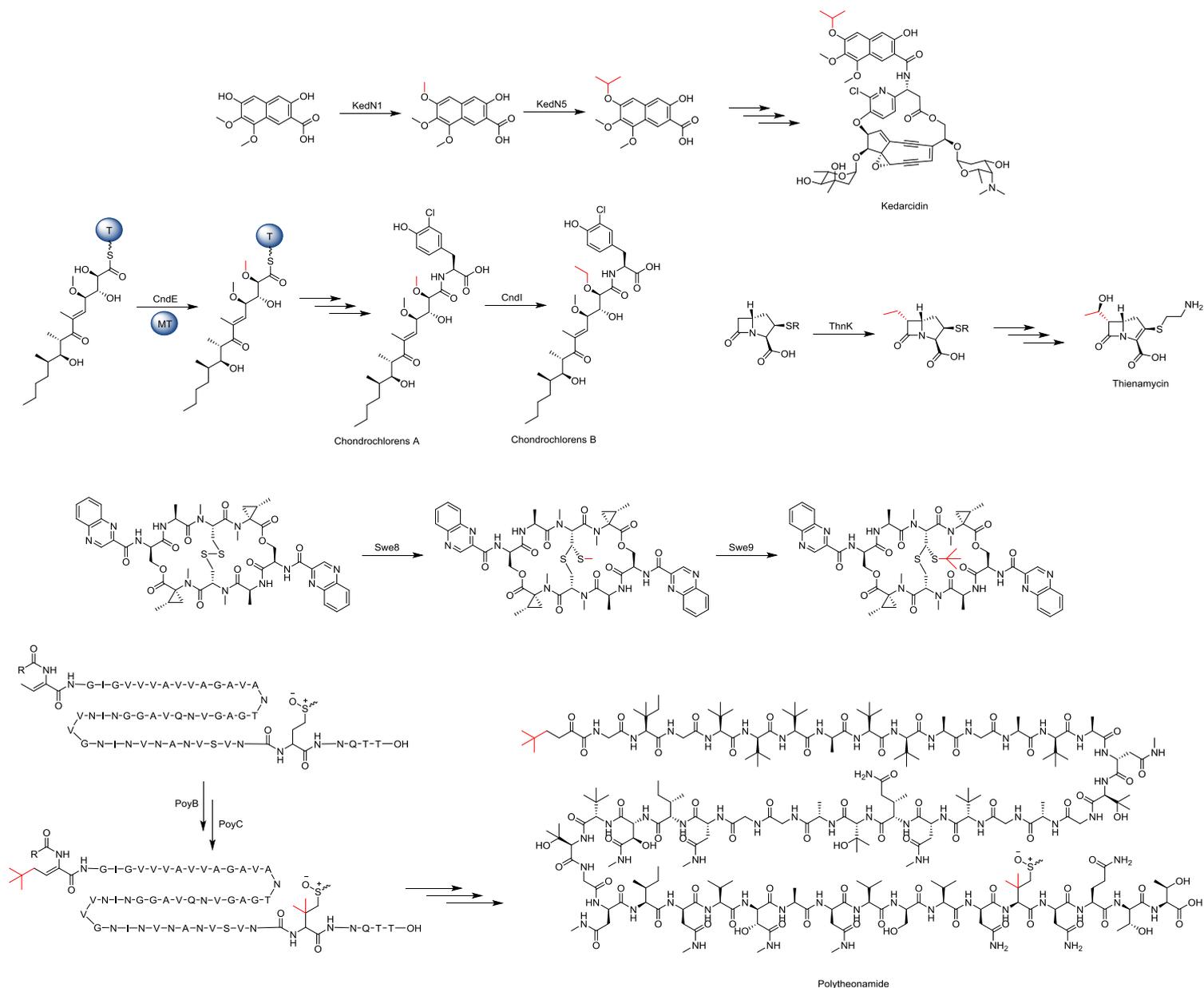


Figure 3.11: Examples of biosynthesis of branched alkyl groups by iterative radical SAM methylation

## 2 Material and Methods

### 2.1 Cloning, overexpression and purification

Standard protocols were used for DNA cloning, transformation, amplification, and purification. Genomic DNA was extracted from *Cystobacter velatus* Cbv34 grown in CTT medium, the 1929 bp DNA fragment encoding CysS was amplified using the forward primer

5' ATCATATGAAACGGTTCTTCAAGCTCCAGTTGC and the reverse primer 5' ATCTCGAGTCAGCGTCCGTGGCTGGATTCG. The amplified DNA fragments were digested with NdeI and XhoI, and cloned into pET-28b with an N-terminal His6 tag. The resulting construct was sequenced to verify that no mutation had been introduced during PCR.

Recombinant protein expression was carried out in *E.coli* BL21 (DE3) grown in LB medium supplemented with 50 µg/ml of kanamycin. The culture was inoculated 1/10 from a fully grown overnight culture and was initially grown at 37°C until it reached an OD of 0.6 to 0.9, at which point the temperature was decreased to 16°C. Induction was initiated after 30 minutes at 16°C with 0,1 mM isopropyl-β,D-thiogalactopyranoside (IPTG). The cells were harvested 16 h after induction by centrifugation at 8000 g for 10 minutes at 4°C and resuspended in lysis buffer (25 mM TRIS pH 7,5; 150 mM NaCl; 20 mM imidazole). The cells were lysed by passage at 1500 bar at 4°C through a M-110P Microfluidizer (Microfluidics) after which the crude extract was centrifuged at 45000 g for 15 minutes at 4°C, the supernatant was treated as soluble protein fraction.

Affinity chromatography was performed on an Äkta Avant system while size exclusion chromatography was performed on an Äkta Pure system. The soluble fraction of the lysed cells was applied to 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole. The pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200 pg SEC column equilibrated in running buffer (25 mM TRIS pH 7,5; 150 mM NaCl; 2 mM DTT). The pooled fractions were controlled via SDS-PAGE, concentrated and frozen at -80°C after addition of 25% glycerol.

To regenerate the iron sulfur clusters CysS was incubated in a glovebox with a 30 fold molar excess of NaS and (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub> and desalted on a PD-10 column to remove excess iron and sulfide. The iron titration was done by colorimetric titration using bathophenanthroline as described previously. (36)

## 2.2 In vitro activity assays

All CysS enzymatic reactions were carried out in an anaerobic chamber containing 95% nitrogen and 5% hydrogen. A typical enzymatic reaction was performed in 100 mM phosphate buffer, pH 7.5 containing 55 µM of CysS with 0.75 mM MeCbl, 2.5 mM SAM, 0.9 mM substrate with different reducing systems, flavodoxin, flavodoxin reductase system (25 µM FldA, FPR; 50 µM FAD, FMN; 3 mM NADPH); 1 mM methyl viologen, 4 mM NADPH; 10 mM dithionite, 10 mM dithiothreitol (DTT); 1 mM methyl viologen, 4 mM NADPH, 50 mM DTT; 1 mM methyl viologen, 4 mM NADPH, 50 mM DTT. The reaction was incubated at room temperature for 12-15 hours. For LC-MS analysis the enzyme was removed by ultrafiltration using a 10kDa cut-off filter (VWR).

## 2.3 B12 feeding

25 ml M-medium was inoculated with 1 mL of a high density overnight culture of Cbv34. 0,5 ml of XAD-7 was added on day 1 and 25 µg or 250 µg of B12 was fed over 4 days. On day 5 the cultures were centrifuged and the pellet and XAD were extracted with 2x10ml methanol. The solvent was removed through rotary evaporation and the extract was redissolved in 2 ml methanol.

LC-ESI-MS was performed using a Thermo Scientific Dionex Ultimate 3000 UHPLC system coupled with a Bruker Amazon speed iontrap mass spectrometer using an ESI in positive mode. LC was performed on a Waters Acquity UPLC BEH C18 50 mm 1.7 µm column. LC conditions: A-Water + 0.1 % formic acid; B-Acetonitrile + 0.1 % formic acid; 0 min: 95% A / 5% B, 0.5 min: 95 % A / 5 % B, 9.5 min: 5 % A / 95 % B, 10.5 min: 5% A / 95 % B, 10.8 min: 95 % A / 5 % B, 11.8 min: 95 % A / 5 % B.

## 3 Results and Discussions

Minor Cystobactamide derivatives that harbor methyl or ethyl groups in place of the isopropoxyls on the two last para-amino benzoates have been observed during fermentation, this led to the hypothesis that the biosynthesis of the isopropoxyl could be performed via iterative methylation of the phenol. A <sup>13</sup>C methionine feeding was performed as preliminary experiment (Cf. chapter 3.1; Figure 3.4) and showed a shift of +7 Da which confirmed the origin of both the isopropoxyl groups as well as the methoxyl carried by the isoasparagine as derived from SAM. *In silico* analysis of the cluster revealed the presence of a class B radical SAM methyltransferase, CysS, which could be a good candidate to perform this type of activity.

### 3.1 *In silico* analysis and protein expression

CysS presents the telltale C194xxxC198xxC201 binding motif for the 4Fe-4S cluster necessary for generating the deoxyadenosine radical in addition to an N terminal B12 binding motif harboring the cobalamine acting as a methyl donor. Radical SAM are known to sometimes harbor additional domains (37) that are believed to play a role in substrate binding and CysS indeed harbors a large C terminal domain of about 200 amino acids which is not structurally related to any characterized radical SAM and does not feature any known conserved motifs. This domain can be found on a few uncharacterized hypothetical proteins in the NCBI database and might be involved in quaternary interactions with the NRPS assembly line but only structural characterization would allow in depth analysis.

To investigate the biochemistry of this reaction, CysS was cloned in a pET28b vector with an N terminal His6 tag, overexpressed in *E.coli* BL21 and purified as a brownish protein which indicates that part of the iron sulfur cluster content is conserved during the original purification. Originally the protein was regenerated by incubation with Fe<sup>II</sup> Iron and S<sup>2-</sup>

sulfide ions under anaerobiosis before desalting on a PD-10 column in the glovebox. This setup allowed confirmation of the presence of the 4Fe-4S cluster but for *in vitro* activity assays, our collaborators at Texas A&M eventually coexpressed a Suf operon with CysS in order to boost the assembly of 4Fe-4S clusters in *E.coli* (38) and resolved to fully anaerobic purification.

### 3.2 Iron sulfur cluster regeneration

To generate the deoxy adenosine radical (dAdo•), radical SAM enzymes require a 4Fe-4S cluster which is prone to oxidation by molecular oxygen hence the brown color featured by CysS after initial purification tends to fade if the protein is left uncovered. Iron sulfur clusters can be regenerated under anaerobic incubation of the protein with Fe<sup>II</sup> ferrous iron and S<sup>2-</sup> sulfide ions, their presence in the protein was assayed both spectrophotometrically and through iron titration.

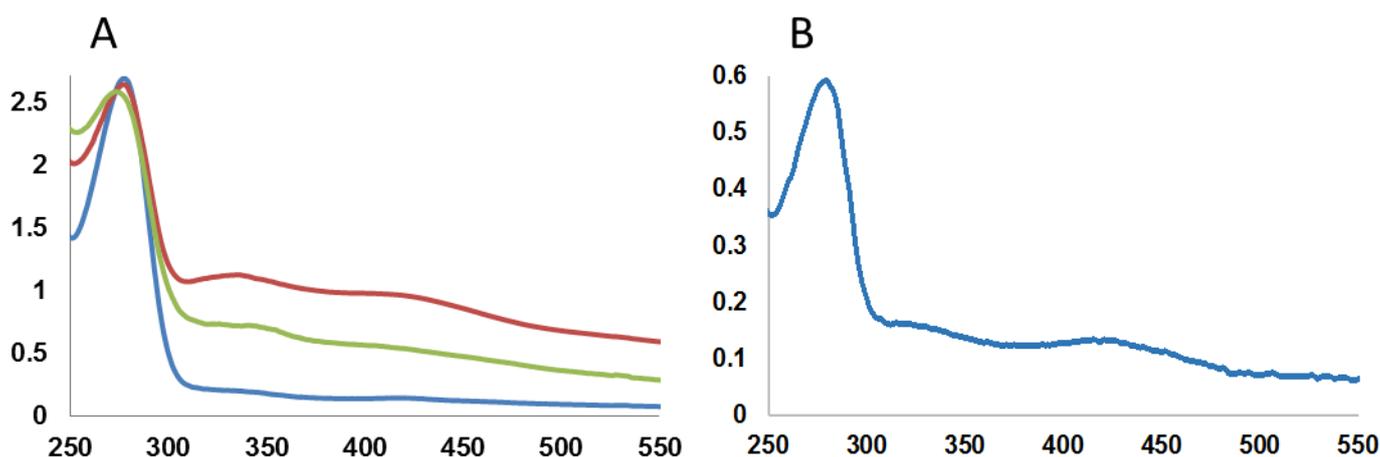


Figure 3.12: A- Spectrophotometric analysis of CysS throughout the regeneration process; CysS before regeneration is in blue, in presence of Fe<sup>2+</sup> and S<sup>2-</sup> is in red and after purification is in green. B- CysS coexpressed with a *suf* operon and purified anaerobically.

Spectrophotometrical analysis of CysS revealed that the shoulders at 328 nm and 420 nm characteristic of 4Fe-4S clusters are barely to be seen before regeneration. After incubation with Fe<sup>2+</sup> and S<sup>2-</sup> and desalting both  $\lambda_{max}$  are remaining at high intensities indicating a successful binding of the clusters to the protein. (Figure 3.12) This result is consistent with the Spectrophotometrical analysis of the protein coexpressed with a Suf operon and purified under anaerobiosis which features these absorption maxima even without regeneration.

In addition to the spectrophotometrical analysis the exact iron content of the protein after regeneration was titrated using a bathophenanthroline colorimetric assay to determine the level of cluster load. The iron concentration was measured at  $2,9 \pm 0,12$  mol Fe per mol of CysS which is consistent with a partial iron-sulfur cluster load, sufficient for *in vitro*

catalysis. The iron concentration in the anaerobic purified protein was also measured and determined at 2,5 irons per monomer of CysS in addition the sulfide concentration was measured at 2,8 mol/mol. These results indicate that both purification methods are equivalent and yield comparably active proteins.

### 3.3 *In vitro* activity reconstitution

We performed preliminary experiments on free 3-methoxy-pABA and related derivatives with the regenerated protein which did not yield any product indicating that the activity was taking place either *in trans* or on CoA bound substrate, in addition *in vitro* reconstitution of radical SAM enzymes usually requires an electron donor chemical to maintain the iron-sulfur cluster as well as the cobalt center in a reduced state able to perform catalysis. Our collaborators managed to perform the *in vitro* biochemical characterization of CysS and proved that its activity is performed *in trans* on PCP loaded substrate since either N-acetylcysteamine or pantetheine thioesters were the only substrates allowing catalytic turnover, the pantetheinyl thioester of 3-methoxy-pABA being processed at a 47 times higher rate than SNAC ester. Interestingly, contrary to CysF, CoA bound substrate was not processed by CysS and only a low thioester hydrolysis activity was observed in this case. In addition various iron-sulfur cluster reducing systems were investigated revealing that addition of flavodoxine / flavodoxine reductase / NADPH or methyl viologen / NADPH are necessary for catalytic turnover but methyl viologen / dithionite did not improve the very low base rate of the reaction indicating that a small portion of the iron-sulfur clusters are in a catalytic reduced state after purification.

*In vitro* activity reconstitution confirmed the iterative methylation activity that was hypothesized after *in silico* analysis, the enzyme yielding both the ethoxyl and the isopropoxyl derivatives when incubated with SNAC 3-methoxyl-pABA. Interestingly the higher reactivity provided by the pentetheine derivatives allowed the reaction to perform a third catalytic cycle yielding either sec-butoxylated or tert-butoxylated pABA derivatives in lower yields, which are not observed *in vivo*. Since the reaction appears to be catalyzed *in trans* on CysG bound substrate it is likely that the specificity of the condensation domains of the fourth and fifth modules is directed towards isopropoxylated intermediates and thus this third catalytic cycle from CysS cannot be performed before condensation of the peptide bonds. The identity of these derivatives was confirmed by retention time comparison with synthetic samples; in addition the reaction was performed using CD<sub>3</sub>-SAM to confirm the origin of the methyls for the ethoxyl and the isopropyl derivatives. A quantitative analysis of the reaction mixture revealed that equal amounts of SAH and dAdo are produced which is consistent with SAM being both the methyl donor and the radical donor

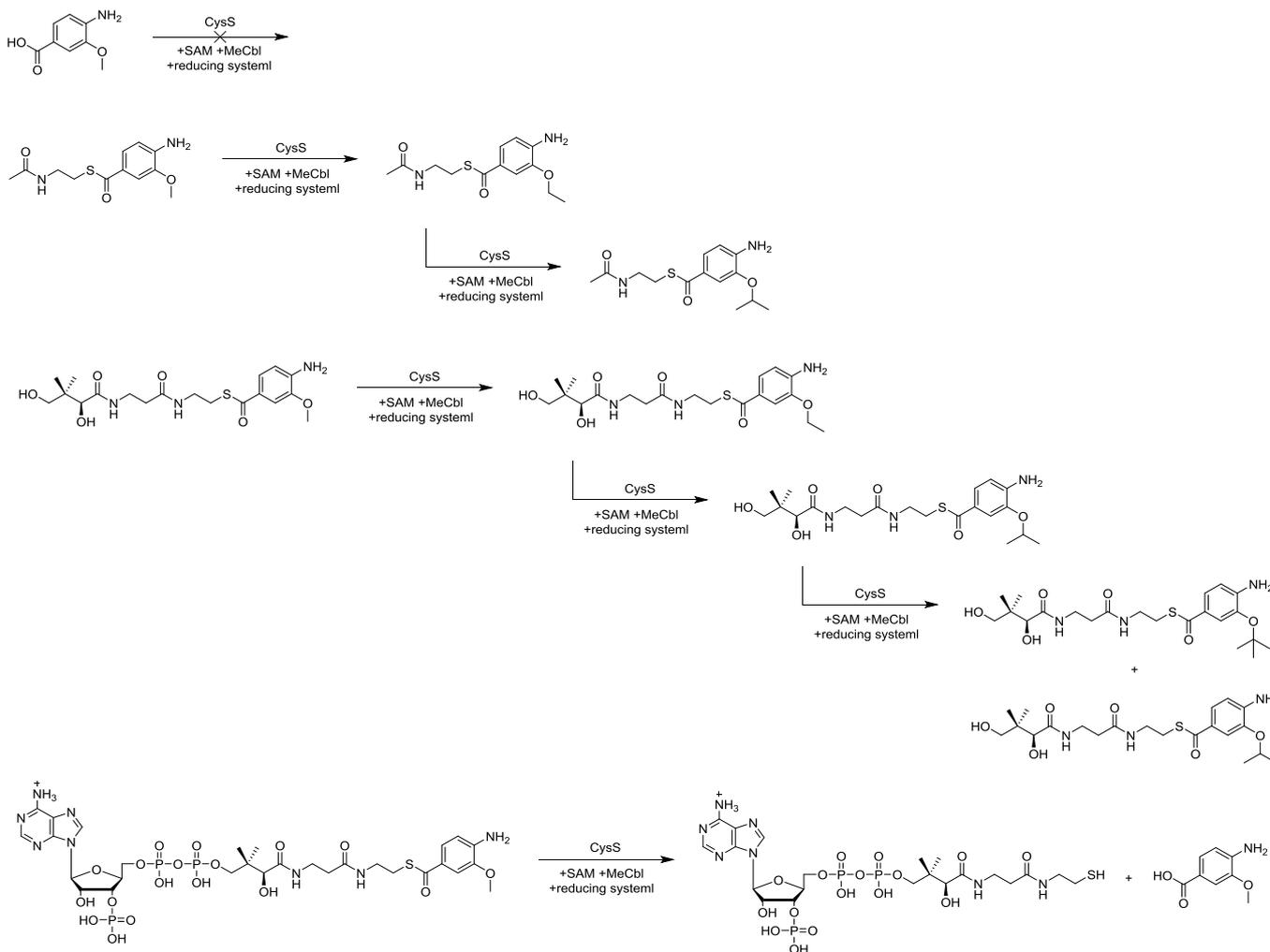


Figure 3.13: *In vitro* activity profile of CysS on 3-methoxy-pABA, SNAC-3-methoxy-pABA, pentheteinyl-3-methoxy-pABA and 3-methoxy-pABA-CoA

### 3.4 B12 feeding

It is well established that medium optimization through the feeding of specific amino acid substrates as well as various cofactors can increase the yield of secondary metabolites isolated from a culture broth. The iterative mode of action of the radical SAM CysS implies a great consumption of methyls from the one carbon metabolism suggesting that this reaction might be a bottleneck in the biosynthesis, this can however be by passed by feeding of large amounts of methyl-cobalamine (B12) to the medium acting directly as methyl donor for this reaction.

The feeding of 1  $\mu\text{g/ml}$  B12, which is a quite moderate amount, already lead to a 33 fold increase in the production of Cystobactamides indicating that the methylation reactions where indeed the first bottleneck in the biosynthesis. However increasing a further tenfold the B12 content of the medium to 10  $\mu\text{g/ml}$  only lead to moderate further increase in yield to 35 fold which means that either another bottleneck has been reached or the greater availability of cofactor does not influence the kinetic of the reaction further.

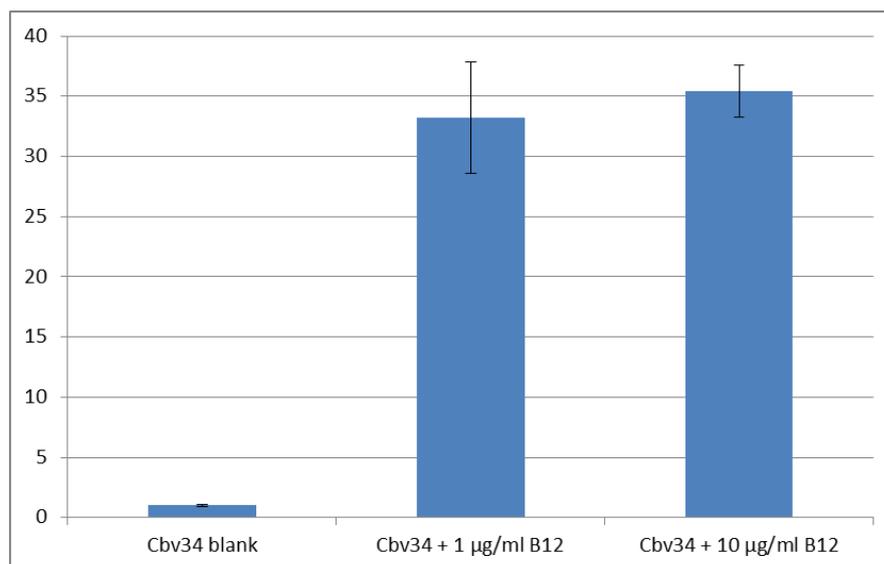


Figure 3.14: Cystobactamide 919-1 yield after B12 feeding (blank was normalized to 1)

## 4 Conclusion

In summary we successfully elucidated the biosynthesis of the isopropoxyl decoration from methoxyl groups in Cystobactamides by *in vitro* characterization of the biochemistry of the iterative radical SAM methyl transferase CysS. The enzyme is active *in trans* on PCP loaded substrate and proceeds stepwise from a methoxyl group to ethoxyl to isopropoxyl and even to sec- or tert-butoxyl, via iterative abstraction of single hydrogen atom followed by a radical substitution on methyl cobalamin. This is the first report describing the biosynthesis of isopropoxyl groups, a rare modification on secondary metabolites and it further extends the catalytic abilities of the diverse group of radical SAM enzymes to iterative C-methylation to yield branched alkoxy groups.

## References

1. Martin, J. L., and McMillan, F. M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold, *Current opinion in structural biology* 12, 783–793.
2. Liscombe, D. K., Louie, G. V., and Noel, J. P. (2012) Architectures, mechanisms and molecular evolution of natural product methyltransferases, *Natural product reports* 29, 1238–1250.
3. Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003) Many paths to methyltransfer. A chronicle of convergence, *Trends in biochemical sciences* 28, 329–335.
4. Cantoni, G. L. (1975) Biological methylation: selected aspects, *Annual review of biochemistry* 44, 435–451.
5. Finkelstein, J. D., and Martin, J. J. (2000) Homocysteine, *The International Journal of Biochemistry & Cell Biology* 32, 385–389.
6. Zhang, Q., van der Donk, Wilfred A, and Liu, W. (2012) Radical-mediated enzymatic methylation: a tale of two SAMs, *Accounts of chemical research* 45, 555–564.
7. Kozbial, P. Z., and Mushegian, A. R. (2005) Natural history of S-adenosylmethionine-binding proteins, *BMC structural biology* 5, 19.
8. Frey, P. A., Hegeman, A. D., and Ruzicka, F. J. (2008) The Radical SAM Superfamily, *Critical reviews in biochemistry and molecular biology* 43, 63–88.
9. Boal, A. K., Grove, T. L., McLaughlin, M. I., Yennawar, N. H., Booker, S. J., and Rosenzweig, A. C. (2011) Structural basis for methyl transfer by a radical SAM enzyme, *Science (New York, N.Y.)* 332, 1089–1092.
10. Dowling, D. P., Vey, J. L., Croft, A. K., and Drennan, C. L. (2012) Structural diversity in the AdoMet radical enzyme superfamily, *Biochimica et biophysica acta* 1824, 1178–1195.
11. Wang, Y., Schnell, B., Baumann, S., Müller, R., and Begley, T. P. (2017) Biosynthesis of Branched Alkoxy Groups: Iterative Methyl Group Alkylation by a Cobalamin-Dependent Radical SAM Enzyme, *Journal of the American Chemical Society* 139, 1742–1745.
12. Walker, A. M., Sattler, S. A., Regner, M., Jones, J. P., Ralph, J., Vermerris, W., Sattler, S. E., and Kang, C. (2016) The Structure and Catalytic Mechanism of Sorghum bicolor Caffeoyl-CoA O-Methyltransferase, *Plant physiology* 172, 78–92.
13. Nelson, J. T., Lee, J., Sims, J. W., and Schmidt, E. W. (2007) Characterization of SafC, a catechol 4-O-methyltransferase involved in saframycin biosynthesis, *Applied and environmental microbiology* 73, 3575–3580.
14. Singh, S., Nandurkar, N. S., and Thorson, J. S. (2014) Characterization of the calicheamicin orsellinate C2-O-methyltransferase CalO6, *ChemBiochem : a European journal of chemical biology* 15, 1418–1421.
15. Berkhan, G., Merten, C., Holec, C., and Hahn, F. (2016) The Interplay between a Multifunctional Dehydratase Domain and a C-Methyltransferase Effects Olefin Shift in

Ambruticin Biosynthesis, *Angewandte Chemie (International ed. in English)* 55, 13589–13592.

16. Guo, X., Crnovcic, I., Chang, C.-Y., Luo, J., Lohman, J. R., Papinski, M., Bechthold, A., Horsman, G. P., and Shen, B. (2018) PokMT1 from the Polyketomycin Biosynthetic Machinery of *Streptomyces diastatochromogenes* Tü6028 Belongs to the Emerging Family of C-Methyltransferases That Act on CoA-Activated Aromatic Substrates, *Biochemistry* 57, 1003–1011.

17. Herbst, D. A., Boll, B., Zocher, G., Stehle, T., and Heide, L. (2013) Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes, *The Journal of biological chemistry* 288, 1991–2003.

18. Schwarzer, D., Mootz, H. D., Linne, U., and Marahiel, M. A. (2002) Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases, *Proceedings of the National Academy of Sciences of the United States of America* 99, 14083–14088.

19. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of applied crystallography* 40, 658–674.

20. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta crystallographica. Section D, Biological crystallography* 60, 2126–2132.

21. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L.-W., Jain, S., Kapral, G. J., Grosse Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2011) The Phenix software for automated determination of macromolecular structures, *Methods (San Diego, Calif.)* 55, 94–106.

22. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta crystallographica. Section D, Biological crystallography* 67, 355–367.

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

24. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

25. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, *Protein engineering* 8, 127–134.

26. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction, *Nature methods* 12, 7–8.

27. Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F., and Miller, N. E. (2001) Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods, *Nucleic acids research* 29, 1097–1106.

- 
28. Wang, S. C., and Frey, P. A. (2007) S-adenosylmethionine as an oxidant: the radical SAM superfamily, *Trends in biochemical sciences* 32, 101–110.
29. Youkwan, J., Sutthivaiyakit, S., and Sutthivaiyakit, P. (2010) Citrusosides A-D and furanocoumarins with cholinesterase inhibitory activity from the fruit peels of *Citrus hystrix*, *Journal of natural products* 73, 1879–1883.
30. Tatsimo, S. J., Tane, P., Srinivas, P. V., Sondengam, B. L., Melissa, J., Okunji, C. O., Schuster, B. M., Iwu, M. M., and Khan, I. A. (2005) Novel antimicrobial diterpenoids from *Turraeanthus africanus*, *Planta medica* 71, 1145–1151.
31. Lohman, J. R., Huang, S.-X., Horsman, G. P., Dilfer, P. E., Huang, T., Chen, Y., Wendt-Pienkowski, E., and Shen, B. (2013) Cloning and sequencing of the kedarcidin biosynthetic gene cluster from *Streptoalloteichus* sp. ATCC 53650 revealing new insights into biosynthesis of the enediyne family of antitumor antibiotics, *Molecular bioSystems* 9, 478–491.
32. Rachid, S., Scharfe, M., Blöcker, H., Weissman, K. J., and Müller, R. (2009) Unusual chemistry in the biosynthesis of the antibiotic chondrochlorens, *Chemistry & biology* 16, 70–81.
33. Marous, D. R., Lloyd, E. P., Buller, A. R., Moshos, K. A., Grove, T. L., Blaszczyk, A. J., Booker, S. J., and Townsend, C. A. (2015) Consecutive radical S-adenosylmethionine methylations form the ethyl side chain in thienamycin biosynthesis, *Proceedings of the National Academy of Sciences of the United States of America* 112, 10354–10358.
34. Freeman, M. F., Gurgui, C., Helf, M. J., Morinaka, B. I., Uria, A. R., Oldham, N. J., Sahl, H.-G., Matsunaga, S., and Piel, J. (2012) Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides, *Science (New York, N.Y.)* 338, 387–390.
35. Watanabe, K., Hotta, K., Nakaya, M., Praseuth, A. P., Wang, C. C. C., Inada, D., Takahashi, K., Fukushi, E., Oguri, H., and Oikawa, H. (2009) *Escherichia coli* allows efficient modular incorporation of newly isolated quinomycin biosynthetic enzyme into echinomycin biosynthetic pathway for rational design and synthesis of potent antibiotic unnatural natural product, *Journal of the American Chemical Society* 131, 9347–9353.
36. Lee, G. F., and Stumm, W. (1960) Determination of Ferrous Iron in the Presence of Ferric Iron With Bathophenanthroline, *Journal - American Water Works Association* 52, 1567–1574.
37. Bridwell-Rabb, J., Zhong, A., Sun, H. G., Drennan, C. L., and Liu, H.-w. (2017) A B12-dependent radical SAM enzyme involved in oxetanocin A biosynthesis, *Nature* 544, 322–326.
38. Takahashi, Y., and Tokumoto, U. (2002) A third bacterial system for the assembly of iron-sulfur clusters with homologs in archaea and plastids, *The Journal of biological chemistry* 277, 28380–28383.



# Chapter 4: From CysH to CysK - The biosynthesis of hydroxy-isoasparagine and its incorporation into the peptide chain

## Introduction

As their name suggests tailoring reactions are often considered as modifications posterior to the biosynthesis of the natural product core from primary metabolism building blocks, but tailoring can actually be performed at every steps of the biosynthetic process. The biosynthesis of an uncommon building block precursor can take place before and after loading on the assembly line, in the latter case the reaction can be performed both *in cis* by a specific domain or *in trans* by an independent tailoring enzyme.(1) Thus, the biosynthesis of building blocks through precursor tailoring, *in cis* and *in trans* tailoring steps can constitute an independent pathway in the global biosynthesis scheme. In the Cystobactamide biosynthesis, production of the  $\beta$ -hydroxy-isoasparagine moiety from asparagine is under the responsibility of three proteins independent of the general assembly line.

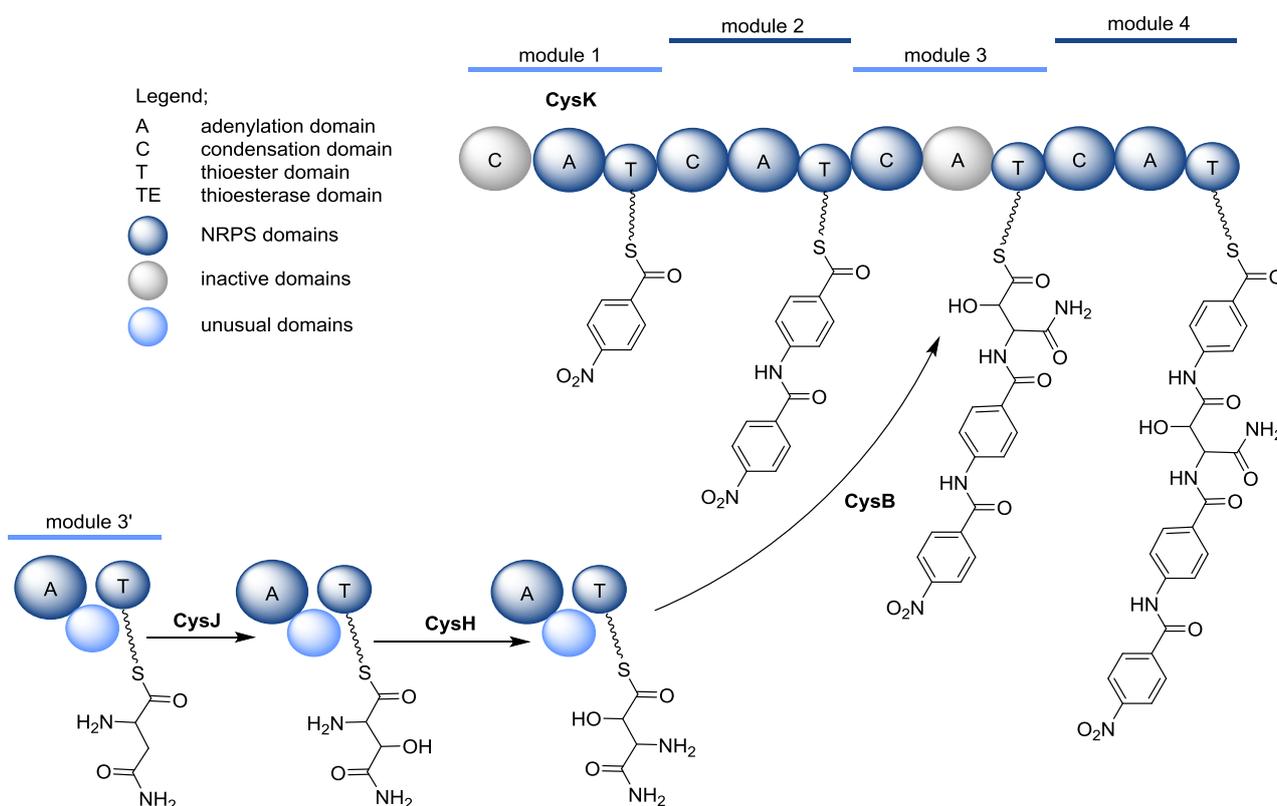


Figure 4.1: Biosynthesis pathway of the  $\beta$ -hydroxy-isoasparagine moiety

## From CysH to CysK - The biosynthesis of hydroxy-isoasparagine and its incorporation into the peptide chain

The  $\beta$ -hydroxylated isoasparagine moiety is unprecedented in natural products which made the understanding of the biosynthetic process that leads to its formation of great interest. The related isoaspartate has been reported twice in natural products, in Cyclocitropside C and Telomycin, but in the first case the isomerization of aspartate was postulated to be spontaneous similarly to the process observed in proteins (2) whereas the Telomycin assembly line seems to activate the side chain acid rather than the  $\alpha$ -carboxylic acid of aspartate leading to this reversed conformation.(3) *In silico* analysis of the cluster revealed two potential proteins that could be responsible for the biosynthesis of  $\beta$ -hydroxy-isoasparagine; the first one was CysH which is a three domain NRPS module independent from the assembly line, it lacks a condensation domain but features an unprecedented domain loosely related to transamidases; the second one is CysJ which shares high similarity with  $\alpha$ -ketoglutarate dependent dioxygenases like the aspartyl hydroxylase SyrP (4) and the taurine hydroxylase TauD.(5)

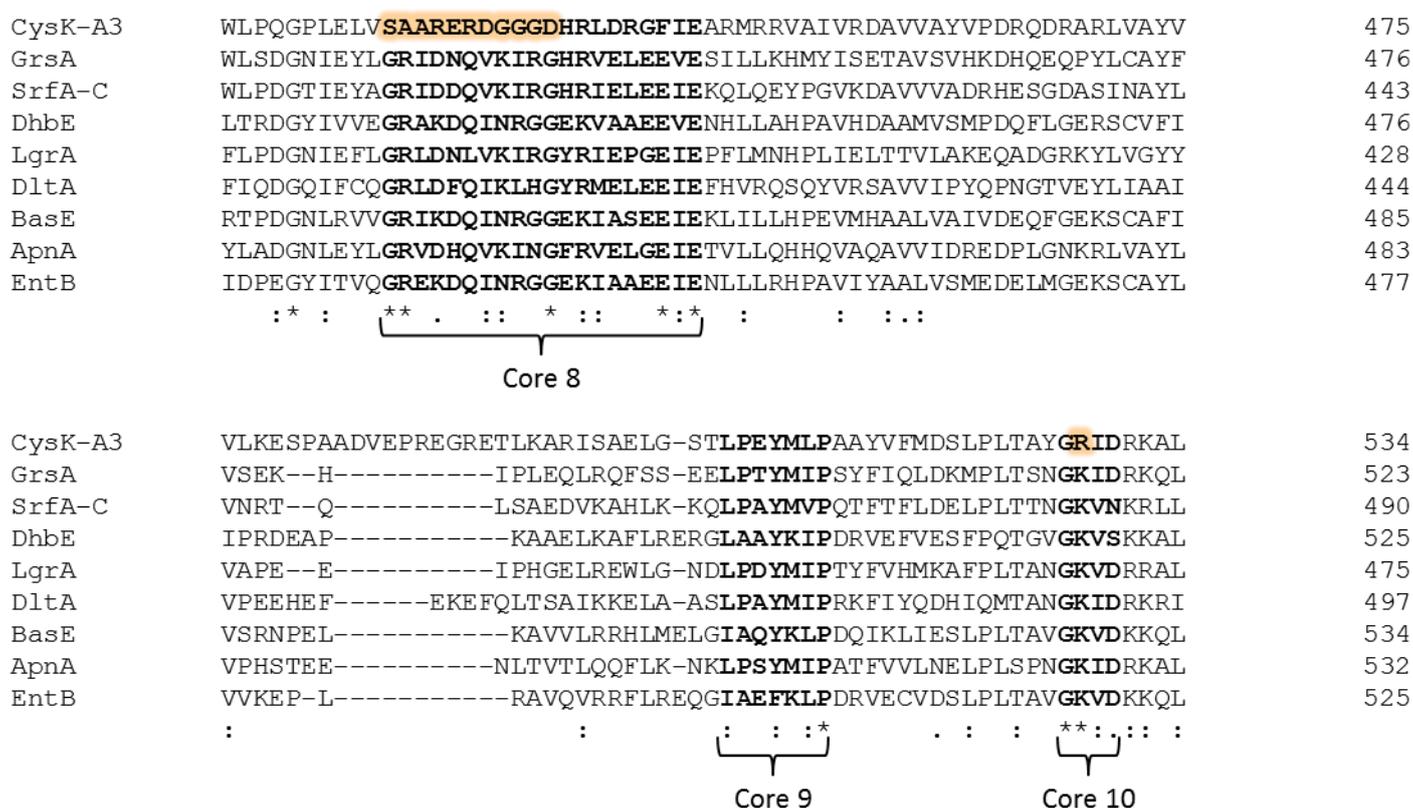


Figure 4.2: CysK-A3 sequence alignment with reference adenylation domain sequences for which structural data is available, the core motives are represented in bold and the mutated residues are highlighted in orange

Further *in silico* analysis revealed that the adenylation domain in the third module of the assembly line is lacking the catalytic lysine in the core motive 10 in addition to harboring a series of mutations in the core motive 8 responsible for the hinge between the core domain and the subdomain.(6) These mutations probably result in the A3 domain being inactive while retaining its fold, thus an alternative loading mechanism must be at play to activate and

bring the building block on the PCP3 position of the assembly line. This process is postulated to be catalyzed by the third enzyme of the  $\beta$ -hydroxy-isoasparagine biosynthesis pathway, CysB, which was speculated to be able to shuttle the building block from CysH to the assembly line. CysB is an unusual thioesterase related acyl transferase homologous to the threonine aminoacyl transferase SyrC.(7)

In this chapter we report the *in vitro* characterization of the biosynthesis of both the  $\beta$ -hydroxy-isoasparagine linker observed in Cystobactamides and of the  $\beta$ -cyano-alanine linker observed in Albicidin by a never before reported *in cis* tailoring NRPS domain, X, in conjunction with an *in trans* tailoring enzyme, CysJ. Furthermore we could experimentally prove the suspected inactivity of the adenylation domain of the third module of the assembly line and investigate the shuttling mechanism performed by a rare aminoacyl transferase, CysB, which circumvents the absence of adenylation and self-loading activity. Finally we could determine that the methylation observed on hydroxyl of  $\beta$ -hydroxy-isoasparagine in the final product cannot be performed by CysQ at any point of this reaction sequence and hence probably happens after release of the peptidic chain. However, the unavailability of a suitable desmethyl-Cystobactamide substrate for this reaction did not allow us to confirm this result.

## Chapter 4.1: CysH and CysJ - In cis and in trans tailoring working hand in hand for hydroxy-isoasparagine biosynthesis

### 1 Introduction

*In cis* tailoring domains in NRPS are widespread and perform various enzymatic activities, they can be found in two distinct positions in the module architecture. While the epimerization domain (E), which performs the racemization of L amino acids into a D configuration, is positioned after the PCP forming a C-A-T-E architecture.(8) The N-methyltransferase domain (N-MT) performing the N-methylation of the amide bond and the oxidation domain (Ox) which performs the oxidation of the thiazoline and oxazoline rings to thiazol or oxazol are found integrated in the A domain shortly after the hinge region separating the A core domain ( $A_{\text{core}}$ ) from the smaller A sub ( $A_{\text{sub}}$ ) domain, in this case the architecture is C- $A_{\text{core}}$ -MT- $A_{\text{sub}}$ -T.(9) The Cystobactamide biosynthesis cluster features an unprecedented *in cis* tailoring domain in the stand alone NRPS module CysH presenting an  $A_{\text{core}}$ -X- $A_{\text{sub}}$ -T architecture. The Adenylation domain specificity predicted through analysis of the Stachelhaus code corresponds to asparagine leading to the hypothesis that CysH might be activating asparagine in order to transform it into the unique  $\alpha$ -hydroxy- $\beta$ -asparagine observed in the major Cystobactamide compound 919-1.(10)

It is noteworthy that while Cystobactamides and Albicidins feature different linker units,  $\alpha$ -hydroxy- $\beta$ -asparagine for the former and  $\beta$ -cyano-alanine for the latter, an X domain can also be observed in the Albicidin cluster embedded in AlbIV in the same way it is observed in CysH. Due to a very high identity in sequence (68,5% identity / 83% similarity) from X domains in Cystobactamide with its Albicidin counterpart, it would be safe to assume that they are performing the same activity. It is, however, also possible that a slight modification of the tailoring pattern would lead to the two radically different products. This could be linked to the absence of an asparagine- $\beta$ -hydroxylase (CysJ) homologue in the Albicidin cluster which implies that the X domain performs its activity on a different substrate which is hydroxylated only in the Cystobactamide pathway. This discrepancy seems to lead to a dramatic change in the reactivity since an amino-hydroxy group exchange on  $\beta$ -hydroxy-asparagine yielding  $\alpha$ -hydroxy- $\beta$ -asparagine is observed in the Cystobactamide biosynthesis and amide dehydration of asparagine is taking place in the Albicidin pathway leading to a  $\beta$ -cyano-alanine. (Figure 4.3)

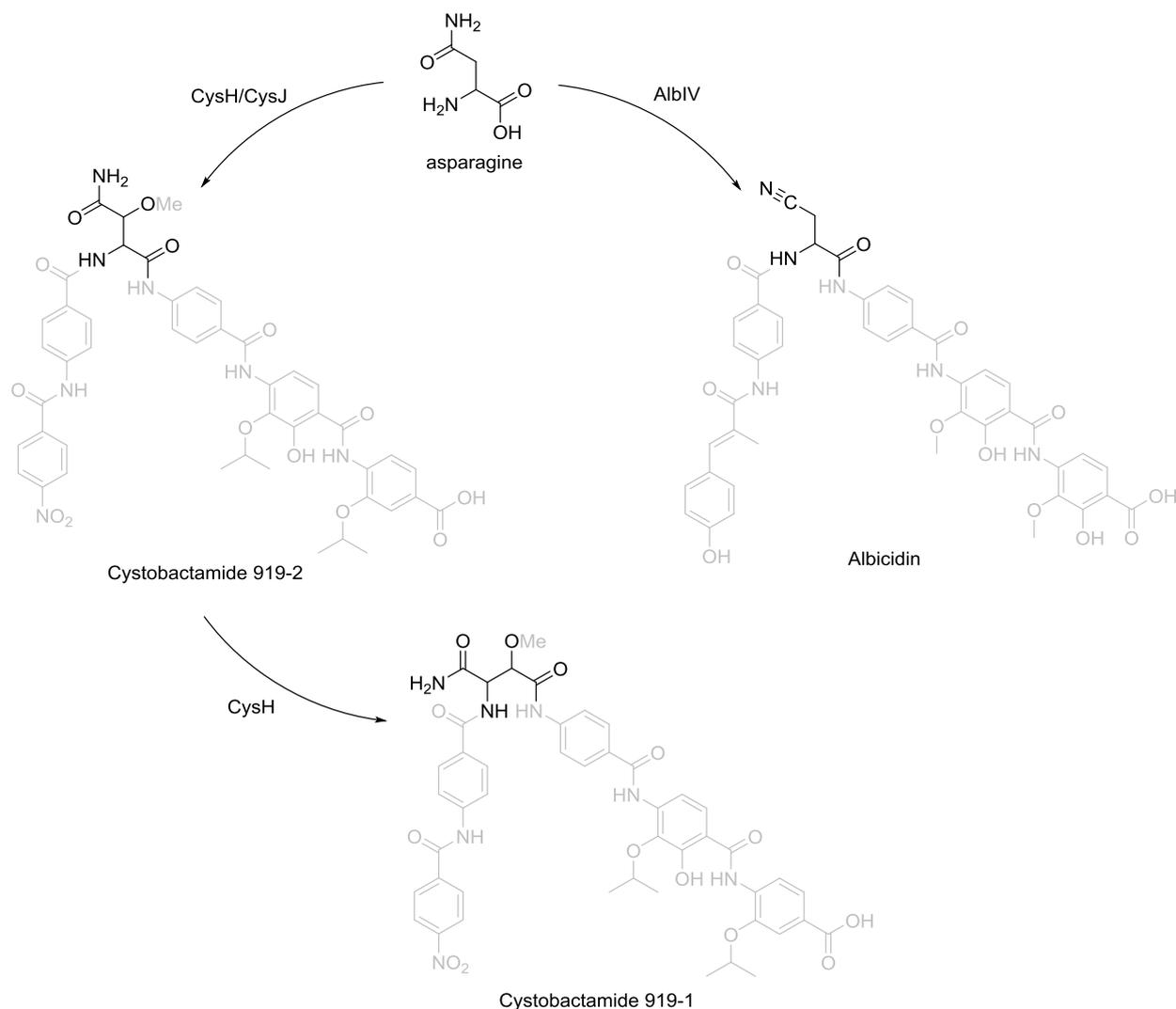


Figure 4.3: Differential integration of asparagine in Cystobactamides and Albicidin, tailored by CysH and CysJ as β-amino-α-hydroxy-asparagine in Cystobactamide 919-1 and by AlbIV alone as β-cyano-alanine in Albicidin

β-hydroxylation of amino acids is a common modification happening both during post translational modification of proteins and during NRPS biosynthesis.<sup>(11)</sup> It is a major source of structural diversity in peptidic natural products since among twenty proteinogenic amino acids, twelve have been reported as β-hydroxylated derivatives in a number of natural products (12–20), and of the eight remaining five cannot carry this modification: serine, threonine and cysteine are already β-hydroxylated or β-thiolated, glycine has no β-carbon and β-hydroxylation of alanine would result in serine. This makes the β-hydroxylation one of the most widespread modifications in NRPS. This process has been proven to be often carried out *in trans* on PCP bound substrate and can be catalyzed by three different type of enzymes; the most widespread being cytochrome P450 monooxygenases, but β-hydroxylation can also be catalyzed by α-ketoglutarate (α-KG) dependent dioxygenases and di-iron monooxygenase.

In silico analysis of CysJ revealed its high similarity to the well described  $\alpha$ -KG dependent dioxygenase taurine hydroxylase TauD (21), and even closer similarity to the aspartyl hydroxylase SyrP (4), which has been shown to be active *in trans* on PCP loaded aspartate.  $\alpha$ -KG dependent dioxygenases rely on a single  $\text{Fe}^{\text{II}}$  center bound in the active site by a conserved His-X-Asp/Glu- $X_n$ -His facial triad (22), the enzyme couples the oxidative decarboxylation of the  $\alpha$ -KG cofactor to the hydroxylation of the substrate.(23) The catalysis relies on the formation of a  $\text{Fe}^{\text{III}}$ -superoxide complex resulting after oxidative decarboxylation of  $\alpha$ -ketoglutarate in a  $\text{Fe}^{\text{IV}}$ -oxo complex able to perform the homolytic cleavage of the  $\text{C}\beta$ -H bond, resulting in a  $\text{Fe}^{\text{III}}$ -hydroxy complex and a radical on  $\text{C}\beta$  which subsequently attacks the hydroxyl ultimately regenerating the  $\text{Fe}^{\text{II}}$  center of the enzyme.(24) (Figure 4.4) Structurally, the protein features a central Jelly Roll fold with 8  $\beta$ -strands for  $\text{Fe}^{\text{II}}$  and  $\alpha$ -KG binding and surrounding additional structural features for substrate specificity and protein-protein interaction.(5)

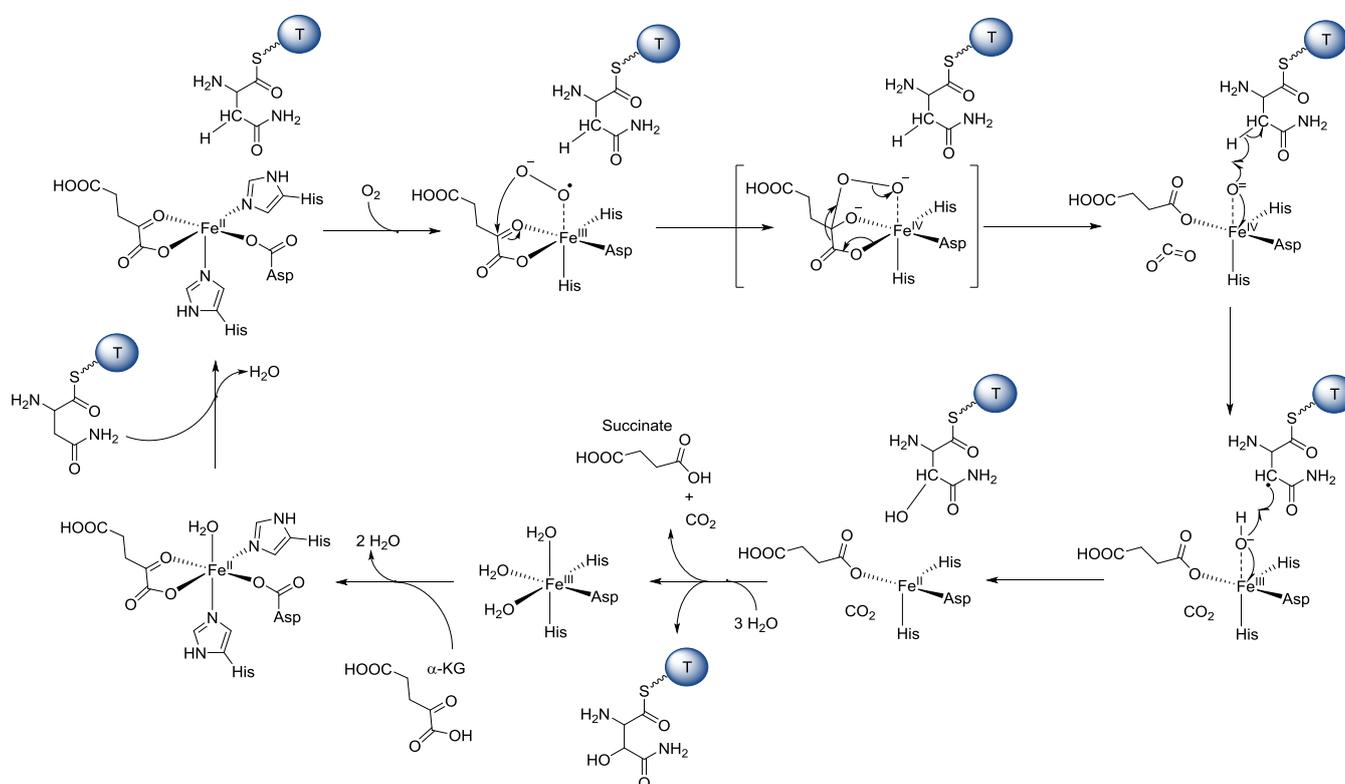


Figure 4.4: CysJ catalytic cycle as example for the mode of action of an  $\alpha$ -KG dependent dioxygenase

Not only has  $\beta$ -hydroxylation repeatedly been observed to be carried out *in trans* on PCP bound intermediates but the occurrence of this particular tailoring step is often coupled with single A-T NRPS modules independent from the assembly line (11), such as in Novobiocin biosynthesis where NovH is an A-T module acting as carrier for the Cyp450 Tyrosine- $\beta$ -hydroxylase NovI.(25) The same process is also observed for Histidine- $\beta$ -hydroxylation by NikP1 and NikQ in Nikkomycin biosynthesis.(26) In the Cystobactamide biosynthesis this particular pathway organization can be observed too, indeed a single A-X-T

NRPS module, CysH, able to load asparagine and isomerize hydroxy-asparagine can be found. Preliminary *in vitro* activity data on CysJ revealed that it was incapable of processing free asparagine or isoasparagine leading to the hypothesis that CysJ could catalyze an *in trans* tailoring step carried out either on CysH or on the assembly line.

## 2 Material and Methods

### 2.1 Cloning, overexpression and purification

Standard protocols were used for DNA cloning, transformation, amplification, and purification. Genomic DNA was extracted from *Cystobacter velatus* Cbv34 grown in CTT medium, the DNA fragment encoding CysJ was amplified using the forward primer 5'TATCATATGACCGGTAATTTGGATAGCGCGG and the reverse primer 5'TATGGATCCTTACGAGCGCCCTGAGTTCGTTGC. The amplified DNA fragments were digested with NdeI and BamHI, and cloned into pET-28b with an N-terminal His6 tag. The resulting constructs were sequenced to verify that no mutation had been introduced during PCR.

The DNA fragment encoding CysH was amplified using the forward primer 5'TATCCATGGACAATCGAGAGATCGC and the reverse primer 5'TATGGATCCTTATCCCCTGTATGCAGGCG. The amplified DNA fragments were digested with NcoI and BamHI, and cloned into pHis-SUMO-TEV with an N-terminal His6 tag. The resulting constructs were sequenced to verify that no mutation had been introduced during PCR.

Recombinant protein expression was carried out in *E.coli* BL21 (DE3) grown in LB medium supplemented with 50 µg/ml of kanamycin. The culture was inoculated 1/10 from a fully grown overnight culture and was initially grown at 37°C until it reached an OD of 0.6 to 0.9, at which point the temperature was decreased to 16°C. Induction was initiated after 30 minutes at 16°C with 0,1 mM isopropyl-β,D-thiogalactopyranoside (IPTG). The cells were harvested 16h after induction by centrifugation at 8000 g for 10 minutes at 4°C and resuspended in lysis buffer (25 mM TRIS pH 7,5; 150 mM NaCl; 20 mM imidazole). The cells were lysed by passage at 1500 bar at 4°C through a M-110P Microfluidizer (Microfluidics) after which the crude extract was centrifuged at 45000 g for 15 minutes at 4°C, the supernatant was treated as soluble protein fraction.

Affinity chromatography was performed on an Äkta Avant system while size exclusion chromatography was performed on an Äkta Pure system. The soluble fraction of the lysed cells was applied to 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10 CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole.

For CysJ the pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200 pg SEC column equilibrated in running buffer (25 mM TRIS pH 7,5; 150 mM NaCl; 2 mM DTT). While for CysH the pooled fractions were applied to a

HiPrep 26/10 desalting column equilibrated in running buffer (25 mM TRIS pH 7,5; 150 mM NaCl; 2 mM DTT). The resulting fractions were pooled ( $\approx$  30 ml) and incubated overnight at 4°C with TEV protease (1 mg/20 mg protein). After  $\approx$  16 h incubation 20 mM imidazole was added to the solution prior to loading on a 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10 CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole. The pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200 pg SEC column equilibrated in running buffer (25 mM TRIS pH7,5; 150 mM NaCl; 2 mM DTT). The pooled fractions were controlled via SDS-PAGE, concentrated and frozen at -80°C after addition of 25% glycerol.

## 2.2 Adenylation assay

The assay was performed at room temperature in 96 well plate with a total volume of 100  $\mu$ L per well. 10  $\mu$ M CysH were mixed with 1 mM ATP, 1 mM amino acid, 2 UI Pyrophosphatase, 10 mM MgCl<sub>2</sub>, 50 mM TRIS, pH 7,6 and 100 mM NaCl. The phosphate liberated during the adenylation reaction was measured with a Malachite green assay. 20  $\mu$ l of the malachite green dye (900  $\mu$ l malachite green mix (60  $\mu$ l H<sub>2</sub>SO<sub>4</sub> +300  $\mu$ l H<sub>2</sub>O + 0,44 g malachite green) , 100  $\mu$ l ammonium molybdate 30 % w/v, 16  $\mu$ l Tween 20) were added to the 100  $\mu$ l reactions and the coloration was left to develop for 10 min. The absorbance at 600 nm was measured in a TECAN infinity 200 pro plate reader.

## 2.3 CysJ Activity assays

In a total volume of 50  $\mu$ l, 5  $\mu$ M CysH was incubated with 1 mM Asparagine, 1 mM ATP and 1 mM  $\alpha$ -KG in 25 mM TRIS pH7,5; 150 mM NaCl; 10 mM MgCl<sub>2</sub>; 50  $\mu$ M Fe<sup>II</sup>SO<sub>4</sub>, finally 500 nM CysJ was added and the reaction was incubated for 1 h at room temperature. The solution was analysed by protein LC-MS or the substrate was unloaded and derivatized for independent LC-MS analysis.

## 2.4 Unloading assays

After the reaction is complete 100 mM cysteamine are added from a newly prepared cysteamine 1M stock solution and the resulting solution is incubated at 30°C for 1h with slow shaking. The reaction mix is transferred in a 2 ml glass vial and 45  $\mu$ l ethanol:pyridine (4:1) solution + 5  $\mu$ l ethyl chloroformiate (ECF) are added and vortexed. After addition of 200  $\mu$ l water the derivatized NN-diethoxycarbonyl  $\beta$ -hydroxyasparaginyl dicysteamine is extracted with 2x300  $\mu$ l EtOAc+1 % ECF. The collected organic layers are dried, redissolved in MeOH and analyzed through LC-MS.

All measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 x 2.1 mm, 1.7  $\mu$ m dp column (Waters, Germany). Separation of 1  $\mu$ l sample was achieved by a linear gradient from (A) H<sub>2</sub>O + 0.1 % FA to (B) MeOH + 0.1 % FA at a flow rate of 600  $\mu$ L/min and 45°C. 0 min: 95 % A / 5 % B, 0.5 min: 95 % A / 5 % B, 1.5 min:

80 % A / 20 % B, 14.5 min: 70 % A / 30 % B, 17,5 min: 5 % A / 95 % B, 20,5 min: 5 % A / 95 % B, 21 min: 95 % A / 5 % B. UV spectra were recorded by a DAD in the range from 200 to 600 nm. MS data was acquired with an Amazon Speed 3D ion trap mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. Mass spectra were acquired in centroid mode ranging from 150 – 1500 m/z.

## 2.5 Synthesis of the di(ethylcarbonyl)asparaginy-dicysteamine references

55,6 mg Trityl chloride (0,2 mmol, 2eq.), 19  $\mu$ l acetic anhydride (0,2 mmol, 2eq.) and 6,1  $\mu$ l concentrated H<sub>2</sub>SO<sub>4</sub> (0,115 mmol, 1,15eq.) are added to 500  $\mu$ l glacial acetic acid until dissolution. Then 35,4 mg IsoAsn (0,1 mmol, 1eq.) and 50  $\mu$ l DMF are added and left to react overnight at 60°C in an oil bath. The solution is slowly poured on 2 ml ice cold H<sub>2</sub>O, the pH is adjusted to 6 by addition of 10M NaOH and left on ice for 1h, the precipitate is then filtrated over silica gel washed with water and redissolved in DCM.

37,4 mg of Trityl-IsoAsn (100  $\mu$ mol, 1eq.) and 27,9  $\mu$ l triethylamine (200  $\mu$ mol, 2eq.) were dissolved in 500  $\mu$ l THF, 34,5  $\mu$ l Di-tert-butyl dicarbonate were added and left to react at room temperature for 2h. Upon completion the reaction was quenched in water, the pH was acidified by addition of 1M HCl and the Boc-IsoAsn(trt)-OH was extracted with DCM.

50 mg Boc-Asn(trt)-OH or Boc-IsoAsn(trt)-OH (105  $\mu$ mol, 1eq.), 35,2  $\mu$ l DIPEA (367  $\mu$ mol, 3,5 eq.) and 60,1 mg PyBop (115  $\mu$ mol, 1,1 eq.) are dissolved in 2ml DCM and left to stir at room temperature for 30min. 29,8 mg Cysteamine hydrochloride (262  $\mu$ mol, 2,5eq.) and 1 ml DMSO are added and the reaction is left overnight at room temperature. The reaction is quenched in H<sub>2</sub>O and the product is extracted twice with DCM, the organic fractions are dried and the white precipitate is washed with pentane.

The previously obtained powder is left to react at room temperature overnight in 2 ml of an 88% TFA; 5% phenol; 5% H<sub>2</sub>O; 2% TIPS solution with addition of 2 eq. of cysteamine to prevent thiol disulfide exchange. The reaction is quenched in water, washed three times with DCM and the aqueous layer is lyophilized to yield pure asparaginy-dicysteamine.

The compound is dissolved in EtOH and left to react with 5% ECF before analysis

## 2.6 Protein MS analysis

CysH was analyzed on a Dionex Ultimate 3000 UPLC system (Thermo Scientific) coupled with an maXis4G Q-TOF MS (Bruker) using an ESI in positive mode. The samples were run on an Aeris Widepore XB-C8, 150 x 2.1 mm, 3.6  $\mu$ m dp column (Phenomenex, USA). LC conditions: A-Water + 0.1% FA; B-Acetonitrile + 0.1 % FA at a flow rate of 300  $\mu$ l/min and 45°C. 0 min: 98 % A / 2 % B, 0.5 min: 98 % A / 2 % B, 10.5 min: 25 % A / 75 % B, 13.5 min: 25 % A / 75 % B, 14 min: 98 % A / 2 % B. The LC flow was split to 75  $\mu$ l/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker

Daltonics, Bremen, Germany) using the standard Bruker ESI source. In the source region, the temperature was set to 180°C, the capillary voltage was 4000 V, the dry-gas flow was 6.0 l/min and the nebulizer was set to 1.1 bar. Mass spectra were acquired in positive ionization mode ranging from 150 – 2500 m/z at 2.5 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm (Copyright 1991-2004 Spectrum Square Associates, Inc.).

## 2.7 Asparagine feeding

25 mL M-medium was inoculated with 1mL of a high density overnight culture of Cbv34. 0,5 mL of XAD-7 was added on day 1 and 12,5 mg of <sup>15</sup>N<sub>2</sub> <sup>13</sup>C<sub>4</sub> labeled asparagine was fed over 4 days. On day 5 the cultures were centrifuged and the pellet and XAD were extracted with 2x10 mL methanol. The solvent was removed through rotary evaporation and the extract was dissolved in 2 mL methanol.

LC-ESI-MS was performed using a Thermo Scientific Dionex Ultimate 3000 UHPLC system coupled with a Bruker Amazon speed iontrap mass spectrometer using an ESI in positive mode. LC was performed on a Waters Acquity UPLC BEH C18 50 mm 1.7 μm column. LC conditions: A-Water + 0.1 % formic acid; B-Acetonitrile + 0.1 % formic acid; 0 min: 95% A / 5% B, 0.5 min: 95 % A / 5 % B, 9.5 min: 5 % A / 95 % B, 10.5 min: 5% A / 95 % B, 10.8 min: 95 % A / 5 % B, 11.8 min: 95 % A / 5 % B.

## 3 Results and Discussions

### 3.1 CysH - Adenylation and thiolation domain activity

CysH is a 108 kDa NRPS module with A<sub>core</sub>-X-A<sub>sub</sub>-T architecture, for *in vitro* investigation we cloned the *cysH* gene in a pSUMO-TEV vector and the protein could be readily overexpressed in *E.coli* BL21, it presented a distinctive brown color that faded after a few freeze-thaw cycles indicating a possible catalytic metal binding site. Three conserved cysteines could serve as Fe binding explaining the color. The specificity of the adenylation domain was predicted through its Stachelhaus code using NRPSpredictor2 to be asparagine which is in accordance with the moiety observed in the major Cystobactamide compounds.

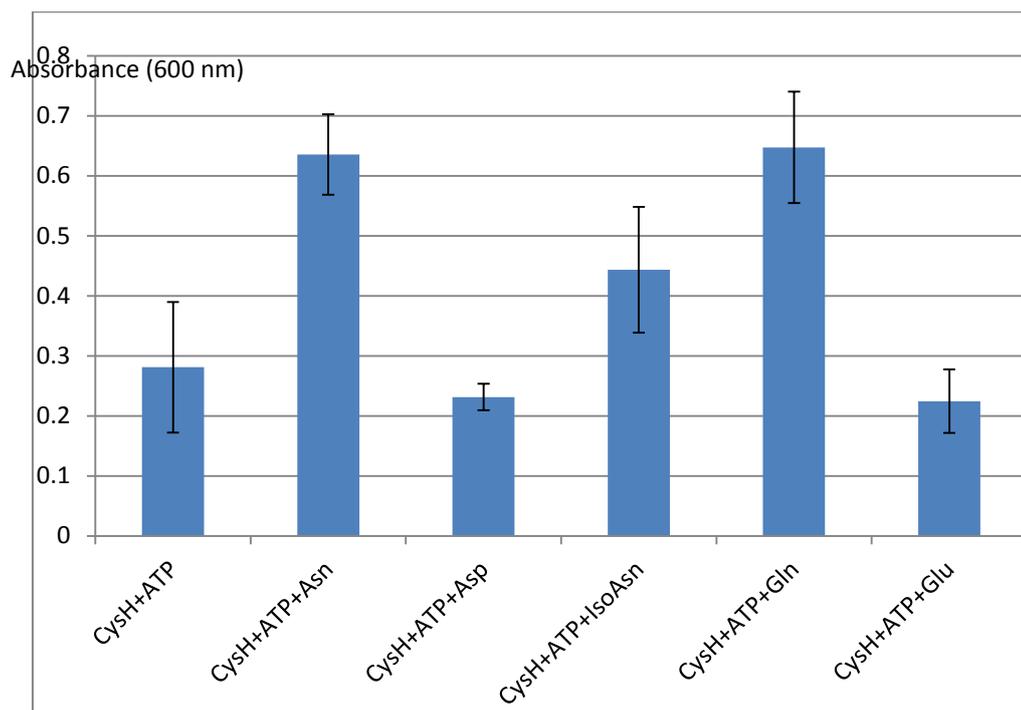


Figure 4.5: Adenylation activity of CysH shown as absorbance at 600 nm measured through Phosphate liberation for different amino acid substrates

The third amino acid, asparagine or aspartate, which was designated as “linker” between the two pABA chains is source of chemical diversity in Cystobactamides, indeed four different linkers are found in the most common Cystobactamide derivatives (Asn/IsoAsn/Asp/ IsoAsp).<sup>(10)</sup> It is however unclear whether the Aspartate and Isoaspartate linkers present in minor derivatives 920-1 and 920-2 are natural or arise through deamination during the extraction and isolation processes. To confirm the *in silico* substrate specificity prediction the adenylation specificity was assayed *in vitro* on purified protein using a malachite green assay to titrate phosphate liberation during the adenylation reaction. It revealed that despite the fact that the Stachelhaus code was directed towards asparagine the closely related glutamine is also a valid substrate for the A domain.

## CysH and CysJ - In cis and in trans tailoring working hand in hand for hydroxy-isoasparagine biosynthesis

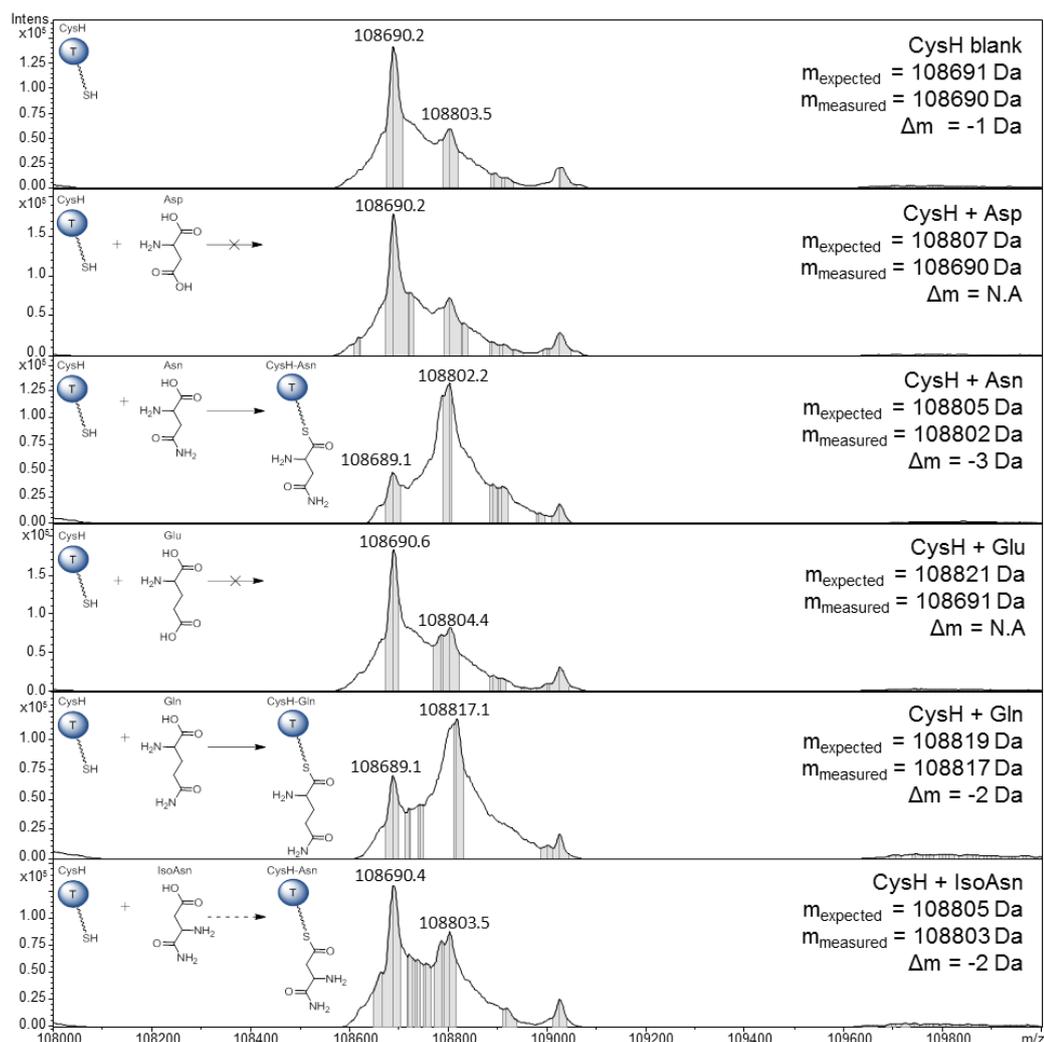


Figure 4.6: Substrate selectivity for the thiolation activity of CysH analyzed through full size protein MS after incubation with amino acids ( $\Delta m$  is in the range of 10-30 ppm which is an acceptable error given the size of the protein). For the sake of simplicity only the T domain of Module CysH A-X-T was depicted.

The substrate specificity prediction of the adenylation domain was confirmed experimentally. (Figure 4.5) Interestingly, Isoasparagine can be adenylated to some extent but since  $\alpha$ -asparagine is a better substrate it seems to imply that contrary to previously reported biosyntheses of natural products containing  $\beta$ -amino acids in this case the adenylation domain is not specific for the  $\beta$ -amino acid but rather for the natural  $\alpha$ -asparagine and thus the isomerization reaction would happen on PCP bound intermediates or on the final product rather than on a free precursor. Aspartate was not a valid substrate for adenylation hinting at an artificial origin of the minor derivatives Cystobactamide 920-1 and 920-2, it is however still possible that Asp could arise through tailoring but no deamidase was identified in the biosynthesis cluster. Another interesting result is that the glutamine can be adenylated despite the fact that no Cystobactamide derivative was observed harboring a glutamine linker, thus a downstream enzymatic process (shuttling or condensation) has to discriminate against glutamine to prevent its integration in the final product.

Protein MS was used to determine whether the protein was fully functional *in vitro* and if the loading specificity could be superposed with adenylation specificity. (Figure 4.6) As expected, aspartate and glutamate which are not activated cannot be loaded while the expected natural substrate asparagine seems to be the preferred substrate. However, glutamine which was unexpectedly adenylated is also accepted by the thiolation domain, furthermore the partially adenylated non-natural substrate isoasparagine can also get loaded to some extent.

### 3.2 CysJ - *in silico* analysis

An in depth *in silico* analysis of CysJ revealed the presence of the His98-Glu100-His206 facial triad responsible for iron binding, backing the initial functional assignment as  $\alpha$ -KG dependent dioxygenase. To confirm the *in silico* data CysJ was overexpressed and purified with high yields of 25 mg/l as a 6xHis tagged fusion protein and experimental confirmation of its activity was obtained. Initial activity assays on free asparagine were performed after the overexpression and analyzed through TLC revealed with ninhydrine. (Figure 4.7) The lack of turnover suggested that the enzyme was unable to process the free amino acid and that this tailoring step would happen *in trans* on Ppant bound substrate. Since CysH could be expressed to high titer and proved to be able to adenylate and load asparagine *in vitro* we were able to use the natural substrate rather than a SNAC ester homologue.

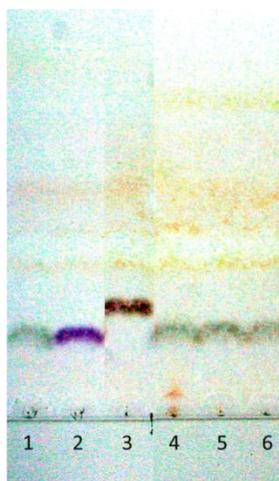


Figure 4.7: TLC showing the absence of visible turnover from asparagine incubated with CysJ,  $\alpha$ -KG and Fe. 1-Asn; 2-Asp; 3-IsoAsn; 4-Asn+CysJ; 5-Asn+CysJ+ $\alpha$ -KG; 6-Asn+CysJ denatured+ $\alpha$ -KG

### 3.3 Unloading assays

*In cis* and *in trans* tailoring is frequent in NRPS and PKS biosynthesis but since these reactions proceeds in a single turnover rather than being catalytic they generate very limited amounts of product to analyze which stays tethered to the carrier protein and are hence very challenging to analyze. This led us to develop a routine unloading procedure via trans-thioesterification by cysteamine.(27,28) While this method has proven efficient to analyze

## CysH and CysJ - In cis and in trans tailoring working hand in hand for hydroxy-isoasparagine biosynthesis

ACP bound fatty acids (29), in the case of asparagine the cysteamine bound analyte is small and very polar it was thus decided to further derivatize both free amines with ethyl carbamates prior to LC-MS analysis.(28) (Figure 4.8) Additionally the expected isomerization of the asparagine is impossible to observe through mass spectrometry and the very limited amounts of derivatized product that can be obtained are not sufficient for NMR analysis. It was thus decided to synthesize reference compounds for unloaded and derivatized asparagine and isoasparagine in order to identify them through their retention time in HPLC.

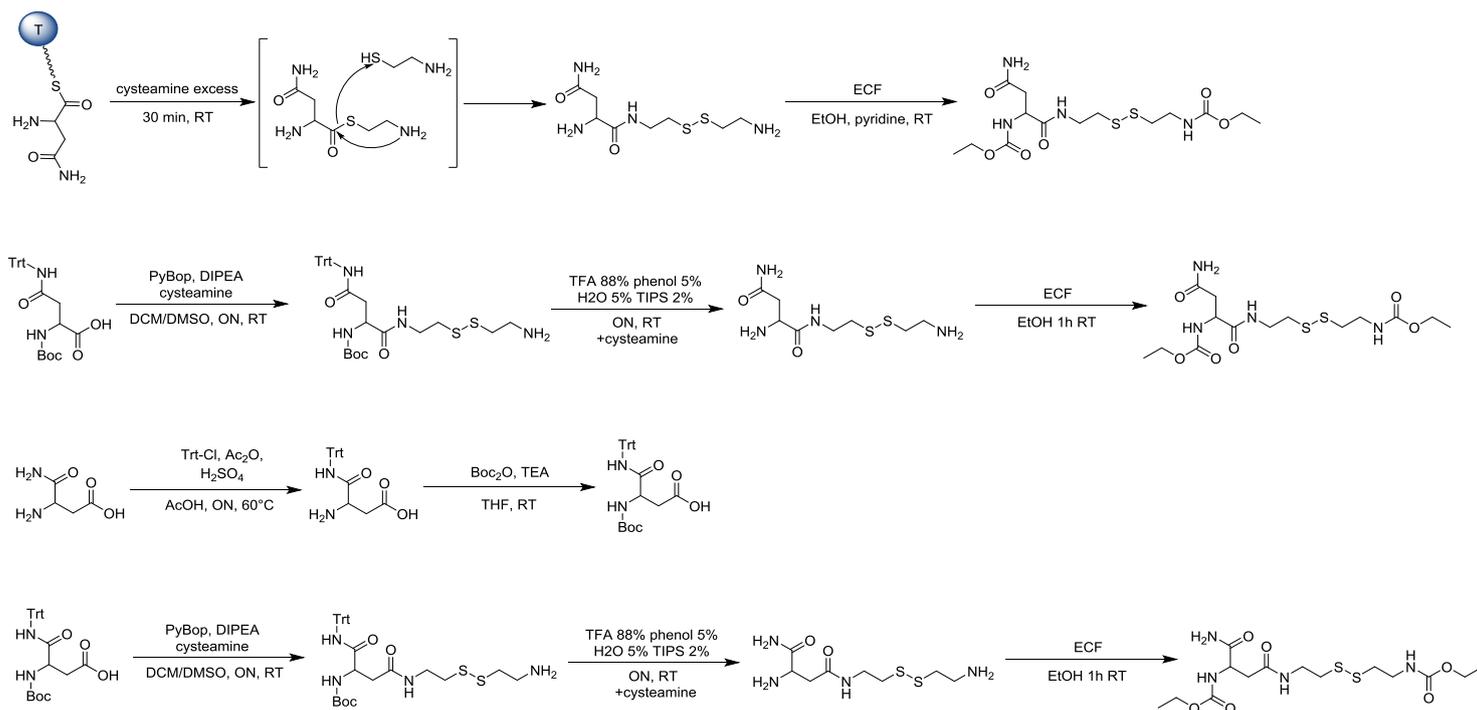


Figure 4.8: derivatization procedure and organic synthesis of reference compounds

Both asparagine and isoasparagine could be detected through this method but upon incubation (Figure 4.9) which CysJ and  $\alpha$ -ketoglutarate the +16 signature mass shift of a hydroxylation could be observed only on the asparagine bond derivative implying that the isoasparaginyl derivative is not a valid substrate for CysJ. This leads to the conclusion that the hydroxylation is the first tailoring step in the biosynthesis of  $\alpha$ -hydroxy- $\beta$ -asparagine. Without references it is however impossible to determine whether the peak observed at 12,2 min retention time is  $\beta$ -hydroxy- $\alpha$ -asparagine or  $\alpha$ -hydroxy- $\beta$ -asparagine but a small peak for the same mass at 11,8 min with a retention time shift comparable to the one between asparagine and isoasparagine can be observed which could imply that we observe only traces of  $\beta$ -hydroxy- $\alpha$ -asparagine and that the peak at 12,2 min corresponds to an already isomerized  $\alpha$ -hydroxy- $\beta$ -asparagine. The synthesis of  $\beta$ -hydroxy- $\alpha$ -asparagine and  $\alpha$ -hydroxy- $\beta$ -asparagine references would be necessary to determine the exact identity of this compound.

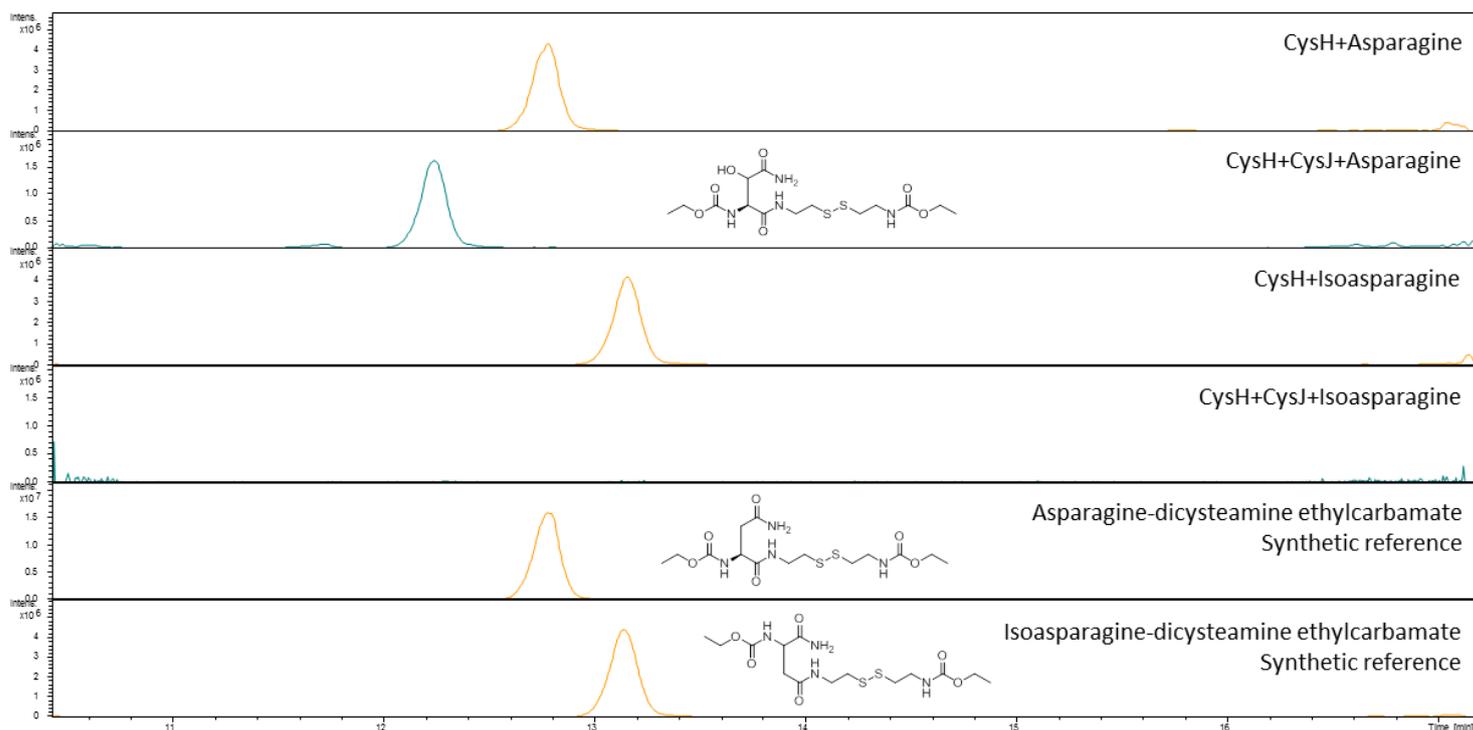


Figure 4.9: Observed derivatives unloaded from CysH upon addition of asparagine or isoasparagine in the presence or absence of CysJ

Furthermore the stereochemistry of the reaction remains elusive and the synthesis of erythro  $\beta$ -hydroxy-L-asparagine and threo  $\beta$ -hydroxy-L-asparagine references would be necessary to confirm without a doubt the exact position of the hydroxyl. Given the orientation on the final Cystobactamide 919-2 product it is likely to yield the threo (S,R)  $\beta$ -hydroxy-L-asparagine diastereoisomer.

### 3.4 Protein MS

Unexpectedly analysis of the reaction mixture through protein MS does not allow a clear identification of the hydroxylation (Figure 4.10 - top) but only what seems to be a partial turn over upon incubation of loaded CysH with CysJ and  $\alpha$ -ketoglutarate as seen by the mass shift of the protein peak, but the quality of the spectrum decreases dramatically which does not allow a conclusive evidence for the hydroxylation and hints at a splitting of the protein population into various derivatives with different masses this might be linked to the interplay between the activities of CysJ and CysH and could imply that reaction intermediates are also observed. Furthermore upon the longer incubation times used for the hydroxylation, the mass of asparagine loaded CysH shifts towards a mass corresponding to a dehydrated product instead of the expected +114 m/z shift corresponding to asparagine which is observed upon shorter incubation periods. This phenomenon could also be related to the activity of the X domain featured by CysH.

## CysH and CysJ - In cis and in trans tailoring working hand in hand for hydroxy-isoasparagine biosynthesis

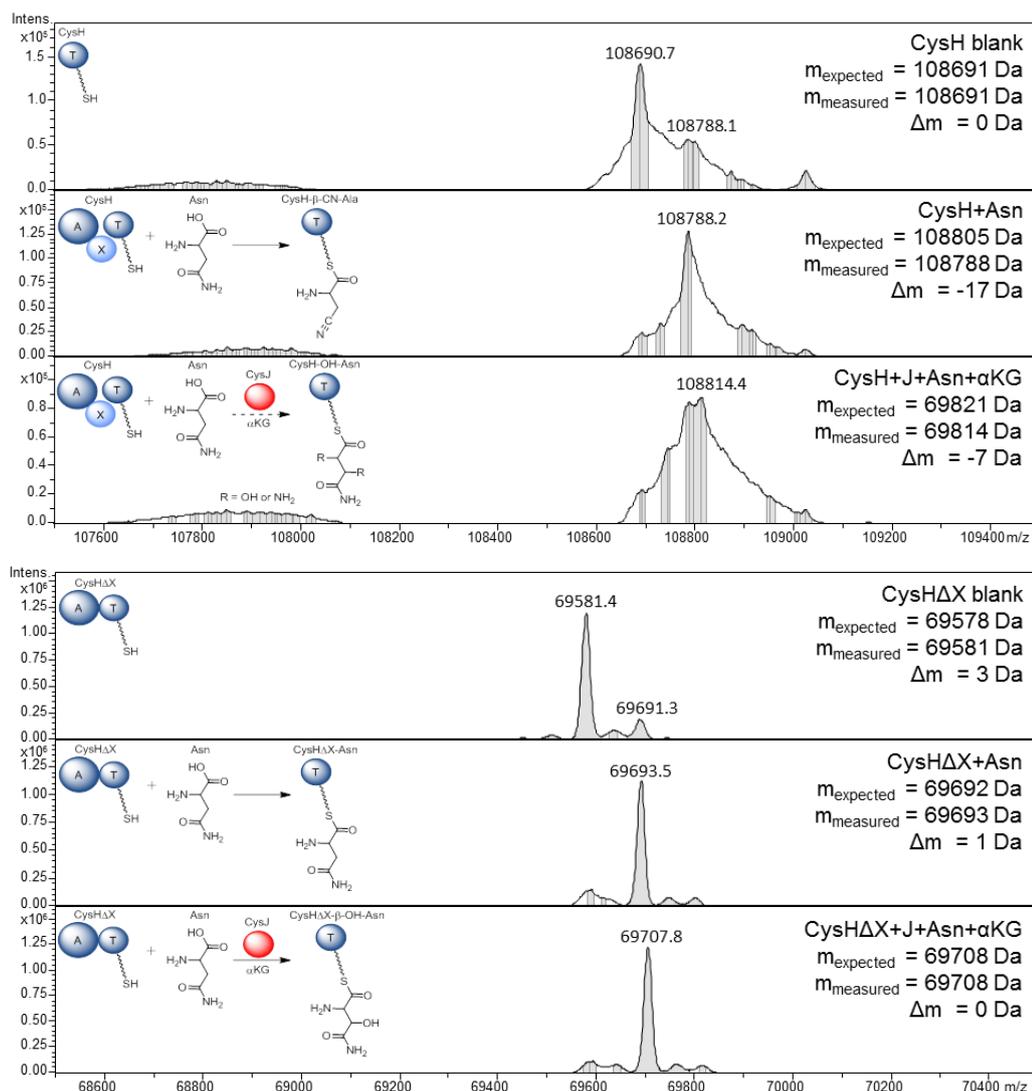


Figure 4.10: Protein MS analysis of CysH (top) and CysH $\Delta$ X (bottom) upon addition of CysJ and  $\alpha$ -ketoglutarate.

To mitigate these effects which probably originate in the activity of the X domain another construct of CysH was generated with a deleted X domain which allowed us to obtain a much better quality in the MS analysis, (Figure 4.10 - bottom) in this case the loading of asparagine was evident with no dehydration, implying that the dehydration process is indeed enzymatical and linked to the X domain. Furthermore a mass shift of +16 m/z is observed upon incubation of CysJ conclusively proving its role as *in trans* asparaginyl hydroxylase. The reaction also showed to be dependent on the presence of  $\alpha$ -ketoglutarate which is in accordance with the *in silico* activity prediction of CysJ as  $\alpha$ -ketoglutarate dependent dioxygenase.

Since protein MS is highly dependent on the homogeneity of the sample regarding derivatives of the protein analyte with small mass differences, the sharpness of the peaks in the second experiment in the absence of X domain indicates that the loading and hydroxylation reaction proceed in high turnover yielding a relatively homogenous population

of product, as does the dehydration reaction by the X domain. However in the presence of both CysJ and the X domain in CysH, the broadening of the signals indicates that a parallel reaction, probably the isomerization, is happening only in the presence of both X and CysJ which provides low turnover and yields a mixture of substrate, product and maybe even intermediates that cannot be resolved by the deconvolution algorithm.

### 3.5 CysH - *In silico* analysis

CysH is a 108 kDa NRPS module with  $A_{\text{core}}\text{-X-A}_{\text{sub}}\text{-T}$  architecture featuring unprecedented in cis tailoring domain of 38 kDa. While it seems to be remotely related to transamidases no conserved domains can be found on primary or secondary sequence level. While this domain is thought to be responsible for an aminomutation type reaction (Figure 4.16) there is no ASG motif that could lead to the 4-methylideneimidazol-5-one cofactor ruling out any homology with TAM and PAM, in addition no CxxCxxx motif as Fe-S cluster binding site can be made out indicating that it is probably not using a radical mode of action, even if a few radical SAM proteins have been characterized that do not harbor this exact motif.(30)

A few homologues of the X domain can be found in the NCBI database, they are all integrated in adenylation domains featuring the same Stachelhaus code specificity (asparagine) the full size proteins can however be quite different depending on whether they harbor a condensation domain or a thioesterase domain and can even be multimodular. Unfortunately however, none of the clusters where these X domains could be found was linked to a known secondary metabolite. *In silico* analysis of the 25 closest homologues show very high sequence conservation (25%) in protein of very diverse origin, this lead to the identification of seven conserved core motives (Figure 4.11Figure 4.12) that might be required for catalysis or folding.



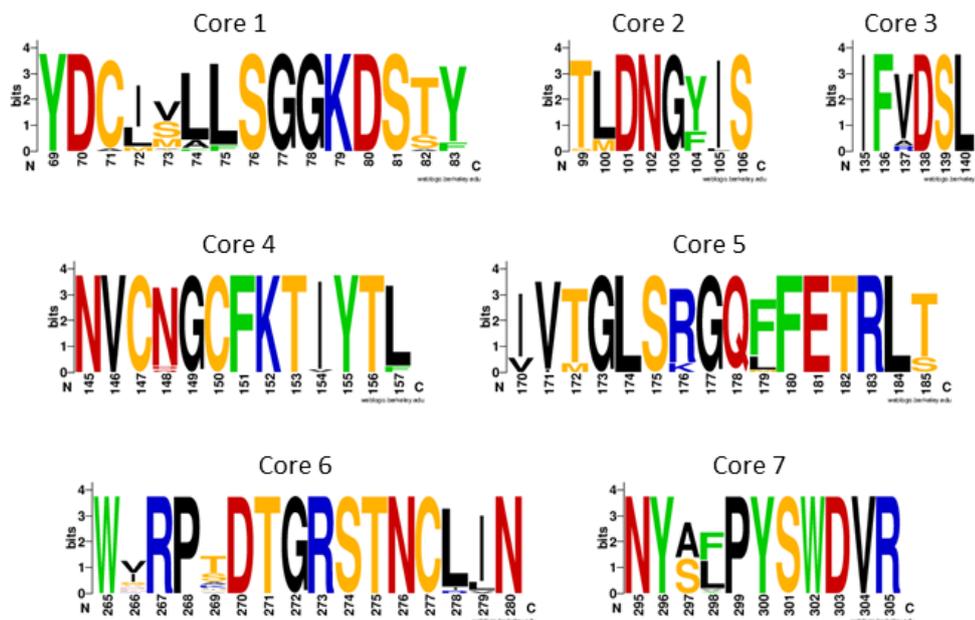


Figure 4.12: conserved core motifs in the X domain

### 3.6 Labeled Asparagine feeding

The origin of the  $\alpha$ -hydroxy- $\beta$ -asparagine moiety in Cystobactamides was confirmed through feeding of the *Cystobacter velatus* Cbv34 with  $^{15}\text{N}_2$   $^{13}\text{C}_4$  labeled asparagine. A mass shift of +6 m/z was observed indicating a full conservation of all carbons and nitrogens from asparagine in  $\beta$ -asparagine. This confirms that the Isoasparagine arises from asparagine through internal rearrangement rather than transamidation reactions with an aspartate intermediate indicating that the  $\beta$ -amino group was transferred from the  $\alpha$ -position by an aminomutase type reaction.

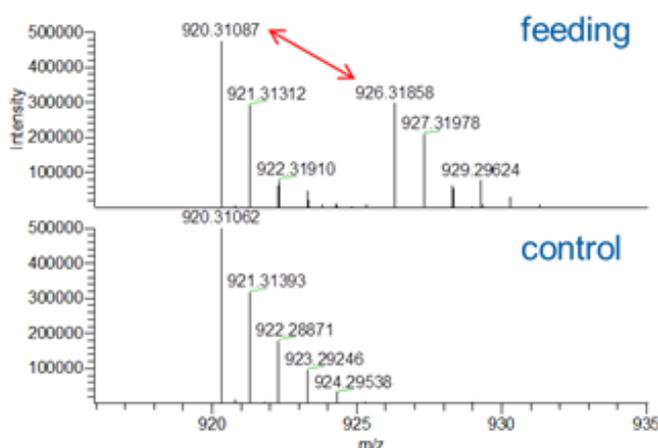


Figure 4.13: 4-isotope pattern of Cystobactamide 919-1 observed in MS upon feeding of  $^{15}\text{N}_2$   $^{13}\text{C}_4$  labeled asparagine. A  $\Delta m/z$  of 6,00771 is measured corresponding to a full incorporation

### 3.7 Dehydration of asparagine

It is noteworthy that while Cystobactamides and Albicidins feature different linker units,  $\alpha$ -hydroxy- $\beta$ -asparagine for the former and  $\beta$ -cyano-alanine for the latter, the domain responsible for this biosynthesis is highly similar between both assembly lines (83,2% similarity / 68,5% identity). It is thus unlikely that the underlying biochemistry would be highly different and most likely a slight modification of the substrate would lead to the two radically different products. We could indeed observe already in the loading assays traces of an apparent dehydration product which can be driven to completion with slightly longer incubation times, indicating that CysH despite yielding almost only  $\alpha$ -hydroxy- $\beta$ -asparagine as product *in vivo* can mimic the activity of AlbIV and produce  $\beta$ -cyano-alanine *in vitro*.

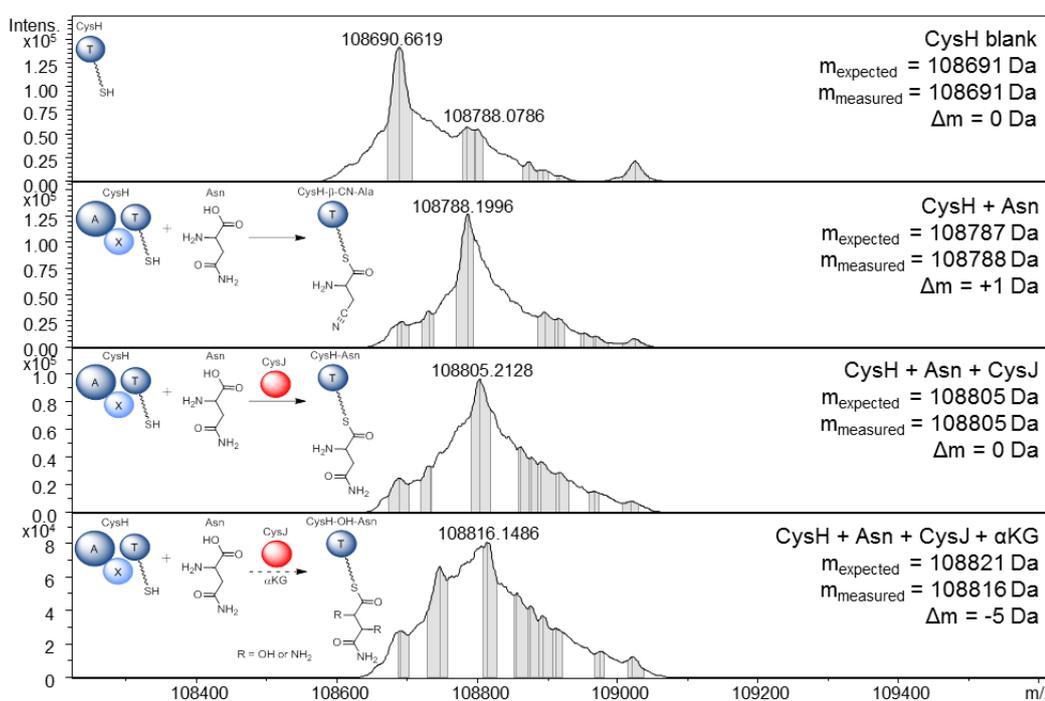


Figure 4.14: Protein MS analysis of CysH upon addition of CysJ and  $\alpha$ -ketoglutarate showing the dehydration activity and its abolishment upon addition of CysJ

Interestingly this dehydration is not observed anymore upon addition of CysJ even in the absence of  $\alpha$ -ketoglutarate and thus of hydroxylation activity, indicating that the simple binding of the PCP-asparaginyl substrate to CysJ is sufficient to prevent any meaningful dehydration turnover from happening. (Figure 4.14) As already shown previously the deletion of the X domain from CysH has the same effect of preventing the dehydration (Figure 4.10) which clearly links this activity to this *in cis* tailoring domain. However, even if the hydroxylation activity of CysJ has been proven upon addition of  $\alpha$ -ketoglutarate, the competing activities of CysJ and CysH do not allow obtaining spectra showing clearly the +16 mass shift expected for a hydroxylation but rather a +11 shift. This is most likely linked to a splitting of the protein population due to the activity of CysH in conjunction with CysJ. Since the homogeneity of the protein analyte is crucial in protein MS especially for large

proteins like CysH, a heterogeneous population of large proteins with small mass differences cannot be resolved and leads after deconvolution to lower intensity and broader peaks with a mass averaging the different protein present in solution like we observe in this case. These variations in the MS pattern upon hydroxylation of the loaded asparagine indicate nonetheless that the presence of the  $\beta$ -hydroxyl installed by this dioxygenase might be the factor that determines the final product of the linker biosynthesis pathway. Indeed CysJ has no equivalent in the Albicidin cluster which could explain why the Albicidins feature a  $\beta$ -cyano-alanine linker unit while Cystobactamides present  $\alpha$ -hydroxy- $\beta$ -asparagine.

### 3.8 CysH - Substrate screening

Additionally to the natural substrates shown in Figure 4.6 two non-natural substrates were assayed for loading on CysH *in vitro*. (Figure 4.15) D-asparagine and  $\beta$ -cyano-alanine were tested because of their structural similarity with the natural substrate asparagine and for the implications they could have in terms of biosynthesis engineering. Additionally, very short reaction times were used to prevent the dehydration from happening. Interestingly CysH does not discriminate for the stereochemistry of the substrate since D-Asparagine was activated and loaded, but it seems to be a much better substrate for the X domain since the observed mass features a shift of -18 m/z corresponding to the dehydrated amino acid to  $\beta$ -cyano-D-alanine despite the fact that incubation times of about 5 minutes were used in this assay. While dehydration is almost complete for D-asparagine only traces of this activity can be observed for the other loaded substrates. It is not clear whether D-asparagine can be used as substrate for CysJ and for the subsequent isomerization but it could imply that a reversal of the stereochemistry of certain Cystobactamides might be achieved biosynthetically.  $\beta$ -cyano-alanine which is the product of the dehydration was only partially loaded which indicates the *in cis* origin of this derivative since a precursor tailoring would lead to a decrease in the efficiency of the biosynthesis.

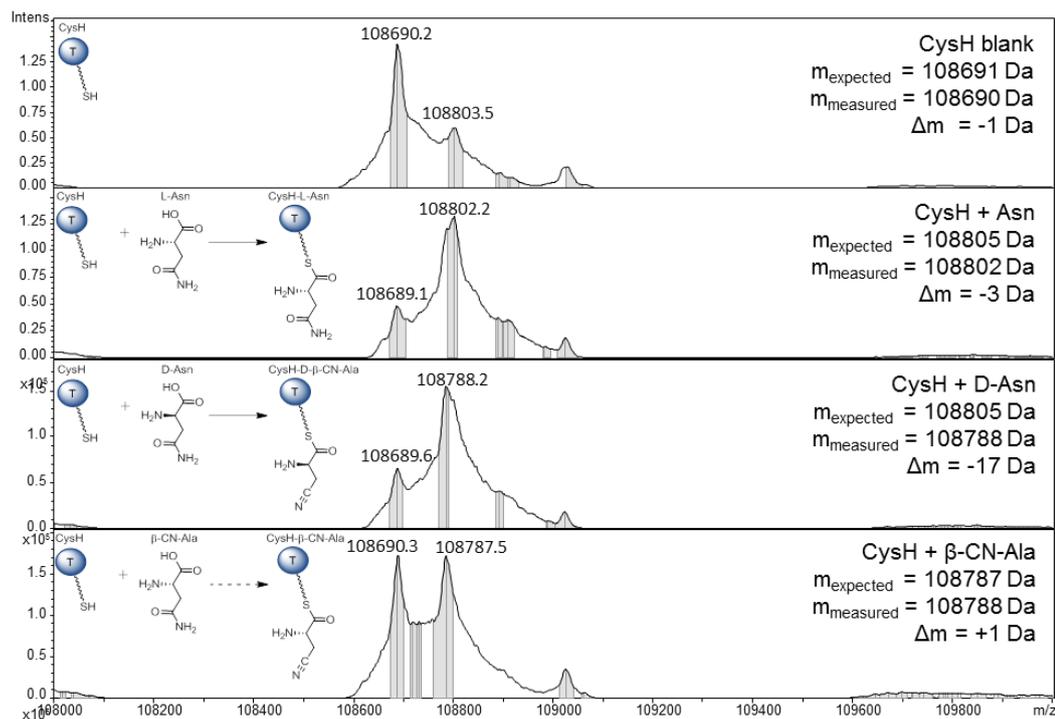


Figure 4.15: Substrate selectivity for the thiolation activity of CysH for D-asparagine and  $\beta$ -cyanoalanine analyzed through full size protein MS after incubation with amino acids. For the sake of simplicity only the T domain of Module CysH A-X-T was depicted.

### 3.9 Mode of Action

Experimental data show that CysH is able to catalyze what seems to be two completely different reactions depending on the substrate that is presented to the X domain. On the one hand, in the presence of an asparagine- $\beta$ -hydroxylase it behaves like an aminomutase performing an isomerization reaction through internal migration of the amine function. On the other hand, in the absence of any other tailoring enzymes it performs a dehydration of asparagine. Both these reactions are unprecedented and hardly compatible regarding the already described modes of action of enzymes showing similar activities. It is however very likely that CysH does not proceed through these mechanisms since it has no homology to either aminomutases or dehydratases.

#### 3.9.1 Aminomutases

While  $\beta$  amino functions in natural products are common, they mostly arise on complex scaffolds and are not linked to natural  $\alpha$  amino acids.<sup>(31)</sup> The presence of a  $\beta$  variant of natural amino acids in NRPS has been reported for four of them only. The most widespread is  $\beta$ -alanine which can be found in a number of natural products such as Theonellamides<sup>(32)</sup> and Destruxins<sup>(33)</sup>, but also in primary metabolism (e.g pantothenate-B5), its biosynthesis however, does not originate in  $\alpha$ -alanine but through  $\alpha$ -decarboxylation of aspartate.<sup>(34)</sup> On the other hand the three less common ones,  $\beta$ -lysine,  $\beta$ -tyrosine and  $\beta$ -phenylalanine, which are found in a few different natural products (e.g Chondramides<sup>(35)</sup>,

Capreomycins (36), Cyclochlorotines (37) respectively), arise from their respective  $\alpha$ -amino acid which are isomerized in free form prior to activation by adenylation domains specific for  $\beta$ -amino acids. This biosynthesis is performed by amino mutase enzymes belonging to two distinct classes.(38)

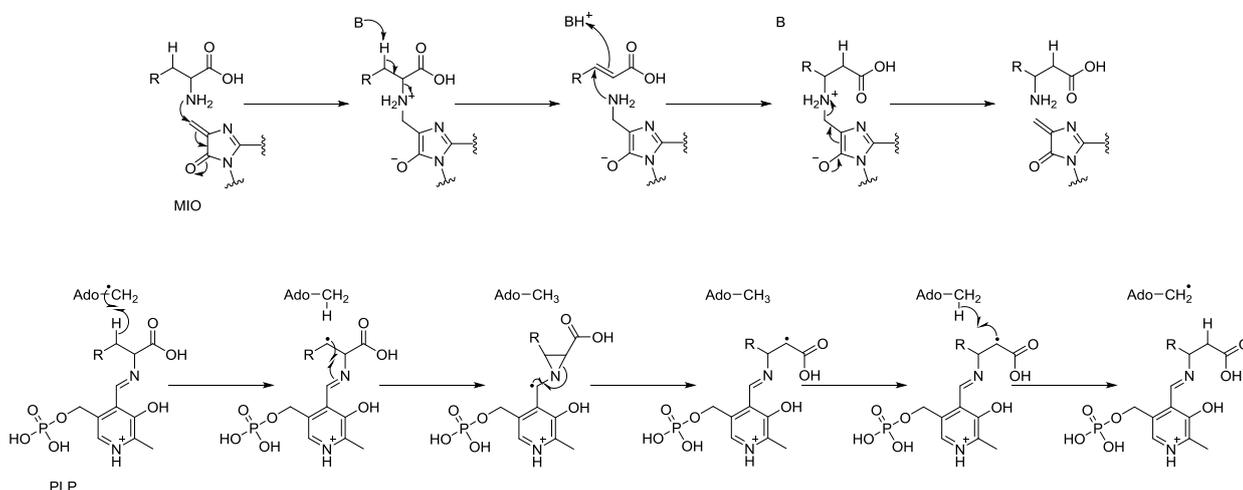


Figure 4.16: 4-methylideneimidazol-5-one (MIO) and Pyridoxal phosphate (PLP) based aminomutases reaction mechanisms

The first amino mutase class encompasses phenylalanine 2,3-aminomutases (PAM) and tyrosine 2,3-aminomutases (TAM) they use a 4-methylideneimidazol-5-one (MIO) prosthetic group in the rearrangement reaction which arises during automodifying maturation of an Ala-Ser-Gly tripeptide motif in the inactive proenzyme.(39) The MIO cofactor undergoes nucleophilic attack by the  $\alpha$ -amino group forming an alkylated aminocomplex, the  $C\alpha$ -N bond cleavage is catalyzed by a nearby base forming a transient  $\alpha,\beta$ -unsaturated intermediate. The amino group is finally transferred back to the olefin by attack on the  $C\beta$  rather than  $C\alpha$  regenerating the MIO group in the process. The second class of amino mutase are radical SAM enzymes containing a 4Fe-4S cluster and a PLP binding motif, the type enzyme being lysine- 2,3-aminomutase (LAM).(40) In this case an  $\alpha$ -amino acid-PLP aldimine is formed before homolytic  $C\beta$ -H bond cleavage by dAdenosine radical, the resulting radical intermediate forms a three-membered aziridine ring allowing the rearrangement of the  $\alpha$ -amino group to the  $\beta$  position, ultimately the radical in position  $\alpha$  can abstract the hydrogen atom back from dAdenosine. (

Figure 4.16) It is noteworthy that a third class of aminomutases exists, but rather than transferring the  $\alpha$  amino function to  $\beta$  position it transfers the terminal amine of lysine from  $\epsilon$  to  $\delta$  position or of ornithine from  $\delta$  to  $\gamma$ .(41) These enzymes are mechanistically similar to the radical SAM aminomutases, they also possess a PLP binding pocket and an 4Fe-4S cluster bound in a TIM barrel but they feature a B12 binding domain and use adenosylcobalamine to generate the deoxyadenosine radical donor. However all these structural features present in every one of the relatively large radical aminomutases are absent from the 37 kDa X domain, it is thus quite unlikely that the X domain would follow the same biochemistry as radical aminomutases.

### 3.9.2 Dehydratases

Dehydratase reactions are mainly known for hydroxyl functions and two modes of action are described but both are based on the abstraction of the  $\beta$  Hydrogen either through a catalytic triad (42) or by a radical mechanism involving a 4Fe-4S cluster and SAM (43), both these reactions then proceed by internal rearrangements involving neighboring functions such as the  $\beta$ -hydroxyl of glycerol (44) or the  $\gamma$ -ketone of isoflavanone.(45) Interestingly while nitrile biosynthesis is usually performed by a dehydratase too, it does not originate in amides but in aldoximes and proceeds through a heme dependent aldoxime dehydratase.(46) Furthermore while dehydration of amides to nitriles was not known to this date, the reverse reaction catalyzed by a heterodimeric nitrile hydratase has been described and relies on a single, cysteine bound iron atom.(47) Amongst all the enzymes described in this paragraph, this is the only one which shares a certain, although low, degree of similarity with the X domain. It is thus possible that the catalysis is performed by a hydratase whose kinetic has been reversed and which performs a dehydration.

### 3.9.3 Mechanism proposal

Even if no *in vitro* data to this date can prove the biochemistry of X domain in regard of the isomerization, *in vivo* experiments with an X domain deletion mutant of the heterologous host harboring the Cystobactamide cluster leads to complete abolishment of the production of Cystobactamides 919-1 and 920-1 both harboring the isoasparagine linker (unpublished results), indicating that this domain is indeed involved in the isomerization of the linker. While the X domain seems to be able to catalyze two radically different reactions leading to either  $\beta$ -cyano-alanine through dehydration of the terminal amide of asparagine or to the isomerization of  $\beta$ -hydroxy-asparagine in  $\beta$ -amino- $\alpha$ -hydroxy-asparagine, both these reactions could theoretically proceed through a similar mode of action involving the abstraction of an  $\text{OH}^-$  ion from the substrate molecule. The central difference lies in the position of the abstracted oxygen atom, since the isomerization takes precedence over the dehydration when both are possible, it is safe to assume that the catalysis seems to be preferentially directed towards the hydroxyl but can be redirected towards the amide if the hydroxyl is absent.

This leads to the following hypothesis regarding the mode of action of the X domain, the rusty brown color of the protein in addition to the presence of three conserved cysteine in the core motifs of all X domains indicate that the activity could rely on a catalytic metal center. It is however not yet clear whether an iron is used or another metal center or maybe even a 2Fe-2S cluster; nevertheless such a catalytic center could help the abstraction of  $\text{OH}^-$  from the substrate. In the absence of a hydroxyl the putative iron center would be able to bind the oxygen from the amide function, which would lead upon attack of a base on one of the hydrogens from the amide to the abstraction of  $\text{OH}^-$  yielding effectively a dehydration of the amide into a nitrile. In the case where an hydroxyl was installed in  $\beta$  position this would be the preferred center for iron interaction, leading upon attack of the  $\alpha$ -amine on the  $\beta$ -carbon to an elimination of the hydroxyl with aziridine ring closure. The reopening of such a ring by

water with the help of the same catalytic base would thus lead to  $\beta$ -amino- $\alpha$ -hydroxy-asparagine, effectively producing an isomerization of the substrate.

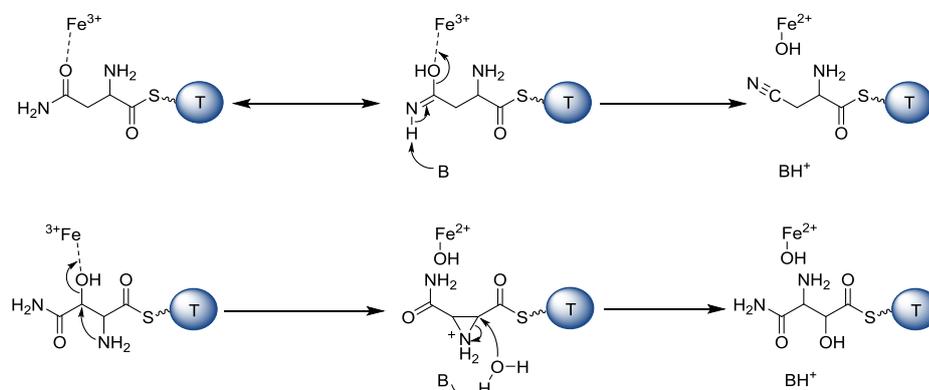


Figure 4.17: Proposed mode of action of the bifunctional asparaginyl dehydratase /  $\beta$ -hydroxy-asparaginyl aminomutase

### 3.9.4 Stereochemistry of the isomerisation

While the stereochemistry of Cystobactamide 919-2 has now been confirmed by synthetic homologues to be (S,R) (48), the absolute configuration of 919-1 is still not completely clarified. The original assignment was made from NMR coupling constants between  $\alpha$  and  $\beta$  hydrogens in combination with ROE coupling of the asparagine or isoasparagine linkers which were misinterpreted as being in “threo” configuration.(10) Later upon re-isolation of Cystobactamide 919-2 “erythro” stereochemistry has been proposed (49) and the absolute configuration (S,R) was later confirmed with synthetic homologues.

Since Cystobactamide 919-2 can be considered as an intermediate leading to 919-1 through one further tailoring step it would be logical that their stereochemistry have to be consistent with one another. In order to lead to the confirmed (S,R) stereochemistry for Cystobactamide 919-2 the pathway would have to start from L-asparagine (S) processed into (S,R)  $\beta$ -hydroxy-asparagine by CysJ. According to the mode of action hypothesis proposed this intermediate can only be isomerized in (R,S)  $\beta$ -amino- $\alpha$ -hydroxy-asparagine yielding an S,R absolute configuration in the final Cystobactamide 919-1 which is contrary to original annotation but consistent with the corrected annotation of Cystobactamide 919-2.

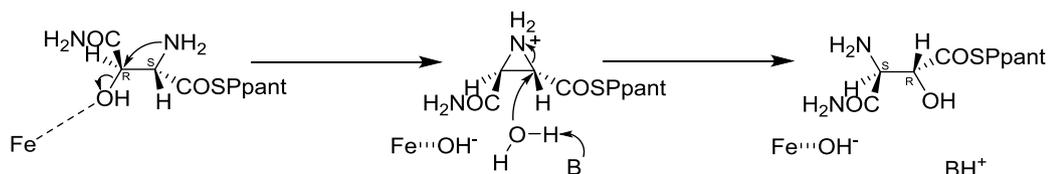


Figure 4.18: Proposed stereochemistry of the isomerization reaction

## 4 Conclusion

We could elucidate the biosynthesis of both the  $\beta$ -hydroxy- $\alpha$ -asparaginyl linker observed in Cystobactamide 919-2 as well as the equivalent  $\beta$ -cyano-alanine linker present in the closely related Albicidin. Furthermore we could propose a biosynthesis for the unique  $\alpha$ -hydroxy- $\beta$ -asparaginyl observed in Cystobactamide 919-1. The biosynthesis of these three linkers is the result of the interplay between a newly discovered NRPS *in cis* tailoring domain featuring a unique bifunctional asparaginyl dehydratase /  $\beta$ -hydroxy-asparaginyl mutase activity, and an *in trans* asparaginyl- $\beta$ -hydroxylase, CysJ.

We confirmed the role of CysJ in the biosynthesis as asparaginyl-PCP  $\beta$  hydroxylase by *in vitro* investigations, these results provide a further example of *in trans*  $\beta$  hydroxylation and strengthen the fact that this type of tailoring reactions tends to be catalyzed *in trans* on standalone A-T domains. The stereochemistry of the reaction is still elusive and would require further investigations to be confirmed. But since the final conformation in Cystobactamide 919-2 was confirmed to be S,R with synthetic analogues, it is thus likely that the hydroxyl is added in erythro conformation (S,R). We could also prove experimentally that hydroxylation is the first tailoring step leading to the  $\beta$ -hydroxy-isoasparagine moiety observed in Cystobactamide 919-1.

The bifunctionality of the X domain of CysH seems to be linked to the nature of the substrate bound to the PCP, with the presence or the absence of a  $\beta$ -hydroxyl on the asparagine being the determining factor. While the X domain is herein described for the first time a few homologues can be found in the NCBI database suggesting that  $\alpha$ -hydroxy- $\beta$ -asparagine or  $\beta$ -cyanoalanine moieties could be produced by a number of silent or unexplored gene clusters, the high level of conservation amongst the sequences allowed to assign a number of core domains which could be implicated in the catalysis and the folding of the domain. However further structural and biochemical investigation of this new domain would be needed to confirm or disprove the mode of action we proposed.

## Chapter 4.2: CysB - A new amino acyl transferase shuttles modified asparagine residues between a standalone NRPS module and the assembly line

### 1 Introduction

The previous subchapter described the processing of asparagine into  $\beta$ -hydroxy isoasparagine carried out in two tailoring steps: the *in trans*  $\beta$ -hydroxylation by CysJ and the *in cis* isomerization by CysH. This independent biosynthetic pathway is performed on CysH, a standalone A-X-T module separate from the assembly line. Since CysH lacks a condensation domain its bound substrate is unlikely to act as acceptor, and since as module 3' it activates and modifies the third acid rather than the first one it is also unlikely to act as donor if it does not carry the peptide chain due to the processivity of condensation reactions.(50) This particular pathway organization implies that the modified amino acid has to be transferred from the independent module in order to reach the assembly line for integration in the peptidic backbone. In the few cases where such processes were described, like for the Nikkomycin (26), Novobiocin (25) or Vancomycin (13) biosyntheses, this activity is carried out by a thioesterase releasing the modified amino acid which is later tethered to the natural product core by a specific ligase or reactivated by a specific adenylation domain to be integrated in the peptidic backbone. However, in the case of Cystobactamides the central position of the hydroxy-isoasparagine moiety in the peptide and the lack of an active adenylation domain at this point of the assembly line implied that an alternative mechanism has to be at play. According to *in silico* analysis of the cluster which revealed the homology of CysB with SyrC, a rare amino acyl transferase type protein, it seemed reasonable to assume that CysB would be able to shuttle the amino acyl moiety between the PCP domains of CysH and CysK<sub>3</sub>.

This phenomenon has been observed only twice before; in the Syrigomycin and in the Coronatine biosyntheses were CysB homologues, SyrC (51) and CmaE (52), are both coupled with a halogenating  $\alpha$ -KG Fe<sup>II</sup> dioxygenase, SyrB2 and CmaB respectively, which are similar to CysJ. From the two, SyrC is the closest homologue to CysB but it displays only 35%/20% sequence similarity/identity. However, these proteins use the ubiquitous  $\alpha/\beta$  hydrolase fold which is known to display high structural similarity although they can have very low sequence identity.(53) Upon alignment the catalytic cysteine 128 as well as the helper histidine 286 were found to be conserved in all three proteins. The cysteine serves as transient binding site for the amino acyl moiety which is accepted as a thioester.

## CysB - A new amino acyl transferase shuttles modified asparagine residues between a standalone NRPS module and the assembly line

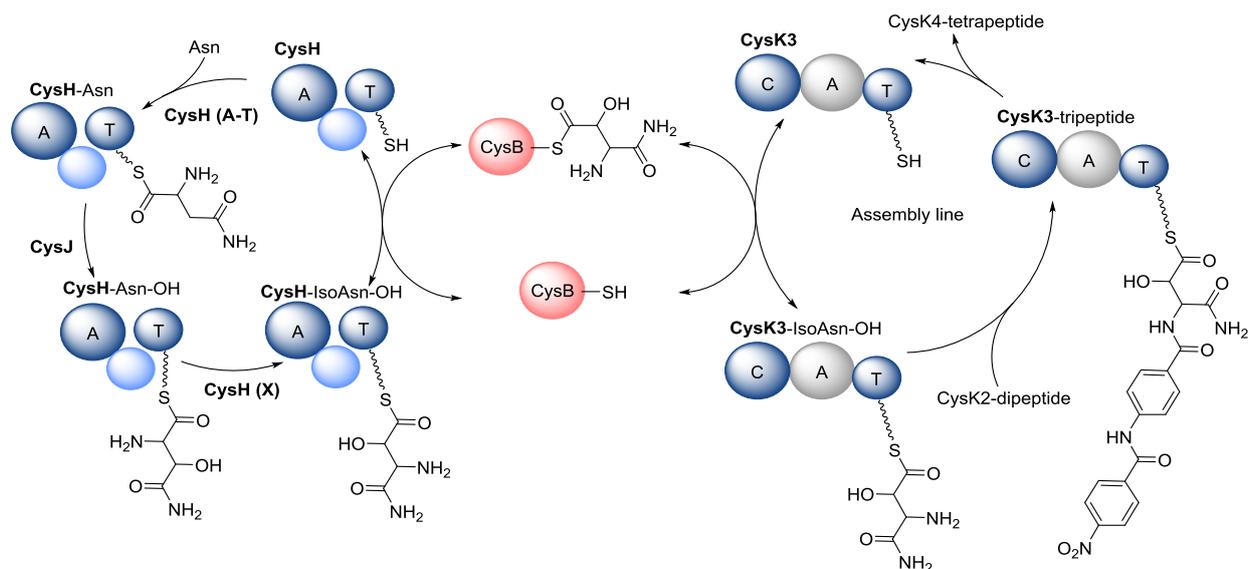


Figure 4.19: CysB as an example for an acyltransferase ping-pong mechanism

This type of acyl transferases is closely related to cysteine hydrolases (54) but features cysteine thioester intermediate stable enough not to be hydrolyzed immediately. This allows them to perform a reversible trans thioesterification reactions leading to a so called ping-pong mechanism where the same reaction is catalyzed in different directions on two carrier proteins. (Figure 4.19) This reaction cycle is probably completely reversible but the dynamic equilibrium is displaced towards a shuttling from CysH to CysK since both are involved in monodirectional reaction cycles.

## 2 Material and Methods

### 2.1 Cloning, overexpression and purification

Standard protocols were used for DNA cloning, transformation, amplification, and purification. Genomic DNA was extracted from *Cystobacter velatus* Cbv34 grown in CTT medium

The DNA fragment encoding CysB was amplified using the forward primer 5'TATCCATGGGTACGCCAGCAGCAGG and the reverse primer 5'TATGGATCCCTACGCTCCCGCCACCGCAG. The amplified DNA fragments were digested with NcoI and BamHI, and cloned into pHis-SUMO-TEV with an N-terminal His6 tag.

The 213 bp DNA fragment encoding CysA was amplified using the forward primer 5'TATCATATGAGCATGAACGGGGACG and the reverse primer 5'TATAGATCTTCAGCAGTTGCTGCGCG. The amplified DNA fragments were digested with NdeI and BglIII and cloned into the second MCS of pETduet-1 in native form.

The DNA fragment encoding CysK3 was amplified using the forward primer 5'TATGGATCCGAAAACCTG-TATTTTCAGAACACCGAGGCGGTGCT containing the ENLYFQG TEV protease recognition site and the reverse primer 5'TATAAGCTTTCACCG-ATGGATCGACGACAC. The amplified DNA fragment was digested with BanHI / HindIII and cloned into the first MCS of pETduet-1 to yield N-terminal His6 tag and TEV protease site fusion constructs. The resulting constructs were sequenced to verify that no mutation had been introduced during PCR.

Recombinant protein expression was carried out in *E.coli* BL21 (DE3) grown in LB medium supplemented with 50 µg/ml of kanamycin. The culture was inoculated 1/10 from a fully grown overnight culture and was initially grown at 37°C until it reached an OD of 0.6 to 0.9, at which point the temperature was decreased to 16°C. Induction was initiated after 30 minutes at 16°C with 0,1 mM isopropyl-β,D-thiogalactopyranoside (IPTG). The cells were harvested 16h after induction by centrifugation at 8000 g for 10 minutes at 4°C and resuspended in lysis buffer (25 mM TRIS pH 7,5; 150 mM NaCl; 20 mM imidazole). The cells were lysed by passage at 1500 bar at 4°C through a M-110P Microfluidizer (Microfluidics) after which the crude extract was centrifuged at 45000 g for 15 minutes at 4°C, the supernatant was treated as soluble protein fraction.

Affinity chromatography was performed on an Äkta Avant system while size exclusion chromatography was performed on an Äkta Pure system. The soluble fraction of the lysed cells was applied to 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10 CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole.

The pooled fractions were applied to a HiPrep 26/10 desalting column equilibrated in running buffer (25 mM TRIS pH 7,5; 150 mM NaCl; 2 mM DTT). The resulting fractions were pooled (≈ 30 ml) and incubated overnight at 4°C with TEV protease (1 mg/20 mg protein). After ≈ 16 h incubation 20 mM imidazole was added to the solution prior to loading on a 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10 CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole. The pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200 pg SEC column equilibrated in running buffer (25 mM TRIS pH7,5; 150 mM NaCl; 2 mM DTT). The pooled fractions were controlled via SDS-PAGE, concentrated and frozen at -80°C after addition of 25% glycerol.

## 2.2 Activity assays

### 2.2.1 Ping - loading from H to B

In a total volume of 50 µl, 1 µM CysH was incubated with 1 mM Asparagine, 1 mM ATP in 25 mM TRIS pH 7,5; 150 mM NaCl; 10 mM MgCl<sub>2</sub>, finally 5 µM CysB was added

and the reaction was incubated for 1 h at room temperature. The solution was analyzed by protein LC-MS.

### 2.2.2 Pong - loading from B to K3

In a total volume of 50  $\mu$ l, 5  $\mu$ M CysK3 and 1  $\mu$ M CysH were incubated with 1 mM Asparagine, 1 mM ATP in 25 mM TRIS pH 7.5; 150 mM NaCl; 10 mM MgCl<sub>2</sub>, finally 500 nM CysB was added and the reaction was incubated for 1 h at room temperature. The solution was analyzed by protein LC-MS or the substrate was unloaded and derivatized for independent LC-MS analysis.

## 2.3 Protein MS

CysH was analyzed on a Dionex Ultimate 3000 UPLC system (Thermo Scientific) coupled with an maXis4G Q-TOF MS (Bruker) using an ESI in positive mode. The samples were run on an Aeris Widepore XB-C8, 150 x 2.1 mm, 3.6  $\mu$ m dp column (Phenomenex, USA). LC conditions: A-Water + 0.1% FA; B-Acetonitrile + 0.1 % FA at a flow rate of 300  $\mu$ l/min and 45°C. 0 min: 98 % A / 2 % B, 0.5 min: 98 % A / 2 % B, 10.5 min: 25 % A / 75 % B, 13.5 min: 25 % A / 75 % B, 14 min: 98 % A / 2 % B. The LC flow was split to 75  $\mu$ l/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany) using the standard Bruker ESI source. In the source region, the temperature was set to 180°C, the capillary voltage was 4000 V, the dry-gas flow was 6.0 l/min and the nebulizer was set to 1.1 bar. Mass spectra were acquired in positive ionization mode ranging from 150 – 2500 m/z at 2.5 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm (Copyright 1991-2004 Spectrum Square Associates, Inc.).

## 2.4 Unloading

After the reaction is complete 10 mM cysteamine are added from a newly prepared cysteamine 1M stock solution and the resulting solution is incubated at 30°C for 1h with slow shaking. The reaction mix is transferred in a 2 ml glass vial and 45  $\mu$ l ethanol:pyridine (4:1) solution + 5  $\mu$ l ethyl chloroformiate (ECF) are added and vortexed. After addition of 200  $\mu$ l water the derivatized NN-diethoxycarbonyl  $\beta$ -hydroxyasparaginyl dicysteamine is extracted with 2x300  $\mu$ l EtOAc+1 % ECF. The collected organic layers are dried, redissolved in MeOH and analyzed through LC-MS.

All measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 x 2.1 mm, 1.7  $\mu$ m dp column (Waters, Germany). Separation of 1  $\mu$ l sample was achieved by a linear gradient from (A) H<sub>2</sub>O + 0.1 % FA to (B) MeOH + 0.1 % FA at a flow rate of 600  $\mu$ l/min and 45°C. 0 min: 95 % A / 5 % B, 0.5 min: 95 % A / 5 % B, 1.5 min: 80 % A / 20 % B, 14.5 min: 70 % A / 30 % B, 17,5 min: 5 % A / 95 % B, 20,5 min: 5 % A / 95 % B, 21 min: 95 % A / 5 % B. UV spectra were recorded by a DAD in the range from 200

to 600 nm. MS data was acquired with an Amazon Speed 3D ion trap mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. Mass spectra were acquired in centroid mode ranging from 150 – 1500 m/z.

### 3 Results and Discussions

After *in silico* analysis and initial assignment of CysB as SyrC like shuttling protein, this hypothesis was tested by investigating the binding between CysH and CysB in solution by size exclusion chromatography. After overexpression and purification of both CysH and CysB they were incubated together and repurified by SEC to account for a possible coelution. (Figure 4.20) The presence of CysB in the CysH fractions confirmed the hypothesis that these proteins were forming a complex in solution which supported the initial annotation of CysB.

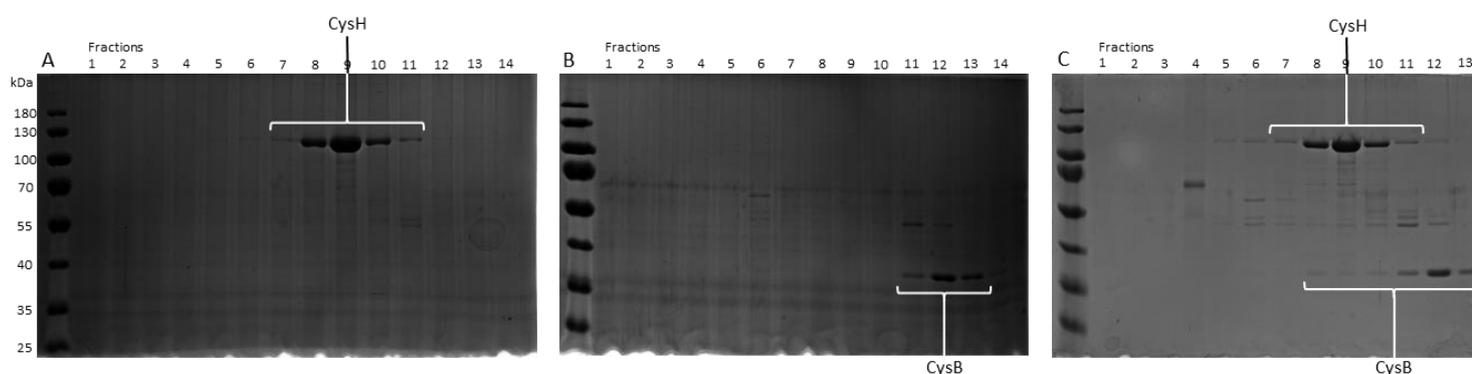


Figure 4.20: SDS-PAGE of the SEC fractions collected for A: CysH; B: CysB; C: an equimolar mixture of CysH and CysB

To investigate the shuttling mechanism both trans thioesterifications involved in the ping-pong mechanism were analysed mainly through protein MS. To this end the third module of the assembly line CysK3 was also cloned as a single module and overexpressed.

#### 3.1 Ping - loading from CysH to CysB

The first part of the two-step aminoacyl transfer reaction is the loading of the Ppant bound isoasparaginyl moiety from CysH to the active cysteine on CysB. Although  $\beta$ -hydroxy-isoasparagine and  $\beta$ -hydroxy-asparagine are the derivatives incorporated into the Cystobactamide scaffold, we used regular asparagine to test for CysB acyltransferase activity since it is readily available and does not require the addition of a supplementary tailoring protein to be produced. CysB was hence incubated with CysH in the presence of ATP and asparagine and analyzed through protein MS revealing a partial loading of CysB dependent on the presence of CysH which implies that CysB cannot load free asparagine but indeed acts as thioester amino acyl transferase. The incomplete loading of CysB results of a dynamic

## CysB - A new amino acyl transferase shuttles modified asparagine residues between a standalone NRPS module and the assembly line

equilibrium between CysH and CysB asparagine loaded intermediates and implies that the reaction is reversible.

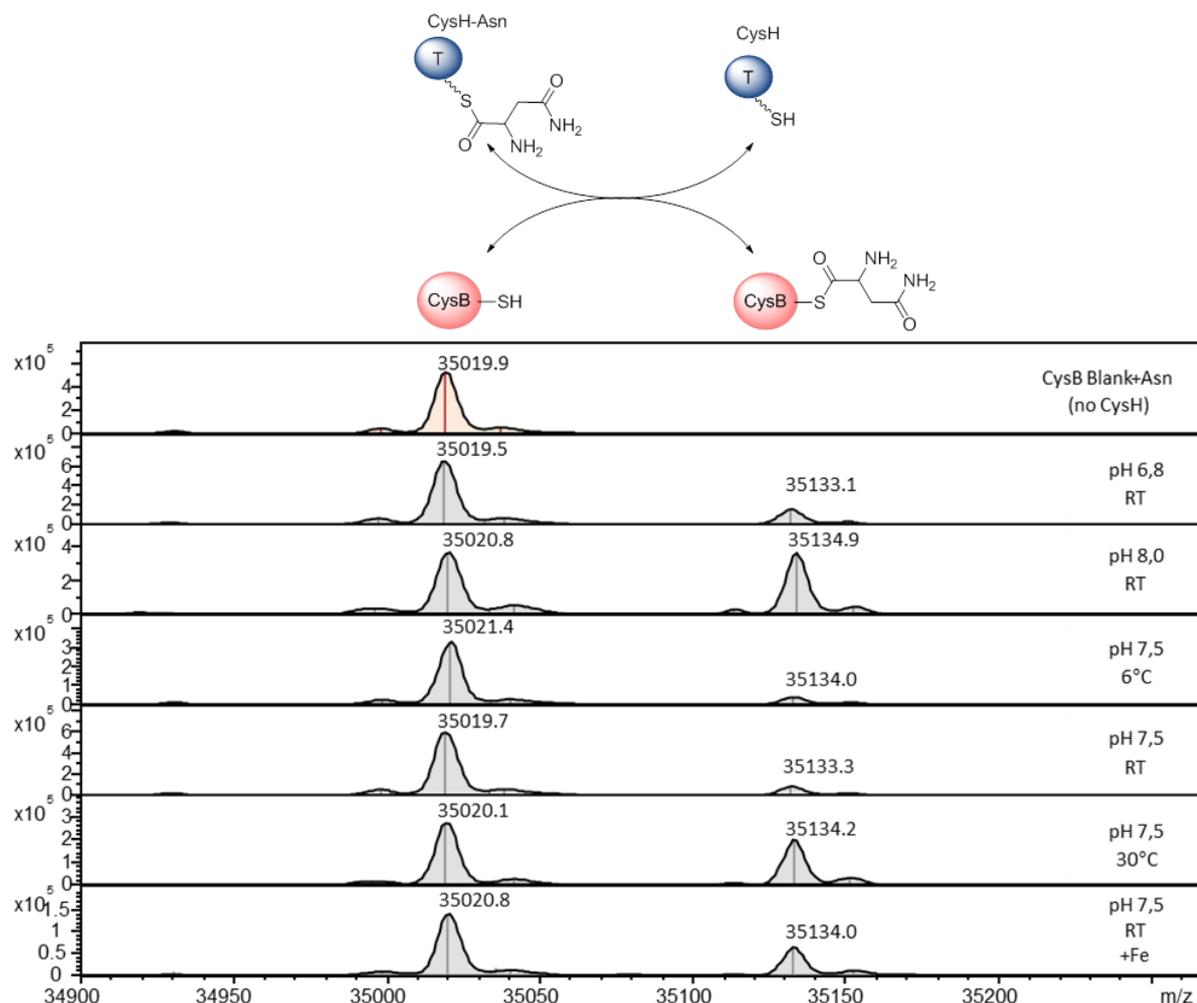


Figure 4.21: protein MS analysis of CysB incubated with CysH and asparagine with varying pH values and temperatures

Since this reaction results in a mixture of apo-CysB and asparagine bound holo-CysB the resulting ratio is very sensitive to the reaction conditions. In order to displace the equilibrium towards an improved loading of CysB the reaction was reproduced with varying reaction conditions, pH values and temperatures turned out to be the major factors influencing the reaction equilibrium. Additionally it is interesting to observe that CysB does not seem to be very substrate specific since asparagine can be taken over while no non-hydroxylated linkers are observed in Cystobactamide derivatives. It would thus be interesting to determine how promiscuous CysB can be and whether modifications of the biosynthesis pathway leading to  $\beta$ -amino- $\alpha$ -hydroxy-asparagine could yield modified Cystobactamide derivatives.

## 3.2 Pong – loading from CysB to CysK3

### 3.2.1 CysK3 overexpression

The second part of the mechanism of CysB involves further shuttling of the asparaginyl moiety onto the third module of the assembly line CysK3. To investigate this reaction *in vitro* this module had to be cloned and overexpressed. The large size of CysK at 507 kDa makes any overexpression attempt of the whole protein in *E.coli* unrealistic, the third module had thus to be cloned separately from the rest of the assembly line. The linker regions between T2 and C3 domains as well as between T3 and C4 were modeled using I-TASSER on the structure of the GrsA T-E didomain protein (PDB: 5ISX) to assign module borders and design fitting primers for the cloning of CysK3 as independent gene. (Further details are provided in Chapter 6) After initial overexpression attempts yielded almost no product, the MbtH type adenylation domain activator protein CysA was coexpressed in a duet vector system. Adenylation domain activators have previously been shown to be necessary for the overexpression and activity of NRPS modules.(55) Despite the fact that the A domain of CysK3 is inactive CysA still proved necessary for its overexpression which implies that MbtH protein association is probably more a factor of stabilization rather than activation of NRPS modules.

### 3.2.2 Protein MS CysK3

To determine whether CysB could mediate the full transfer from CysH to CysK3 we used a similar assay as described above for the first aminoacyl transfer but with the addition of the final acceptor CysK3 as analyte for protein LC-MS and CysB in catalytic amounts. As expected CysK3 is unable to activate or load asparagine on its own due to its mutated adenylation domain but loading could be restored in presence of both CysH and CysB implying that both parts of the aminoacyl transfer are functional *in vitro* and carried out by CysB. Contrary to the first part of the shuttling where even after condition optimization the ratio between free CysB and loaded CysB is at best of 1:1, in this case the loading of CysK3 is almost complete, implying that this half of the reaction is probably driving the cycle towards the transfer from CysH to CysK3 rather than the other way around. The differences in the kinetic of the ping and the pong reactions in addition to the apparent sensibility to pH of this reaction imply that it could be driven by pI difference between the PCP domains of CysH (pI=5,6) and CysK3 (pI=6,3) as it was also proposed in the case of CmaE. The presence of a condensation domain on CysK3 accepting the loaded aminoacyl thus preventing its return on CysB could also play a role in the dynamic equilibrium.

## CysB - A new amino acyl transferase shuttles modified asparagine residues between a standalone NRPS module and the assembly line

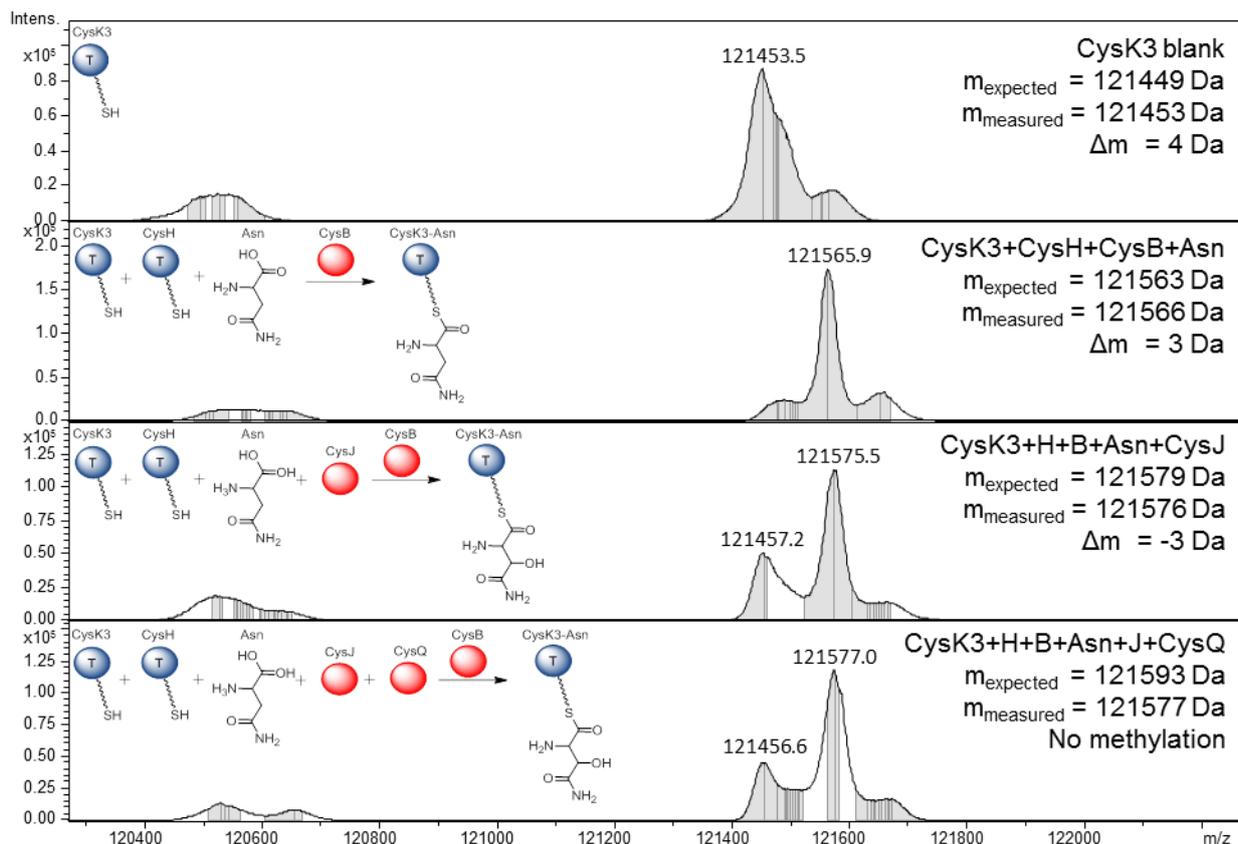


Figure 4.22: MS analysis of CysK3 incubated with CysB, CysH, asparagine, CysJ and CysQ. For the sake of simplicity only the T domain of Module CysH A-X-T and Module CysK3 C-A-T were depicted.

## 4 Conclusion

We report a new asparaginyl transferase, CysB expanding the family of amino acyl transferases “shuttling proteins”. *In vitro* analysis of this rarely studied class of proteins was rendered possible through the overexpression and purification of both NRPS modules partners CysH and CysK3. CysB is the missing link between the standalone A-X-T NRPS modules CysH were both *in trans* tailoring by the non-heme  $\text{Fe}^{\text{II}}$  dioxygenase CysJ and *in cis* tailoring by the X domain happen and the assembly line where the peptidic backbone is assembled. CysB seems to display complex kinetic parameters that could deserve further investigation. However this process has proven to be quite challenging to investigate since it relies on a complex three part reaction mixture involving two large NRPS modules in addition to the target protein.

---

## References

1. Walsh, C. T., Chen, H., Keating, T. A., Hubbard, B. K., Losey, H. C., Luo, L., Marshall, C. G., Miller, D. A., and Patel, H. M. (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines, *Current opinion in chemical biology* 5, 525–534.
2. Lacroix, D., Prado, S., Kamoga, D., Kasenene, J., Zirah, S., and Bodo, B. (2012) Unprecedented occurrence of isoaspartic acid in a plant cyclopeptide, *Organic letters* 14, 576–579.
3. Fu, C., Keller, L., Bauer, A., Brönstrup, M., Froidbise, A., Hammann, P., Herrmann, J., Mondesert, G., Kurz, M., Schiell, M., Schummer, D., Toti, L., Wink, J., and Müller, R. (2015) Biosynthetic Studies of Telomycin Reveal New Lipopeptides with Enhanced Activity, *Journal of the American Chemical Society* 137, 7692–7705.
4. Singh, G. M., Fortin, P. D., Koglin, A., and Walsh, C. T. (2008) beta-Hydroxylation of the aspartyl residue in the phytotoxin syringomycin E: characterization of two candidate hydroxylases AspH and SyrP in *Pseudomonas syringae*, *Biochemistry* 47, 11310–11320.
5. Elkins, J. M., Ryle, M. J., Clifton, I. J., Dunning Hotopp, J. C., Lloyd, J. S., Burzlaff, N. I., Baldwin, J. E., Hausinger, R. P., and Roach, P. L. (2002) X-ray Crystal Structure of *Escherichia coli* Taurine/ $\alpha$ -Ketoglutarate Dioxygenase Complexed to Ferrous Iron and Substrates †,‡, *Biochemistry* 41, 5185–5192.
6. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases, *Chemistry & biology* 6, 493–505.
7. Singh, G. M., Vaillancourt, F. H., Yin, J., and Walsh, C. T. (2007) Characterization of SyrC, an aminoacyltransferase shuttling threonyl and chlorothreonyl residues in the syringomycin biosynthetic assembly line, *Chemistry & biology* 14, 31–40.
8. Bloudoff, K., and Schmeing, T. M. (2017) Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity, *Biochimica et biophysica acta* 1865, 1587–1604.
9. Hornbogen, T., Riechers, S.-P., Prinz, B., Schultchen, J., Lang, C., Schmidt, S., Mügge, C., Turkanovic, S., Süßmuth, R. D., Tauberger, E., and Zocher, R. (2007) Functional characterization of the recombinant N-methyltransferase domain from the multienzyme enniatin synthetase, *Chembiochem : a European journal of chemical biology* 8, 1048–1054.
10. Baumann, S., Herrmann, J., Raju, R., Steinmetz, H., Mohr, K. I., Hüttel, S., Harmrolfs, K., Stadler, M., and Müller, R. (2014) Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity, *Angewandte Chemie (International ed. in English)* 53, 14605–14609.
11. Chen, H., Thomas, M. G., O'Connor, S. E., Hubbard, B. K., Burkart, M. D., and Walsh, C. T. (2001) Aminoacyl- S -Enzyme Intermediates in  $\beta$ -Hydroxylations and  $\alpha,\beta$ -Desaturations of Amino Acids in Peptide Antibiotics †, *Biochemistry* 40, 11651–11659.

12. Hou, J., Robbel, L., and Marahiel, M. A. (2011) Identification and characterization of the lysobactin biosynthetic gene cluster reveals mechanistic insights into an unusual termination module architecture, *Chemistry & biology* 18, 655–664.
13. Hubbard, B. K., and Walsh, C. T. (2003) Vancomycin assembly: nature's way, *Angewandte Chemie (International ed. in English)* 42, 730–765.
14. Hüttel, W. (2017) Structural diversity in echinocandin biosynthesis: the impact of oxidation steps and approaches toward an evolutionary explanation, *Zeitschrift für Naturforschung. C, Journal of biosciences* 72, 1–20.
15. Kurome, T., Inami, K., Inoue, T., Ikai, K., Takesako, K., Kato, I., and Shiba, T. (1993) Total Synthesis of Aureobasidin A, an Antifungal Cyclic Depsipeptide, *Chem. Lett.* 22, 1873–1876.
16. Rami Reddy, M.V., Harper, M. K., and Faulkner, D.J. (1998) Eurypamide A, a cyclic tripeptide from the Palauan sponge *Microciona eurypa*, *Tetrahedron* 54, 10649–10656.
17. Nussbaum, F. von, Anlauf, S., Freiberg, C., Benet-Buchholz, J., Schamberger, J., Henkel, T., Schiffer, G., and Häbich, D. (2008) Total synthesis and initial structure-activity relationships of longicatenamycin A, *ChemMedChem* 3, 619–626.
18. Shen, B., Du, L., Sanchez, C., Edwards, D. J., Chen, M., and Murrell, J. M. (2002) Cloning and characterization of the bleomycin biosynthetic gene cluster from *Streptomyces verticillus* ATCC15003, *Journal of natural products* 65, 422–431.
19. Wohlrab, A., Lamer, R., and VanNieuwenhze, M. S. (2007) Total synthesis of plusbacin A3: a depsipeptide antibiotic active against vancomycin-resistant bacteria, *Journal of the American Chemical Society* 129, 4175–4177.
20. Zhang, C., Kong, L., Liu, Q., Lei, X., Zhu, T., Yin, J., Lin, B., Deng, Z., and You, D. (2013) In vitro characterization of echinomycin biosynthesis: formation and hydroxylation of L-tryptophanyl-S-enzyme and oxidation of (2S,3S)  $\beta$ -hydroxytryptophan, *PloS one* 8, e56772.
21. Bollinger, J. M., Price, J. C., Hoffart, L. M., Barr, E. W., and Krebs, C. (2005) Mechanism of Taurine. A-Ketoglutarate Dioxygenase (TauD) from *Escherichia coli*, *Eur. J. Inorg. Chem.* 2005, 4245–4254.
22. Hegg, E. L., and Que, L. (1997) The 2-His-1-carboxylate facial triad--an emerging structural motif in mononuclear non-heme iron(II) enzymes, *European journal of biochemistry* 250, 625–629.
23. O'Brien, J. R., Schuller, D. J., Yang, V. S., Dillard, B. D., and Lanzilotta, W. N. (2003) Substrate-induced conformational changes in *Escherichia coli* taurine/alpha-ketoglutarate dioxygenase and insight into the oligomeric structure, *Biochemistry* 42, 5547–5554.
24. Hausinger, R. P. (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes, *Critical reviews in biochemistry and molecular biology* 39, 21–68.
25. Chen, H., and Walsh, C. T. (2001) Coumarin formation in novobiocin biosynthesis: beta-hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI, *Chemistry & biology* 8, 301–312.

26. Chen, H., Hubbard, B. K., O'Connor, S. E., and Walsh, C. T. (2002) Formation of beta-hydroxy histidine in the biosynthesis of nikkomycin antibiotics, *Chemistry & biology* 9, 103–112.
27. Belecki, K., and Townsend, C. A. (2013) Biochemical determination of enzyme-bound metabolites: preferential accumulation of a programmed octaketide on the enediyne polyketide synthase CalE8, *Journal of the American Chemical Society* 135, 14339–14348.
28. Hušek, P. (1991) Amino acid derivatization and analysis in five minutes, *FEBS Letters* 280, 354–356.
29. Fu, C., Auerbach, D., Li, Y., Scheid, U., Luxenburger, E., Garcia, R., Irschik, H., and Müller, R. (2017) Solving the Puzzle of One-Carbon Loss in Ripostatin Biosynthesis, *Angewandte Chemie (International ed. in English)* 56, 2192–2197.
30. Parent, A., Guillot, A., Benjdia, A., Chartier, G., Leprince, J., and Berteau, O. (2016) The B12-Radical SAM Enzyme PoyC Catalyzes Valine C $\beta$ -Methylation during Polytheonamide Biosynthesis, *Journal of the American Chemical Society* 138, 15515–15518.
31. Kudo, F., Miyanaga, A., and Eguchi, T. (2014) Biosynthesis of natural products containing  $\beta$ -amino acids, *Natural product reports* 31, 1056–1073.
32. Matsunaga, S., Fusetani, N., Hashimoto, K., and Walchli, M. (1989) Theonellamide F. A novel antifungal bicyclic peptide from a marine sponge *Theonella* sp, *J. Am. Chem. Soc.* 111, 2582–2588.
33. Pedras, M. S. C., Irina Zaharia, L., and Ward, D. E. (2002) The destruxins: synthesis, biosynthesis, biotransformation, and biological activity, *Phytochemistry* 59, 579–596.
34. Leonardi, R., and Jackowski, S. (2007) Biosynthesis of Pantothenic Acid and Coenzyme A, *EcoSal Plus* Epub 2007. DOI: 10.1128/ecosalplus.3.6.3.4.
35. Rachid, S., Krug, D., Weissman, K. J., and Müller, R. (2007) Biosynthesis of (R)-beta-tyrosine and its incorporation into the highly cytotoxic chondramides produced by *Chondromyces crocatus*, *The Journal of biological chemistry* 282, 21810–21817.
36. Felnagle, E. A., Rondon, M. R., Berti, A. D., Crosby, H. A., and Thomas, M. G. (2007) Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin, *Applied and environmental microbiology* 73, 4162–4170.
37. Schafhauser, T., Kirchner, N., Kulik, A., Huijbers, M. M. E., Flor, L., Caradec, T., Fewer, D. P., Gross, H., Jacques, P., Jahn, L., Jokela, J., Leclère, V., Ludwig-Müller, J., Sivonen, K., van Berkel, Willem J H, Weber, T., Wohlleben, W., and van Pée, K.-H. (2016) The cyclochlorotine mycotoxin is produced by the nonribosomal peptide synthetase CctN in *Talaromyces islandicus* ('*Penicillium islandicum*'), *Environmental microbiology* 18, 3728–3741.
38. Wu, B., Szymański, W., Heberling, M. M., Feringa, B. L., and Janssen, D. B. (2011) Aminomutases: mechanistic diversity, biotechnological applications and future perspectives, *Trends in biotechnology* 29, 352–362.
39. Krug, D., and Müller, R. (2009) Discovery of additional members of the tyrosine aminomutase enzyme family and the mutational analysis of CmdF, *Chembiochem : a European journal of chemical biology* 10, 741–750.

40. Ruzicka, F. J., and Frey, P. A. (2007) Glutamate 2,3-aminomutase: a new member of the radical SAM superfamily of enzymes, *Biochimica et biophysica acta 1774*, 286–296.
41. Tang, K.-H., Casarez, A. D., Wu, W., and Frey, P. A. (2003) Kinetic and biochemical analysis of the mechanism of action of lysine 5,6-aminomutase, *Archives of biochemistry and biophysics 418*, 49–54.
42. Webb, N. A., Mulichak, A. M., Lam, J. S., Rocchetta, H. L., and Garavito, R. M. (2004) Crystal structure of a tetrameric GDP-D-mannose 4,6-dehydratase from a bacterial GDP-D-rhamnose biosynthetic pathway, *Protein science : a publication of the Protein Society 13*, 529–539.
43. Grabowski, R., Hofmeister, A. E., and Buckel, W. (1993) Bacterial L-serine dehydratases: a new family of enzymes containing iron-sulfur clusters, *Trends in biochemical sciences 18*, 297–300.
44. O'Brien, J. R., Raynaud, C., Croux, C., Girbal, L., Soucaille, P., and Lanzilotta, W. N. (2004) Insight into the mechanism of the B12-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization, *Biochemistry 43*, 4635–4645.
45. Akashi, T., Aoki, T., and Ayabe, S.-I. (2005) Molecular and biochemical characterization of 2-hydroxyisoflavanone dehydratase. Involvement of carboxylesterase-like proteins in leguminous isoflavone biosynthesis, *Plant physiology 137*, 882–891.
46. Liao, R.-Z., and Thiel, W. (2012) Why is the oxidation state of iron crucial for the activity of heme-dependent aldoxime dehydratase? A QM/MM study, *The journal of physical chemistry. B 116*, 9396–9408.
47. Huang, W., Jia, J., Cummings, J., Nelson, M., Schneider, G., and Lindqvist, Y. (1997) Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold, *Structure 5*, 691–699.
48. Hüttel, S., Testolin, G., Herrmann, J., Planke, T., Gille, F., Moreno, M., Stadler, M., Brönstrup, M., Kirschning, A., and Müller, R. (2017) Discovery and Total Synthesis of Natural Cystobactamid Derivatives with Superior Activity against Gram-Negative Pathogens, *Angewandte Chemie (International ed. in English) 56*, 12760–12764.
49. Kim, Y. J., Kim, H.-J., Kim, G.-W., Cho, K., Takahashi, S., Koshino, H., and Kim, W.-G. (2016) Isolation of Coralmycins A and B, Potent Anti-Gram Negative Compounds from the Myxobacteria *Coralloccoccus coralloides* M23, *Journal of natural products 79*, 2223–2228.
50. Stachelhaus, T., Mootz, H. D., Bergendahl, V., and Marahiel, M. A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain, *The Journal of biological chemistry 273*, 22773–22781.
51. Singh, G. M., Vaillancourt, F. H., Yin, J., and Walsh, C. T. (2007) Characterization of SyrC, an aminoacyltransferase shuttling threonyl and chlorothreonyl residues in the syringomycin biosynthetic assembly line, *Chemistry & biology 14*, 31–40.
52. Strieter, E. R., Vaillancourt, F. H., and Walsh, C. T. (2007) CmaE: a transferase shuttling aminoacyl groups between carrier protein domains in the coronamic acid biosynthetic pathway, *Biochemistry 46*, 7549–7557.

53. Heikinheimo, P., Goldman, A., Jeffries, C., and Ollis, D. L. (1999) Of barn owls and bankers: A lush variety of  $\alpha/\beta$  hydrolases, *Structure* 7, R141-R146.
54. Cantu, D. C., Chen, Y., and Reilly, P. J. (2010) Thioesterases: a new perspective based on their primary and tertiary structures, *Protein science : a publication of the Protein Society* 19, 1281–1295.
55. Herbst, D. A., Boll, B., Zocher, G., Stehle, T., and Heide, L. (2013) Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes, *The Journal of biological chemistry* 288, 1991–2003.



## Chapter 5: CysP - Cystobactamide resistance protein

### 1 Introduction

Cystobactamides are a new class of potent broad-spectrum antibacterial compounds which exhibit bactericidal activity on a number of strains from the ESKAPE panel.(1) Some derivatives show a remarkable nanomolar minimum inhibitory concentration on notoriously resistant Gram negative bacteria such as *Pseudomonas aeruginosa* or *Acinetobacter baumannii*. Cystobactamides are structurally close to Albicidins (2) with which they share a similar activity profile, and are also active on the same target: the bacterial gyrase. It is also noteworthy that no cross resistance is observed on fluoroquinolone resistant mutants which led to the assumption that they bind to another part of the target protein. Gyrase is a topoisomerase with the ability to negatively supercoil DNA.(3) It operates as a heterotetramer GyrA<sub>2</sub>GyrB<sub>2</sub> and introduces a double strand break in a DNA segment before allowing a second segment to travel through a series of gates at the dimer interface and finally re-tethering of the opened strand after having performed the supercoiling. This enzyme is essential to DNA replication, transcription, and gene regulation, in addition, the inhibition of the catalytic cycle can lead to double-strand DNA break making it a potent target for antibiotics.

Albicidins are produced by the causative agent of the sugarcane leaf scald, *Xanthomonas albilineans*, and three different self-resistance mechanisms have been reported in this organism. The first that was reported is AlbF, a specific DHA14 efflux pump encoded in the cluster allowing active transport of the antibiotic out of the cell.(4) The overexpression of this pump in *E. coli* led to a leap of 3000 fold increase in resistance making AlbF a very powerful self-resistance mechanism. The second resistance factor is the pentapeptide repeat protein AlbG, an unusual resistance protein also found in fluoroquinolone resistance.(5) Structural investigation of the pentapeptide repeat protein revealed that the protein is a DNA mimic that interacts with gyrase either to prevent the binding of the antibiotics or to prevent any harm to DNA while the antibiotic is bound to gyrase.(6) This protein only improved resistance four to six fold when overexpressed in *E.coli*. As a third self-resistance mechanism to Albicidin the GyrA subunit of the gyrase in *X. albilineans* shows significant mutations, especially with an insertion of 43 amino acids that seems to confer a high level of resistance to Albicidin as well as other antibiotics such as ciprofloxacin.(7)

*In silico* comparison of the Albicidin and Cystobactamide cluster reveal that they are mostly similar apart from a few key aspects of the biosynthesis. (Cf. chap. 1) While the gene encoding the pentapeptide repeat protein is present in both clusters (CysO and AlbG), the gene encoding the DHA14 pump in the Albicidin cluster, AlbF, is heavily truncated in the Cystobactamide cluster. CysM is 38 amino acids long and aligns only with one of the fourteen transmembrane helices reported for AlbF. In addition alignment of the gyrase of several Cystobactamide producer strains with *X. albilineans*, *E. coli* and non-producer myxobacteria show that the key insertion region conferring resistance in *X. albilineans* is

absent in myxobacteria. In this context it is likely that Cystobactamide producers evolved other self-resistance mechanisms and the only further unexplained difference between both biosynthesis clusters is the presence of the CysP gene in the Cystobactamide cluster.

## 2 Materials and methods

### 2.1 Cloning, overexpression and purification

Standard protocols were used for DNA cloning, transformation, amplification, and purification. Genomic DNA was extracted from *Cystobacter velatus* Cbv34 grown in CTT medium, the 702 bp DNA fragment encoding CysP was amplified using the forward primer 5'TATCATATGAAACGGTCTTCAAGCT and the reverse primer 5'TATCTCGAGTCAGCG-TCCGTGGCTGGATT and the 612 bp DNA fragment encoding CysO was amplified using the forward primer 5'TATCATATGCCCGCCCGCTCCACTCC and the reverse primer 5'TATCTC-GAGTCAGGGTGTCTGATAGCCGTCGA. The amplified DNA fragments were digested with NdeI and XhoI, and cloned into pET-28b with an N-terminal His6 tag. The resulting constructs were sequenced to verify that no mutation had been introduced during PCR.

Recombinant protein expression was carried out in *E.coli* BL21 (DE3) grown in LB medium supplemented with 50 µg/ml of kanamycin. The culture was inoculated 1/10 from a fully grown overnight culture and was initially grown at 37°C until it reached an OD of 0.6 to 0.9, at which point the temperature was decreased to 16°C. Induction was initiated after 30 minutes at 16°C with 0,1 mM isopropyl-β, D-thiogalactopyranoside (IPTG). The cells were harvested 16h after induction by centrifugation at 8000 g for 10 minutes at 4°C and resuspended in lysis buffer (25 mM TRIS pH7,5; 150 mM NaCl; 20 mM imidazole). The cells were lysed by a passage at 1500 bar at 4°C through a M-110P Microfluidizer (Microfluidics) after which the crude extract was centrifuged at 45000 g for 15 minutes at 4°C, the supernatant was treated as soluble protein fraction.

Affinity chromatography was performed on an Äkta Avant system while size exclusion chromatography was performed on an Äkta Pure system. The soluble fraction of the lysed cells was applied to 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10 CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole. The pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200 pg SEC column equilibrated in running buffer (25 mM TRIS pH 7,5; 150 mM NaCl; 2 mM DTT). The pooled fractions were controlled via SDS-PAGE, concentrated and frozen at -80°C after addition of 25% glycerol.

## 2.2 MIC measurements

*E.coli* BL21 harboring CysP, CysO or an empty pET28b plasmid were inoculated into 5 ml LB medium supplemented with 50 µg/ml of kanamycin and grown overnight. 50 ml LB medium supplemented with 50 µg/ml of kanamycin were inoculated 1/10 from the fully grown overnight culture and was initially grown at 37°C for about 2h until it reached an OD of 0.6 to 0.9, at which point the temperature was decreased to 30°C. Induction was initiated after 30 minutes at 30°C with 1 mM isopropyl-β,D-thiogalactopyranoside (IPTG) and protein expression was performed for 2h30 at 30°C. The cultures were diluted to an OD of 0,01 in LB medium in 96 wells plates and increasing concentrations of Cystobactamide were added to the medium in DMSO. OD measurements were performed after 16h growth to determine the MIC.

## 2.3 Binding studies - SEC

In 500 µl of buffer (50 mM TRIS pH7,5; 100 mM NaCl; 5 mM DTT and 5 % DMSO) 10 µM purified CysP was incubated 1h at RT with 100 µM fluorescent Cystobactamide tripeptide mimics **214** and **215** and injected on a Superdex Increase 10/300GL SEC column equilibrated in running buffer (25 mM TRIS pH 7,5; 150 mM NaCl; 2 mM DTT). 24 fractions of 1 ml each were collected and 20 µl of fractions 11 to 22 were analyzed through SDS-PAGE to confirm the presence of CysP, and 200 µl were transferred to a 96 wells plate to measure fluorescence in a TECAN infinite 200 plate reader equipped with a DAD, excitation was performed at 370 nm and emission was measured at 420 nm.

## 2.4 Binding studies - Fluorescence

In a volume of 100 µl in a 96 wells plate 10 µM **215** were incubated with increasing concentrations of CysP or BSA for 10 minutes at 30°C. The fluorescence was measured in a TECAN infinite 200 plate reader equipped with a DAD, excitation was performed at 370 nm and emission was measured at 420 nm.

In a volume of 100 µl in a 96 wells plate 5 µM **215** in 10 µM CysP were incubated for 10 min at 30°C before adding increasing concentrations of **17** and **241** which were incubated for further 10 minutes at 30°C. The fluorescence was measured in a TECAN infinite 200 plate reader equipped with a DAD, excitation was performed at 370nm and emission was measured at 420nm.

### 3 Results and discussions

#### 3.1 *In silico* analysis

CysP features two domains, both related to the vicinal oxygen chelate (VOC) superfamily which mainly encompasses type I extradiol dioxygenases but also glyoxylase I and the Bleomycin resistance protein.(8) This diverse family of proteins is known to show very low sequence identities while retaining the same fold. Bleomycin resistance is conferred by the *ble* system (9) commonly horizontally transferred between bacteria on multidrug resistance plasmids, this protein forms a dimer before binding to the natural product at the dimer interface in order to prevent the double strand DNA break Bleomycin can cause.(10) CysP homologues can be found in a few myxobacteria with about 40 % identity where they might play a role in drug resistance too.

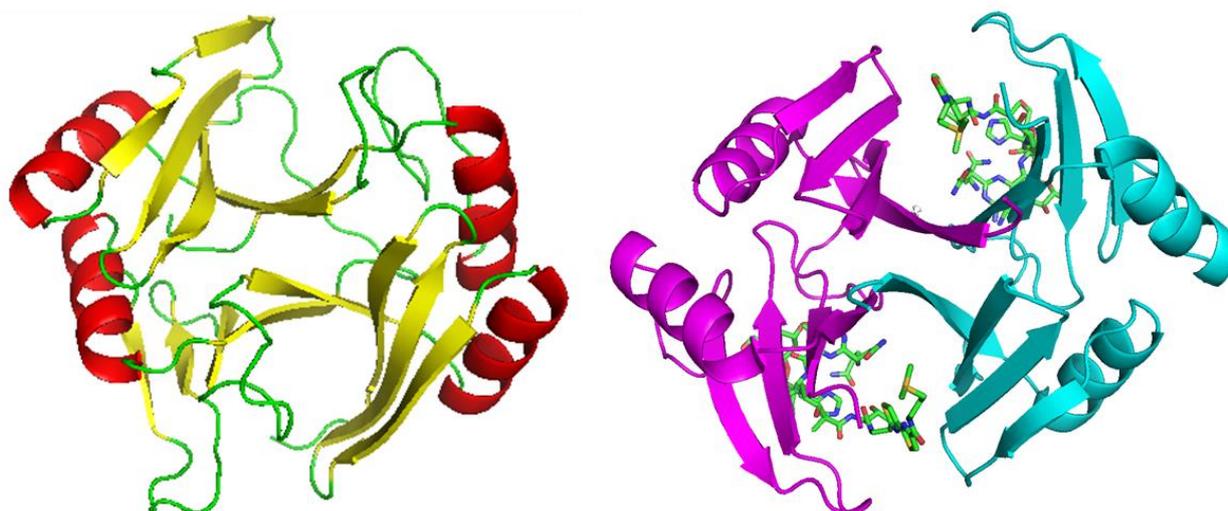


Figure 5.1: Model structure of CysP (helices are shown in red and strands in yellow) along with the dimer structure of the bleomycine resistance protein (monomers are shown in magenta and blue) binding two Bleomycin molecules (shown in green)

Two structures of VOC domain containing proteins can be found that share the same general architecture close to CysP on secondary structure despite very low sequence identity. They belong to the transposon based Tn5 *ble* dimer and the monomeric Rv0577 from *Mycobacterium tuberculosis* (11) which is, like CysP, a fused dimer. While Tn5 *ble* is a resistance protein sequestering Bleomycin, Rv0577 is thought to be a virulence factor that could be involved either in interacting with the host immune response or act as a receptor activating a phosphorylation cascade involved in secondary metabolism regulation. Structurally they are formed of a pseudotetramer of four repetitive  $\beta\uparrow\alpha\beta\uparrow\beta\downarrow\beta\uparrow$  subunits leading to two saddle-shaped binding sites facing away from each other on both sides of the protein, both constituted by an eight-stranded  $\beta$  sheet topped by an  $\alpha$  helix. A model for the structure of CysP was obtained with the I-TASSER online tool using RV0577 as a template

structure. Despite the very low sequence identity between both proteins a good model could be established with relatively good C-score of 1,00 and a RMSD at 3,6 Å. (Figure 5.1) The structure presents the complete pseudo-tetramer with the two saddle-shaped binding sites. The structure of Cystobactamide with two linear pABA chains linked by an  $\alpha$  or a  $\beta$  asparagine linker naturally tends to form a chevron in 3D which fits perfectly each binding site since they form a groove about 40 Å long which can fit a Cystobactamide 919-1 which spans 38,4 Å.

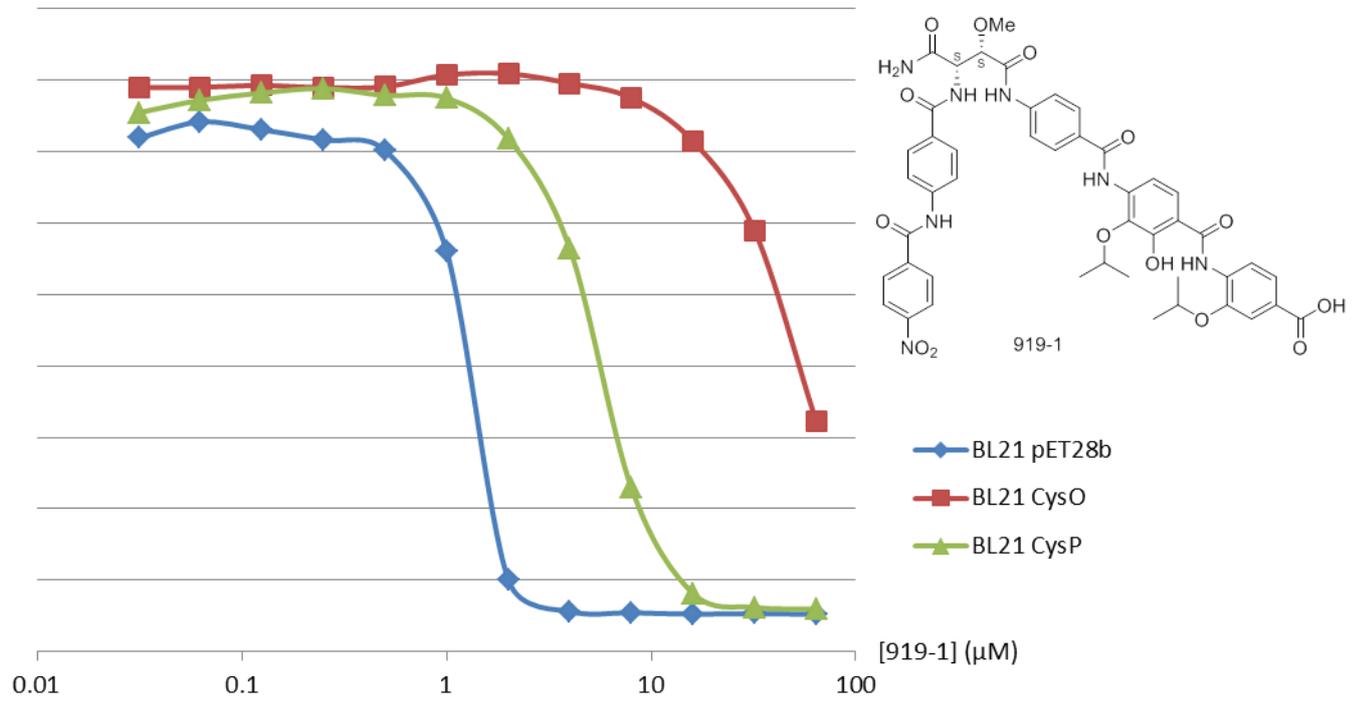
### 3.2 MIC comparisons

The absence of the DHA14 pump and of the gyrase mutation in Cystobactamide producers which are thought to be the most important reported Albicidin resistance mechanisms lead us to investigate the self-resistance mechanisms underlying the Cystobactamide production. Both putative resistance genes CysP and CysO were overexpressed in *E. coli* BL21 and the cultures were tested for MIC against the two major Cystobactamide derivatives, 919-1 and 919-2, which are isomers differing only by the nature of the linker, isoasparagine for 919-1 and asparagine for 919-2. (Figure 5.2) In both cases the MIC was increased about a 128 fold if CysO was expressed and about 16 fold if CysP was expressed. CysP is thus confirmed to be a resistance protein but its activity is accessory if compared to CysO, it is not clear whether both proteins might have a synergistic effect or whether CysP might have a further function apart from self-resistance.

Against *E. coli* BL21 Cystobactamide 919-2 showed an MIC of about 3.15 nM which is 300 fold lower than the MIC for Cystobactamide 919-1 at about 1  $\mu$ M. It is not clear why the minor compound 919-2 which is produced in only about one fifth of the concentration of 919-1 is much more active than the last, additionally 919-1 is the product of a complete tailoring while 919-2 is lacking one tailoring step. It is possible that the self-resistance mechanisms would not be sufficient to cope with a higher titer of 919-2. Furthermore a few resistance mechanisms to Albicidin have been reported and some could be transferable to Cystobactamides, the derivative cocktail produced might be aimed at overcoming these resistances in target strains.

## CysP - Cystobactamide resistance protein

[919-1] (μM)	64	32	16	8	4	2	1	0.5	0.25	0.125	0.0625	0.03125
BL21 pET28b (OD)	0.052	0.053	0.052	0.054	0.056	0.100	0.560	0.702	0.716	0.731	0.742	0.719
BL21 CysO (OD)	0.322	0.589	0.715	0.775	0.796	0.809	0.808	0.791	0.790	0.793	0.790	0.790
BL21 CysP (OD)	0.060	0.062	0.082	0.229	0.564	0.718	0.775	0.779	0.789	0.782	0.772	0.755



[919-2] (μM)	6.4	3.2	1.6	0.8	0.4	0.2	0.1	0.05	0.025	0.0125	0.00625	0.003125
BL21 pET28b (OD)	0.076	0.078	0.078	0.075	0.076	0.076	0.077	0.075	0.080	0.086	0.153	1.008
BL21 CysO (OD)	0.079	0.080	0.079	0.081	0.153	0.435	0.748	0.914	0.959	0.990	1.002	1.008
BL21 CysP (OD)	0.082	0.086	0.083	0.080	0.086	0.089	0.098	0.158	0.426	0.706	0.935	1.000

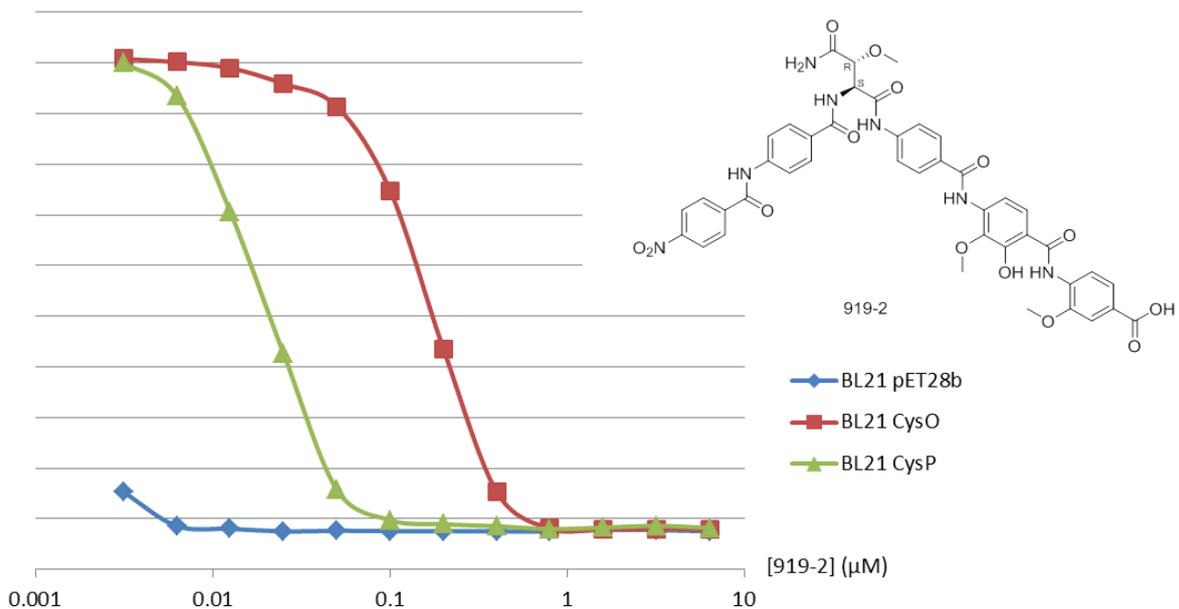


Figure 5.2: MIC measurements, OD reached by *E. coli* BL21 in the presence of increasing concentrations of Cystobactamide 919-1 or 919-2

Another fundamental discrepancy is observed since the pentapeptide repeat protein CysO conferred an increase in resistance of about 128 fold. In comparison, in the case of Albicidin the pentapeptide repeat protein AlbG improved the resistance in *E.coli* by only 6 fold but AlbG and CysO share only 30% identity which could imply differences in their ability to interact with the gyrase. This could be explained by the fact that the pentapeptide repeat protein has to interact with the host DNA gyrase and has thus to display a certain degree of specificity for the target rather than for the antibiotic. However in the case of *X. albilineans* the modification of the gyrase A subunit could imply that AlbG has a significantly lower binding affinity to the *E.coli* gyrase than CysO, and thus confer lower resistance levels. Furthermore, it is noteworthy that the expression rate of both genes in the heterologous host *E.coli* might be different.

### 3.3 Binding with tripeptides

To investigate the binding properties of CysP towards Cystobactamides we took advantage of the natural fluorescence displayed by some of the synthetic tripeptide mimics of Cystobactamide **507**.(12) (Figure 5.3) CysP was thus incubated with Cystobactamide **215** and **214** and repurified from the mixture by size exclusion chromatography. Fluorescence can be observed specifically in the fractions containing the protein and protein aggregates when using a tripeptidic Cystobactamide mimics harboring a free acid **215** implying a binding with CysP while its methyl ester counterpart **214** did not show any binding. (Figure 5.4) This result, although limited due to the relatively large structural difference between the tripeptide used and a full sized hexapeptidic Cystobactamide, infers that the resistance conferred by CysP might indeed be based on sequestration of the antibiotic in a similar fashion as is observed in the bleomycin resistance system.

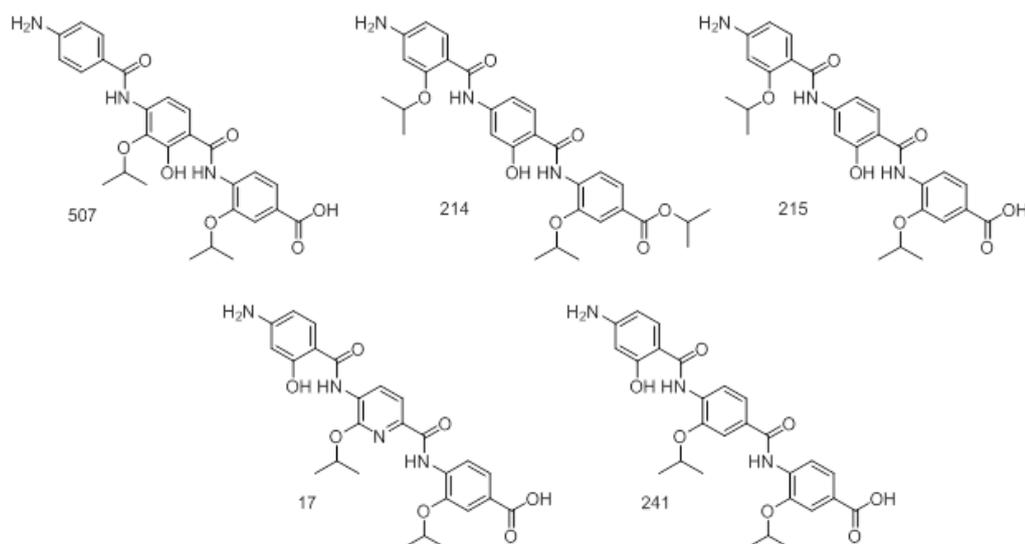


Figure 5.3: Structure of Cystobactamide 507 along with the synthetic derivatives used in this study; 214 and 215 are fluorescent while 17 and 241 are not

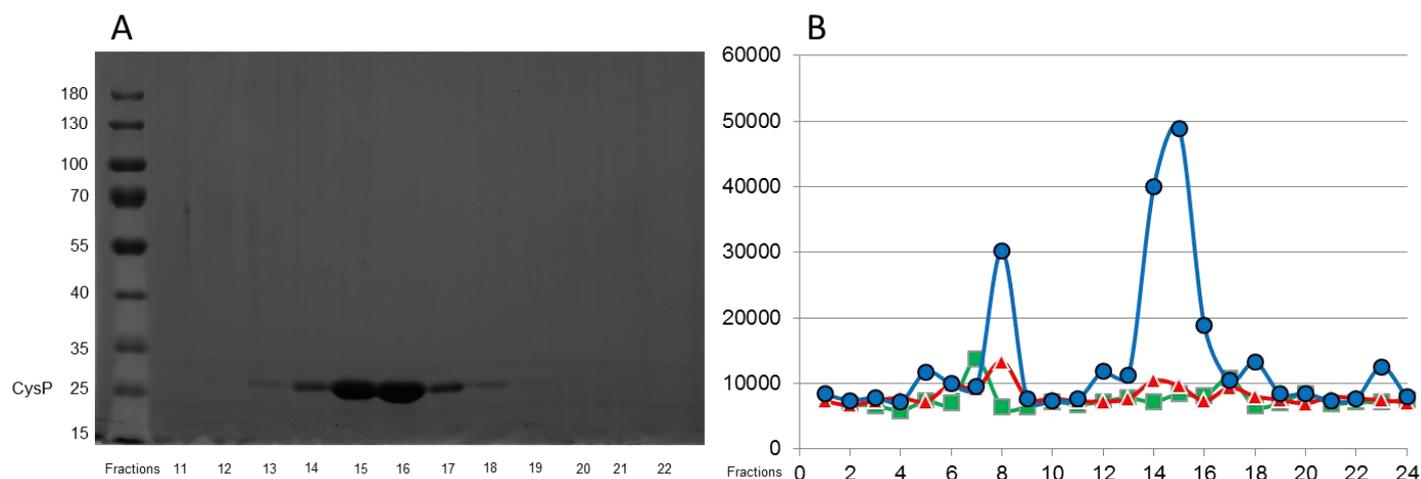


Figure 5.4: A: SDS gels showing SEC fractions containing CysP; B: fluorescence measurement of the same fractions; blue trace: CysP + 215; red trace: CysP + 214; green trace: CysP blank

Further assays using the tripeptide **215** showed that a fluorescence increase can be observed upon binding to CysP, we used this property to explore the stoichiometry of the binding and the possible unspecific binding **215** might display. (Figure 5.5) Upon incubation of **215** with increasing concentrations of CysP a plateau was reached at stoichiometric concentrations implying a 1:1 binding ratio between Cystobactamide and CysP. The fluorescence increase upon unspecific binding to BSA was also very low and only observed at higher concentrations validating the results. The reversibility of the binding was also assayed using non fluorescent tripeptides with similar structures (Cystobactamide **17** and **241**) incubated at increasing concentration on an equimolar mixture of CysP and **215**. The displacement of **215** can be observed via a concentration dependent fluorescence quenching that increases significantly when the non-fluorescent mimic reached equimolar concentration to the fluorescent one, confirming the reversible nature of the binding between CysP and Cystobactamides.

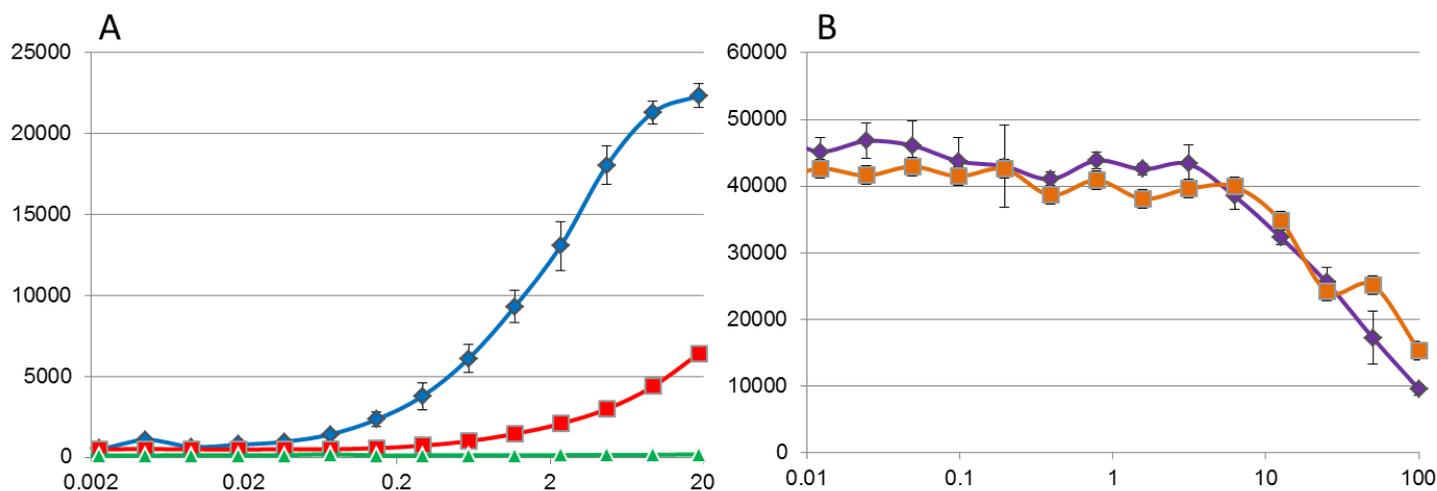


Figure 5.5: Fluorescence increase of **215** with increasing concentrations of proteins; blue trace: **215** + CysP; red trace: **215** + BSA; green trace: **215** blank; B: fluorescence quenching upon displacement of **215** from CysP by non-fluorescent Cystobactamide mimics; purple trace: CysP and **215** + **17**; orange trace: CysP and **215** + **241**

## 4 Conclusion

We could assign a function to the CysP protein encoded in the Cystobactamide biosynthesis cluster, CysP was proven to be the second self-resistance factor in the Cystobactamide biosynthesis along with the pentapeptide repeat protein CysO. The conferred resistance is probably linked to sequestration of the antibiotic as observed for the related bleomycin resistance protein. Self-resistance to Cystobactamide and Albicidin seems to follow significantly different routes with a well-documented preponderance of efflux and target mutation in the case of Albicidin, while the Cystobactamide self-resistance is based on a more efficient pentapeptide repeat protein and sequestration of the antibiotic. It is, however, likely that both mechanisms could show a synergistic effect. Further investigation of the binding between CysP and full-size Cystobactamide through MST or SPR spectroscopy techniques would be needed to confirm the binding observed with tripeptide mimics. Furthermore even if a lot of data could be inferred from a structural model of CysP further insights would be gained from cocrystallization of the protein with Cystobactamide. While *in vitro* experiments are essential to understand the biochemical properties of the protein further *in vivo* analysis in myxobacteria such as MIC evaluation would be needed to shed light on the exact impact of CysP on resistance in the producer organism.

## 5 References

1. Baumann, S., Herrmann, J., Raju, R., Steinmetz, H., Mohr, K. I., Hüttel, S., Harmrolfs, K., Stadler, M., and Müller, R. (2014) Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity, *Angewandte Chemie (International ed. in English)* 53, 14605–14609.
2. Cociancich, S., Pesic, A., Petras, D., Uhlmann, S., Kretz, J., Schubert, V., Vieweg, L., Duplan, S., Marguerettaz, M., Noëll, J., Pieretti, I., Hügelland, M., Kemper, S., Mainz, A., Rott, P., Royer, M., and Süßmuth, R. D. (2015) The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine, *Nature chemical biology* 11, 195–197.
3. Collin, F., Karkare, S., and Maxwell, A. (2011) Exploiting bacterial DNA gyrase as a drug target: current state and perspectives, *Applied microbiology and biotechnology* 92, 479–497.
4. Bostock, J. M., Huang, G., Hashimi, S. M., Zhang, L., and Birch, R. G. (2006) A DHA14 drug efflux gene from *Xanthomonas albilineans* confers high-level albicidin antibiotic resistance in *Escherichia coli*, *Journal of applied microbiology* 101, 151–160.
5. Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A., and Birch, R. G. (2007) The phytotoxin albicidin is a novel inhibitor of DNA gyrase, *Antimicrobial agents and chemotherapy* 51, 181–187.
6. Hegde, S. S., Vetting, M. W., Roderick, S. L., Mitchenall, L. A., Maxwell, A., Takiff, H. E., and Blanchard, J. S. (2005) A fluoroquinolone resistance protein from *Mycobacterium tuberculosis* that mimics DNA, *Science (New York, N.Y.)* 308, 1480–1483.
7. Hashimi, S. M., Huang, G., Maxwell, A., and Birch, R. G. (2008) DNA gyrase from the albicidin producer *Xanthomonas albilineans* has multiple-antibiotic-resistance and unusual enzymatic properties, *Antimicrobial agents and chemotherapy* 52, 1382–1390.
8. He, P., and Moran, G. R. (2011) Structural and mechanistic comparisons of the metal-binding members of the vicinal oxygen chelate (VOC) superfamily, *Journal of inorganic biochemistry* 105, 1259–1272.
9. Gatignol, A., Durand, H., and Tiraby, G. (1988) Bleomycin resistance conferred by a drug-binding protein, *FEBS Letters* 230, 171–175.
10. Maruyama, M., Kumagai, T., Matoba, Y., Hayashida, M., Fujii, T., Hata, Y., and Sugiyama, M. (2001) Crystal structures of the transposon Tn5-carried bleomycin resistance determinant uncomplexed and complexed with bleomycin, *The Journal of biological chemistry* 276, 9992–9999.
11. Buchko, G. W., Echols, N., Flynn, E. M., Ng, H.-L., Stephenson, S., Kim, H.-B., Myler, P. J., Terwilliger, T. C., Alber, T., and Kim, C.-Y. (2017) Structural and Biophysical Characterization of the *Mycobacterium tuberculosis* Protein Rv0577, a Protein Associated with Neutral Red Staining of Virulent Tuberculosis Strains and Homologue of the *Streptomyces coelicolor* Protein KbpA, *Biochemistry* 56, 4015–4027.
12. Moreno, M., Elgaher, W., Herrmann, J., Schläger, N., Hamed, M., Baumann, S., Müller, R., Hartmann, R., and Kirschning, A. (2015) Synthesis and Biological Evaluation of

Cystobactamid 507. A Bacterial Topoisomerase Inhibitor from *Cystobacter* sp, *Synlett* 26, 1175–1178.



# Chapter 6: CysK and CysG - Study of the assembly line and attempts at *in vitro* reconstitution

## 1 Introduction

In the frame of the investigation of the Cystobactamide biosynthesis we decided to go beyond the review of the biochemistry of the tailoring enzymes and aim for a complete *in vitro* reconstitution of the biosynthesis. Reconstitution of large assembly line type biosynthesis has rarely been reported mostly because of the challenge of expressing very large proteins in an active state in addition to the difficulties implied by complex protein mixtures.(1–3) Moreover in the particular case of NRPS assembly lines, it is noteworthy that despite the biochemistry of the adenylation domain being relatively well known (4) the condensation domain lacks that level of understanding.(5) Thus, the biochemistry is still only postulated to happen largely spontaneously with no indispensable catalytic residue identified. It has also been shown that C domains display a certain level of specificity for the donor and even more for the acceptor substrate.(6) But in the absence of substrate-bound co-crystal structures of C domains no insights exist towards the mechanism of this specificity and about the protein-substrate interactions happening in the active site.

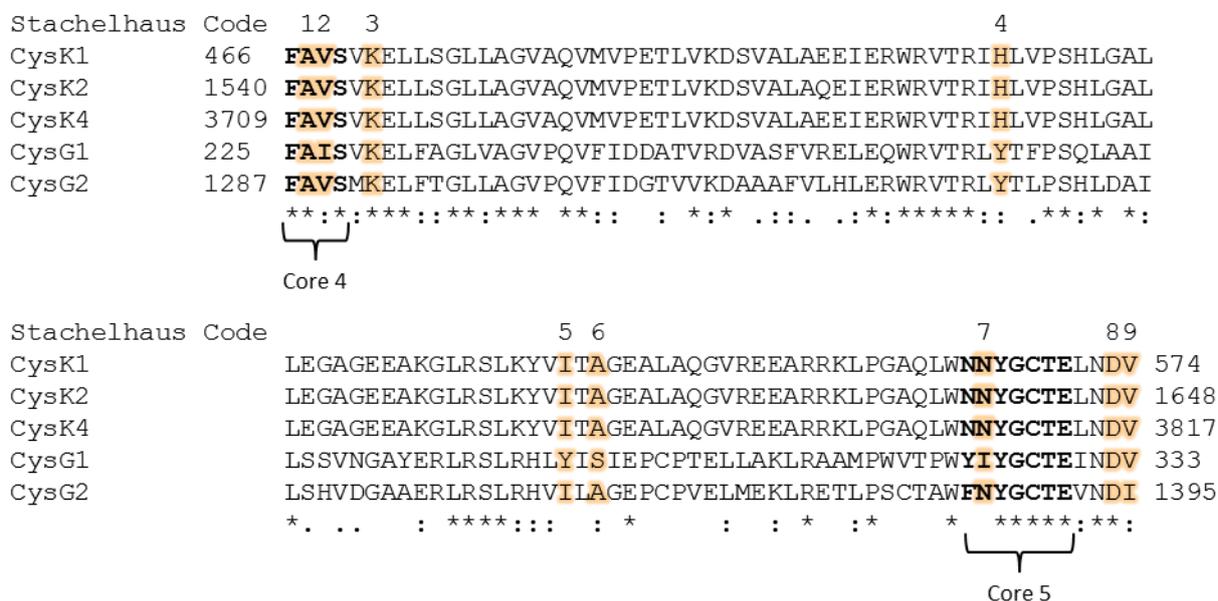


Figure 6.1: Alignment of A domains from CysK and CysG between Core domains 4 and 5; The Stachelhaus code residues are highlighted in orange

NRPS specificity prediction algorithms such as *NRPSpredictor2* (7) identify the substrate of the Adenylation domains as aromatic and mostly predict phenylalanine. The substrate specificity of the different adenylation domains is reflected in the Stachelhaus code

(8), indeed all three A domains from CysK have the same code as they probably activate the same substrate, pABA, while CysG1 and CysG2 display significant differences. (Figure 6.1) Analysis of the Stachelhaus code reveals that the first residue of the code, which is usually a conserved Aspartate responsible for stabilizing the alpha-amino group, is in the case of the Cystobactamide Adenylation modules mutated to Alanine, this is a common feature of Adenylation domains activating nonalpha amino acids.(9) The conserved aspartate seems to be shifted to position 8 and probably stabilizes the para-amino function from there. Most residues are hydrophobic as expected for aromatic substrates and the 4<sup>th</sup> position harbors either histidine for CysK or tyrosine for CysG both able to stabilize aromatic substrates through  $\pi$ - $\pi$  stacking. Unexpectedly there is quite a large variation between G1 and G2 despite them activating very similar substrates differing only in one hydroxyl.

## 2 Material and methods

### 2.1 Cloning, overexpression and purification

Standard protocols were used for DNA cloning, transformation, amplification, and purification. Genomic DNA was extracted from *Cystobacter velatus* Cbv34 grown in CTT medium

The 213 bp DNA fragment encoding CysA was amplified using the forward primer 5'TATCATATGAGCATGAACGGGGACG and the reverse primer 5'TATAGATCTTCAGCA-GTTGCTGCGCG. The amplified DNA fragment was digested with NdeI and BglIII, and cloned in the second multiple cloning site of pETduet-1 without any affinity tag.

The DNA fragment encoding CysK and CysG modules were amplified using the forward primer and the reverse primer described in table 6.1, the forward primer contains the ENLYFQG TEV protease recognition site. The amplified DNA fragments were digested with BamHI and HindIII, and cloned into the first multiple cloning site of pETduet-CysA to yield N-terminal His6 tag and TEV protease site fusion constructs. The resulting constructs were sequenced to verify that no mutation had been introduced during PCR

The 732 bp DNA fragment encoding CysE was amplified using the forward primer 5'TACATATGATGATTGCGTTCAACCCGCA and the reverse primer 5'TACTCGAGCTAC-AACACACACTCAAGCA. The amplified DNA fragments were digested with NdeI and XhoI, and cloned into pET-28b with an N-terminal His6 tag. The resulting construct was sequenced to verify that no mutation had been introduced during PCR.

Table 6.1: List of primers used in assembly line cloning

CysK1 for	5'TATGGATCCGGAAAACCTGTATTTTCAGGGCATGCTGCTGGAGGGAGAGCT
CysK1 rev 1	5'TATAAGCTTTCACTGCGTCCGCTCGACC
CysK1 rev 1/2	5'TATAAGCTTTACGGCACAACGTTGGAGAGC
CysK2 for 0	5'TATGGATCCGGAAAACCTGTATTTTCAGGGCGAGATTCCGCTCTCCTACCTGC
CysK2 rev 1	5'TATAAGCTTTCACTCGGTGTTCCCCGACG
CysK2 for 1/2	5'TATGGATCCGGAAAACCTGTATTTTCAGGGCGTTCGAGCGGACGCAGG
CysK2 rev 1/2	5'TATAAGCTTTACGACGGAGCGAGGGC
CysK3 for 0	5'TATGGATCCGGAAAACCTGTATTTTCAGGGCGCGGTGCTCTCGTTCGC
CysK3 rev 1	5'TATAAGCTTCCGATGGATCGACGACAC
CysK3 for 1/2	5'TATGGATCCGGAAAACCTGTATTTTCAGGGCAACACCGAGGCGGTGCT
CysK3 rev 1/2	5'TATAAGCTTTCCGGTCCGCTCCACC
CysK4 for 0	5'TATGGATCCGGAAAACCTGTATTTTCAGGGCCCGCTGCCTCTGGCGTA
CysK4 for 1/2	5'TATGGATCCGGAAAACCTGTATTTTCAGGGCGTGGAGCGGACCGGAC
CysK4 rev	5'TATAAGCTTTACCGAGCCCCCAGC
CysG1 for	5'TATGGATCTGGAAAACCTGTATTTTCAGGGCATGGCCACCAAATTGTCTGACTTC
CysG1 rev 1	5'TATAAGCTTTACATGCTGATCAGCCTCTGCG
CysG1 rev 1/2	5'TATAAGCTTTCACTGGCCGCGGTCCGT
CysG2 for 0	5'TATGGATCCGGAAAACCTGTATTTTCAGGGCGCGCTTCCGCTGTTCGC
CysG2 for 1/2	5'TATGGATCCGGAAAACCTGTATTTTCAGGGCACGGACCGCGGCCAG
CysG2 rev	5'TATAAGCTTTACGAAGCTCGCGTCCTC

Recombinant protein expression was carried out in *E.coli* BL21 (DE3) grown in LB medium supplemented with 50 µg/ml of kanamycin. The culture was inoculated 1/10 from a fully grown overnight culture and was initially grown at 37°C until it reached an OD of 0.6 to 0.9, at which point the temperature was decreased to 16°C. Induction was initiated after 30 minutes at 16°C with 0,1 mM isopropyl-β,D-thiogalactopyranoside (IPTG). The cells were harvested 16h after induction by centrifugation at 8000 g for 10 minutes at 4°C and resuspended in lysis buffer (25 mM TRIS pH 7,5; 150 mM NaCl; 20 mM imidazole). The cells were lysed by passage at 1500 bar at 4°C through a M-110P Microfluidizer (Microfluidics) after which the crude extract was centrifuged at 45000 g for 15 minutes at 4°C, the supernatant was treated as soluble protein fraction.

Affinity chromatography was performed on an Äkta Avant system while size exclusion chromatography was performed on an Äkta Pure system. The soluble fraction of the lysed cells was applied to a 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole. The pooled fractions were applied to a HiPrep 26/10 desalting column equilibrated in running buffer (25 mM Tris-HCl pH 7,5; 150 mM NaCl; 2 mM DTT). In the case of CysK and CysG the resulting fractions were pooled (≈30 ml) and incubated overnight at 4°C with TEV protease (1 mg/20 mg protein) and CysE (1 mM). After ≈16 h incubation 20mM imidazole was added to the solution prior to loading on a 5 ml Ni-NTA cartridge (GE) equilibrated with lysis buffer. The Ni-NTA column was washed with 10CV lysis buffer prior to elution with elution buffer containing 250 mM imidazole. The pooled fractions were concentrated to less than 5 ml and loaded on a HiLoad 16/600 Superdex 200

pg SEC column equilibrated in running buffer (25 mM TRIS pH7,5; 150 mM NaCl; 2 mM DTT). The pooled fractions were controlled via SDS-PAGE, concentrated and frozen at -80°C after addition of 25% glycerol.

## 2.2 Loading and condensation assays

To load NRPS modules, in a total volume of 50  $\mu$ l, 5  $\mu$ M of protein was incubated with 1 mM substrate and 1 mM ATP in 25 mM Tris-HCl pH 7,5; 150 mM NaCl, 10 mM  $MgCl_2$  for two hours at room temperature. The solution was then analyzed by LC-MS.

Condensation assay solutions were prepared with the same procedure and adding respective SNAC esters after 10 minutes incubation at room temperature, the mixture was left to incubate for two hours at room temperature before analysis by LC-MS. Various pH conditions were assayed (25 mM Bis-TRIS pH 6,8; TRIS pH 7,5 and TRIS pH 8), reactions were also incubated at room temperature or 30°C.

## 2.3 Protein MS

CysH was analyzed on a Dionex Ultimate 3000 UPLC system (Thermo Scientific) coupled with an maXis4G Q-TOF MS (Bruker) using an ESI in positive mode. The samples were run on an Aeris Widepore XB-C8, 150 x 2.1 mm, 3.6  $\mu$ m dp column (Phenomenex, USA). LC conditions: A-Water + 0.1% FA; B-Acetonitrile + 0.1 % FA at a flow rate of 300  $\mu$ l/min and 45°C. 0 min: 98 % A / 2 % B, 0.5 min: 98 % A / 2 % B, 10.5 min: 25 % A / 75 % B, 13.5 min: 25 % A / 75 % B, 14 min: 98 % A / 2 % B. The LC flow was split to 75  $\mu$ l/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany) using the standard Bruker ESI source. In the source region, the temperature was set to 180°C, the capillary voltage was 4000 V, the dry-gas flow was 6.0 l/min and the nebulizer was set to 1.1 bar. Mass spectra were acquired in positive ionization mode ranging from 150 – 2500 m/z at 2.5 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm (Copyright 1991-2004 Spectrum Square Associates, Inc.).

## 2.4 Synthesis of the SNAC esters

1,5 mmol N-acetyl cysteamine (15,95  $\mu$ l) and 1 equivalent of triethylamine (13,94  $\mu$ l) were dissolved in 10 ml DMF at room temperature and 185 mg of para-nitrobenzoate chloride (1 mmol) was added and left to stir at room temperature overnight. After completion, the reaction was quenched by addition of 20 ml 1N HCl in water and pNBA-SNAC thioester was extracted with DCM (3 x 30 mL). The combined organic extract was washed with brine, dried over anhydrous  $MgSO_4$ , and the solvent was removed by vacuum distillation.

147,6 mg of pNBA-SNAC (550  $\mu$ mol) was dissolved in 20 ml ethanol and heated to 55°C and 6 equivalents of metallic iron (181,5 mg) were added. 1,5 equivalents of ammonium chloride (88,275 mg) were dissolved in 5 ml water and added to the reaction mix

before heating at 90°C for 2 h. Reaction completion was confirmed via LC-MS and the iron was filtered out while hot, the filtrate was concentrated in vacuo. The residue was diluted with water (15 mL) and basified by NH<sub>4</sub>OH to pH 9. The pABA-SNAC was extracted with DCM (3 × 20 mL). The combined organic extract was washed with brine, dried over anhydrous MgSO<sub>4</sub>, and the solvent was removed by vacuum distillation.

137,6 mg of pABA-SNAC (578 μmol) and 1,5 equivalents of triethylamine (55,05 μl) were dissolved in 5 ml DMF at room temperature and 129,5 mg of para-nitrobenzoate chloride (700 μmol) was added and left to stir at room temperature overnight. After completion, the reaction was quenched by addition of 15 ml 1N HCl in water and pNBA-SNAC thioester was extracted with DCM (3 × 20 mL). The combined organic extract was washed with brine, dried over anhydrous MgSO<sub>4</sub>, and the solvent was removed by vacuum distillation.

## 3 Results and discussions

### 3.1 Linker modelling

The large size of the NRPS proteins CysK (507 kDa) and CysG (220 kDa) make an overexpression of the full sized proteins unrealistic. Thus, we took advantage of the modular architecture with separate domains to express each module separately. In this frame, however, separately overexpressed modules will have to interact in the right order in vitro, thus it is critical that the linker region between modules is optimally designed. In this cluster, these linkers are 23 amino acids long except for the T2-C3 linker which is only 21 amino acids long. To design the constructs, models were generated on existing crystal structures. In this case, only two structures are published that cover this particular inter-module region (10,11) one of them is probably not in a catalytic conformation and thus not usable as a template structure. The other one is a structure of a Thiolation-Epimerization didomain from GrsA in the gramicidin biosynthesis, with the epimerization domain being structurally very close to the condensation domain this structure offers a satisfying template for a model of this linker.

CysK3 was used as a screening platform to find the optimal linker length on both N and C terminus of the C-A-T module. The structures of T2-C3 and T3-C4 were modeled (Figure 6.2) using the GrsA structure (PDB ID: 5ISX) as the template. The modeling was done using I-TASSER online tool (12) and gave models with C scores of respectively 1,55 and 1,34 which indicate relatively high confidence in the quality of the models. Since the linker region seems to associate more tightly with the downstream C domain rather than with the PCP, multiple constructs were designed in regard to the length of the linkers. On the N terminus of the module the minimal linker is constituted from the AVLS sequence which is buried between an alpha helix on its right and a beta sheet on its left, the solvent accessibility of these residues has been calculated to be 0 during the modeling which indicates that they could be considered to be part of the C domain. An intermediate linker length would encompass

GNTEAVLS that strongly interacts with the C domain and a rational full-length linker would go until TQRELALAPSGNTEAVLS to include the flexible region without including the residues interacting with the T domain. (Figure 6.3)

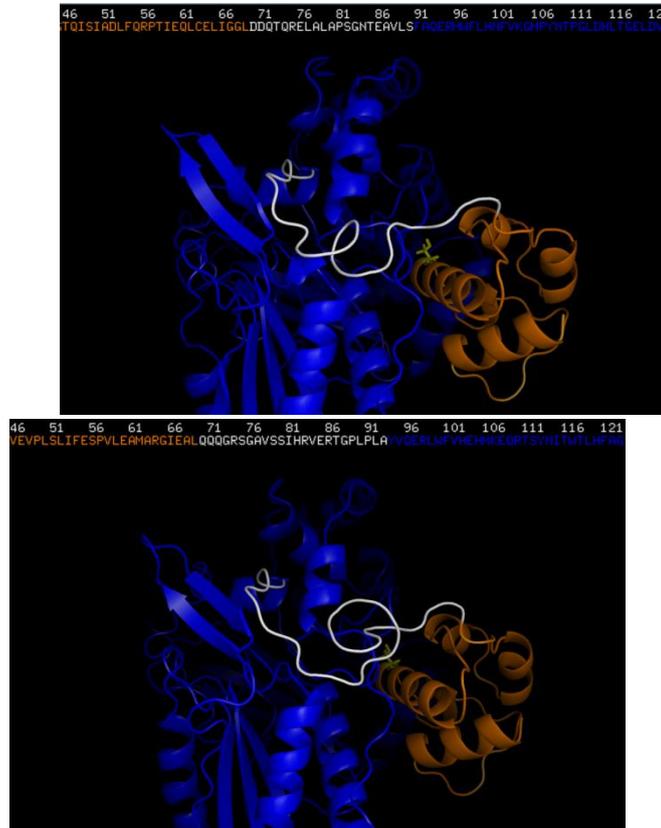


Figure 6.2: Models of the T2-C3 and T3-C4 domains; T domain is shown in orange, linker is shown in white and C domain in blue

In the C terminus linker only the first and third Q are weakly interacting with the T domain, QQQ was thus chosen as the minimal linker. In order to get a potential *in vitro* interaction between separately expressed modules 3 and 4 the linker should probably be kept on the C terminus of T3 in order to get an interaction with C4, thus the intermediate length was chosen as QQQGRSGAVSSIHR which encompasses the flexible region weakly interacting with the C domain. And the full length linker was chosen as QQQGRSGAVSSIHRVERTG to include part the strongly interacting but excluding the few amino acids buried inside the C domain. (Figure 6.3) While CysK3 was cloned with every possible combination of linkers only two series of modules were designed for the other modules with either half linkers on each side of the protein to promote stability or a full size linker on the N-terminus and none on the C-terminus in order to promote *in vitro* protein-protein interaction.

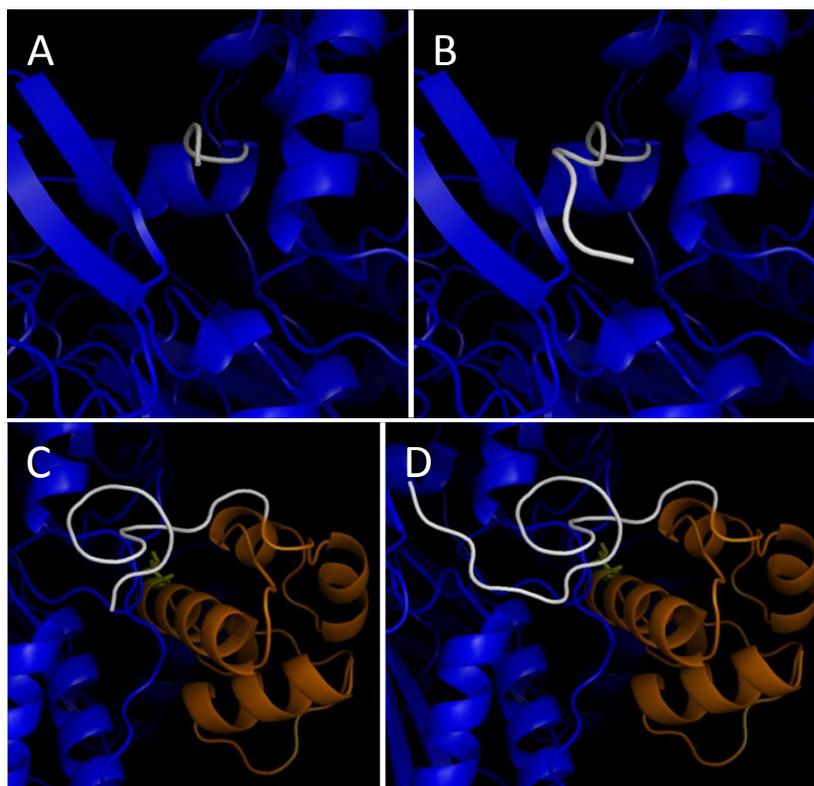


Figure 6.3: linker length used for overexpression: A- minimal linker on the N terminus; B- half linker on the N terminus; C- half linker on the C terminus; D- full linker on

## 3.2 CysA

The coexpression of the MbtH type Adenylation domain activator present in the cluster CysA proved to be indispensable for the expression of the NRPS modules since initial overexpression attempts with pHis cloned NRPS modules yielded only very limited amounts of recombinant protein apart for the smaller CysG1. This problem could not be solved after the initial screening phase for expression conditions and purification buffers. It was thus decided to coexpress the MbtH protein of the cluster, CysA, in a pET duet-1 vector system. An untagged native CysA was coexpressed with the His<sub>6</sub>-TEV tagged NRPS modules which allowed overexpression of all the modules in sufficient to very high yields and subsequent separation of the two proteins.

CysA is a small 8 kDa protein belonging to the MbtH type adenylation domain activators. These proteins do not possess any known catalytic activity but are known to form a quaternary complex with the Adenylation domain of NRPS modules.<sup>(13)</sup> It is not yet clear whether the A domain activators promote the expression of NRPS modules by helping the folding or preventing the degradation by proteases. MbtH proteins are known to bind A domains of which they are thought to be activators, in some cases, they were reported to impact their substrate specificity but there is only little data to support the hypotheses regarding their mode of action.<sup>(14)</sup>

### 3.3 CysE

Protein MS analysis of all the purified NRPS modules except CysK<sub>3</sub>, which possesses an inactive A domain, revealed that they were partially loaded with pABA *in vivo* during the overexpression. This phenomenon is common since NRPS and PKS assembly lines derive their building blocks from the primary metabolism, therefore they are readily available in the overexpression host cell. This problem is usually overcome by expressing the protein in absence of a phosphopantetheine transferase and by making this key post translational modification *in vitro*, however, in the case of myxobacterial carrier proteins this strategy often fails due to the ability of the *E.coli* phosphopantetheine transferase to activate myxobacterial assembly lines.

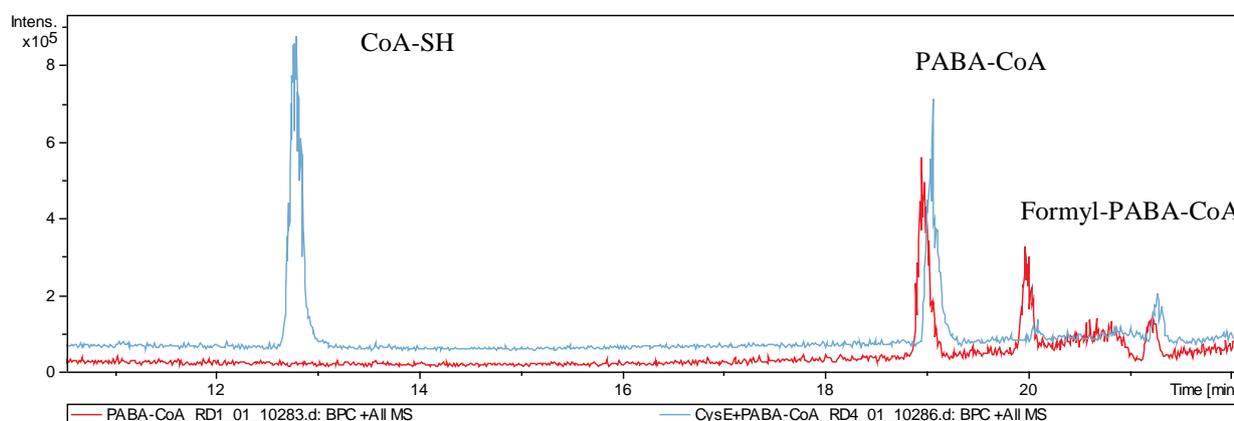


Figure 6.4: BPC of pABA-CoA and formyl-pABA-CoA without addition of CysE (red) or with addition of CysE (blue)

*In silico* analysis revealed that CysE was an  $\alpha/\beta$  hydrolase homologous to type II thioesterases. Since this type of enzyme is thought to have an editing activity to remove the non-cognate substrate from the assembly line, this *in silico* function assignment was confirmed by incubating CysE with a formylated pABA-CoA derivative in parallel to regular pABA-CoA. The enzyme selectively recognized and hydrolyzed the modified pABA-CoA corroborating the *in silico* analysis. (Figure 6.4) It has nevertheless been observed that type II thioesterases tend to be rather unspecific *in vitro* and that even cognate aminoacyl thioesters can be hydrolyzed with a lower turnover.(15)

Despite the fact that pABA is probably the cognate substrate of most modules of the Cystobactamide assembly line the unspecific thioesterase activity of CysE was confirmed in parallel to a chemical treatment with cysteamine. (Figure 6.5) Hence, CysE was proven to be useful as a tool for the unloading of the thioesterified amino acyls from phosphopentetheines. In subsequent overexpressions of the NRPS modules a His<sub>6</sub> tagged CysE was incubated with freshly purified NRPS modules in parallel with the TEV protease in order to unload the pABA from the PCP at the same time than tag cleavage allowing an easy separation during the second Ni-NTA affinity chromatography.

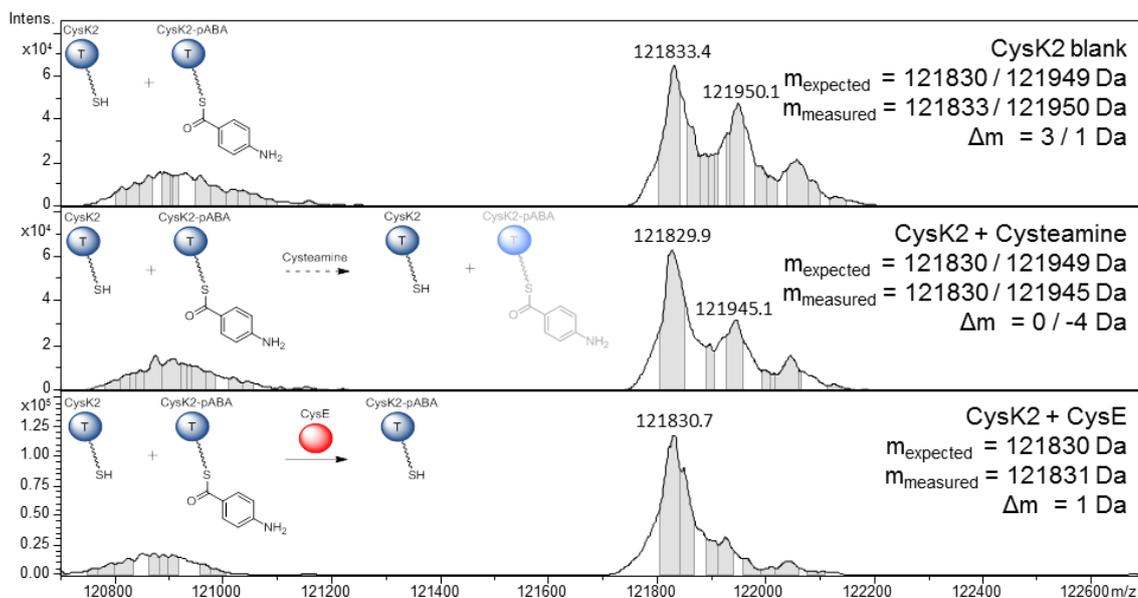


Figure 6.5: Protein MS of CysK2 incubated with cysteamine or CysE to unload pABA. The blank spectrum shows a high concentration of the loaded protein (second peak  $m/z = 121950$  Da) in comparison to the unloaded form (first peak  $m/z = 121833$  Da) this prevalence decreases slightly upon incubation with cysteamine through trans-thioesterification and the loaded specie completely disappears upon incubation with CysE which catalyzes the hydrolysis. For the sake of simplicity only the T domain of Module CysK2 C-A-T was depicted.

### 3.4 CysK sequencing

Original Illumina sequencing of the Cystobactamide cluster has proven difficult especially in the CysK gene because of the extremely high sequence identity of 99,3 % between the repetitive modules CysK<sub>1</sub>, CysK<sub>2</sub>, and CysK<sub>4</sub>. The *Cystobacter velatus* Cbv34 was thus resequenced using 454 sequencing to assign the correct module order. Regular cloning procedures include a sequencing step prior to overexpression to confirm the absence of PCR induced mutations in the cloned gene. Sequencing of the pETDuet-1 constructs after cloning of the NRPS modules CysK<sub>1</sub>, CysK<sub>2</sub> and CysK<sub>4</sub> revealed a high level of single nucleotide polymorphism in comparison with the original sequencing data with a total of 71 possible mutated positions. Since the number of mutations was significantly above the expected error rate of the phusion polymerase used for PCR and the constructs were also proven to successfully overexpress, multiple clones for both linker versions of each of the three modules were sequenced to confirm the data. The Single Nucleotide Polymorphism proved to be consistent indicating failure in the original sequencing process probably due to the high repetitively of the sequence. The only non-repetitive parts in CysK being the linkers were the primers for cloning were chosen, the confidence in the new sequence is a lot higher than the original data allowing for safe sequence correction. (Figure 6.6)

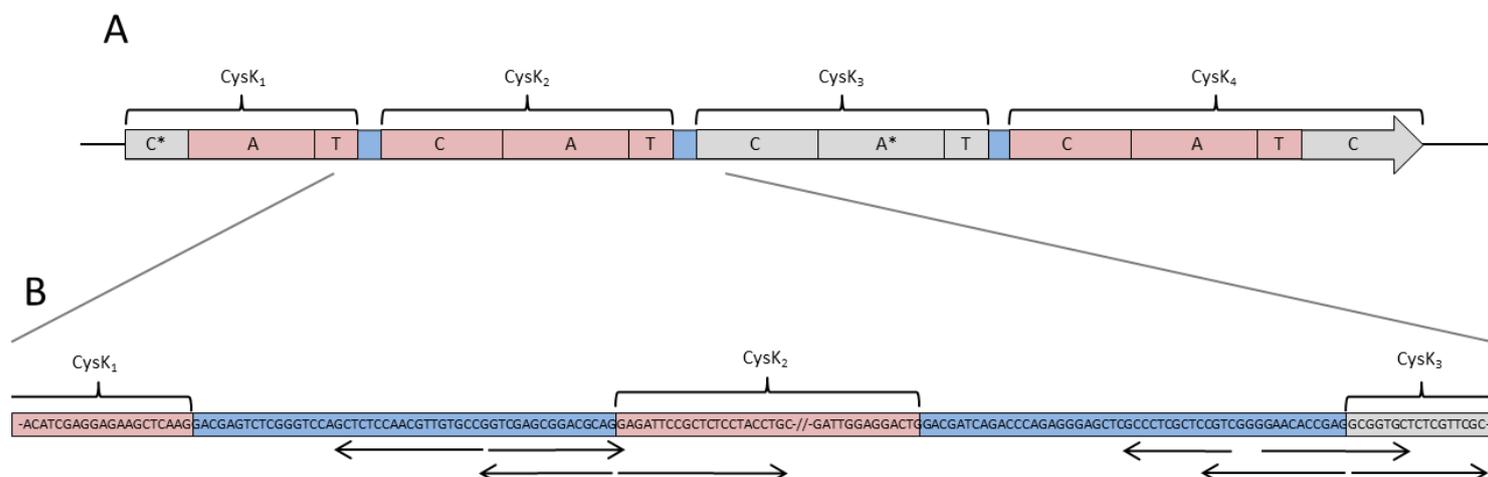


Figure 6.6: Cloning strategy used for CysK modules; A: schematic view of CysK with repetitive sequences marked in red and linkers marked in blue; B: primer binding sites in the  $K_1$ - $K_2$  and  $K_2$ - $K_3$  linkers

### 3.5 Loading Assays

*In vitro* overexpression of the third module of the assembly line CysK3 was originally carried out to provide an acceptor to the shuttling reaction by CysB, in addition the overexpression of the two last modules CysG1 and CysG2 was also performed to provide PCP bound substrate to confirm *in trans* tailoring reactions for the methyltransferases CysF and the radical SAM CysS. Nevertheless once overexpressed the individualized NRPS modules were tested to confirm the *in silico* assigned adenylation and loading specificity. (Figure 6.8-Figure 6.15) Cystobactamides are characteristic for having a backbone made of polymerized pABA units which are not proteinogenic and very uncommon in NRPS since they are encountered only in one other compound, Candicidine, apart from Cystobactamides and the closely related Albicidins.

Another interesting question was the activity of CysR, this N-dioxygenase is homologous to the thoroughly studied AurF a di-iron pABA N-oxygenase present in the Aureothine biosynthesis pathway.<sup>(16)</sup> Because of the high similarity between CysR and AurF, we originally postulated that it would carry out the same activity and use free pABA as substrate. This was however in discrepancy with the Stachelhaus code of CysK1 which is identical to CysK2 and CysK4 and the general identity of the three domains (97%) which would indicate that they are actually all activating and loading pABA which would in the case of the first module be subsequently oxidized *in trans*. This was confirmed experimentally through loading assays showing that pNBA is not a valid substrate for CysK1, indicating that the oxygenation is indeed either an *in trans* tailoring step or carried out on the final product after chain release. Surprisingly CysG1 seems to have a more relaxed substrate specificity than the other modules and is able to load pNBA.

As already described in Chapter 3, PCP bound 3-OHpABA was a valid substrate for *in trans* O-methylation by CysF it is however noteworthy that already methylated 3-OMepABA could also be loaded indicating that the methylation might happen both before or

after the loading depending on reaction kinetics of both enzymes. Unexpectedly, however, all modules were revealed to share the same substrate specificity despite differences in the Stachelhaus code between adenylation domains from CysK and CysG. The absence of any Cystobactamides containing a hydroxylated or methoxylated pABA in position one to four indicates that for an efficient biosynthesis either the tailoring would have to happen slower than the adenylation leaving it as a purely *in trans* process, or the C domains of modules 2, 3 and 4 would have to discriminate strongly against loaded hydroxylated and methoxylated pABA jamming the assembly line until the type II thioesterase could unload these residues.

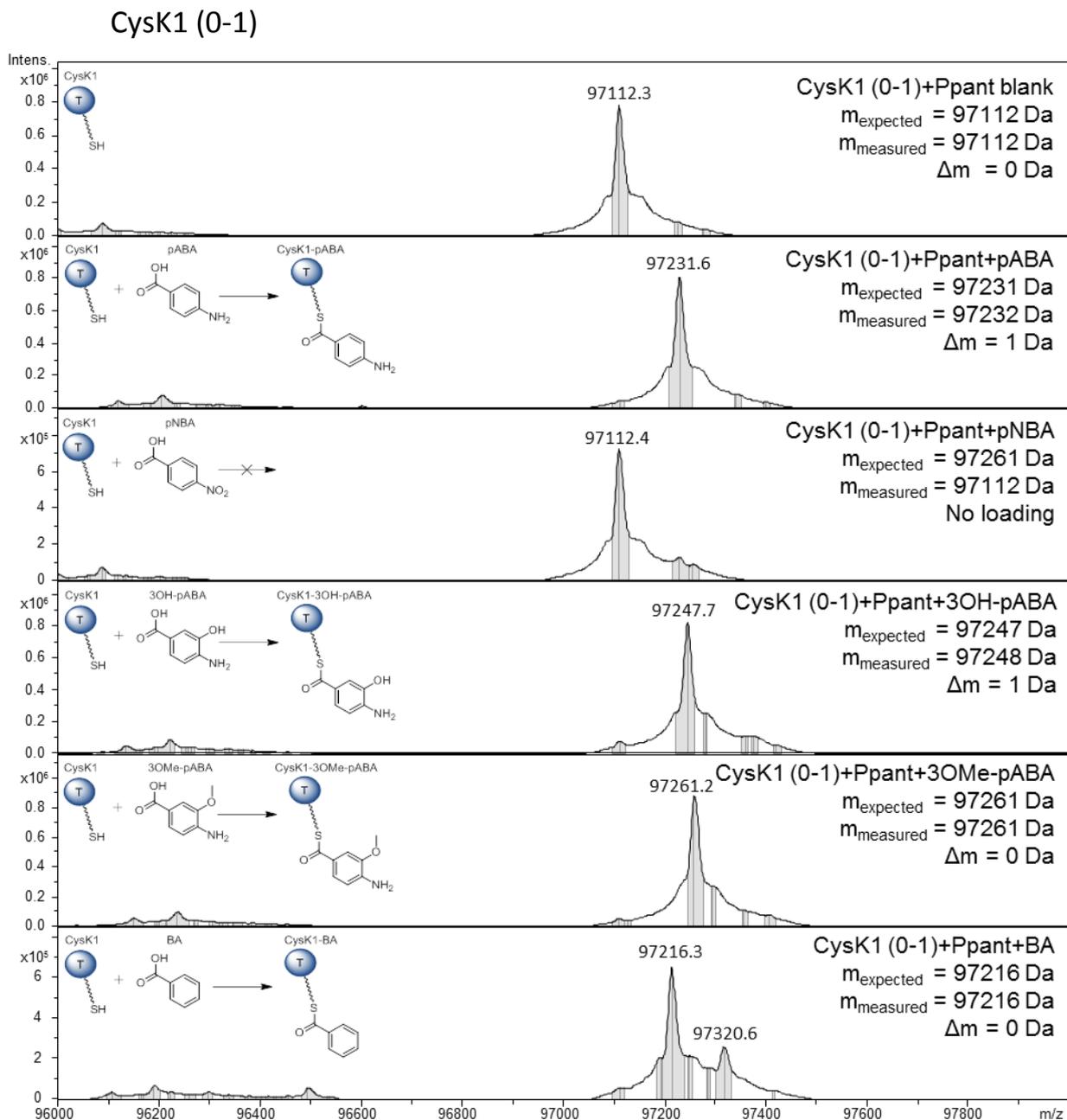


Figure 6.7: Protein MS analysis of adenylation and loading experiment for module CysK1 with a long C-terminal linker and no N-terminal linker in the presence of pABA, pNBA, 3-OHpABA, 3-OMe-pABA and BA. For the sake of simplicity only the T domain of Module CysK1 C-A-T was depicted.

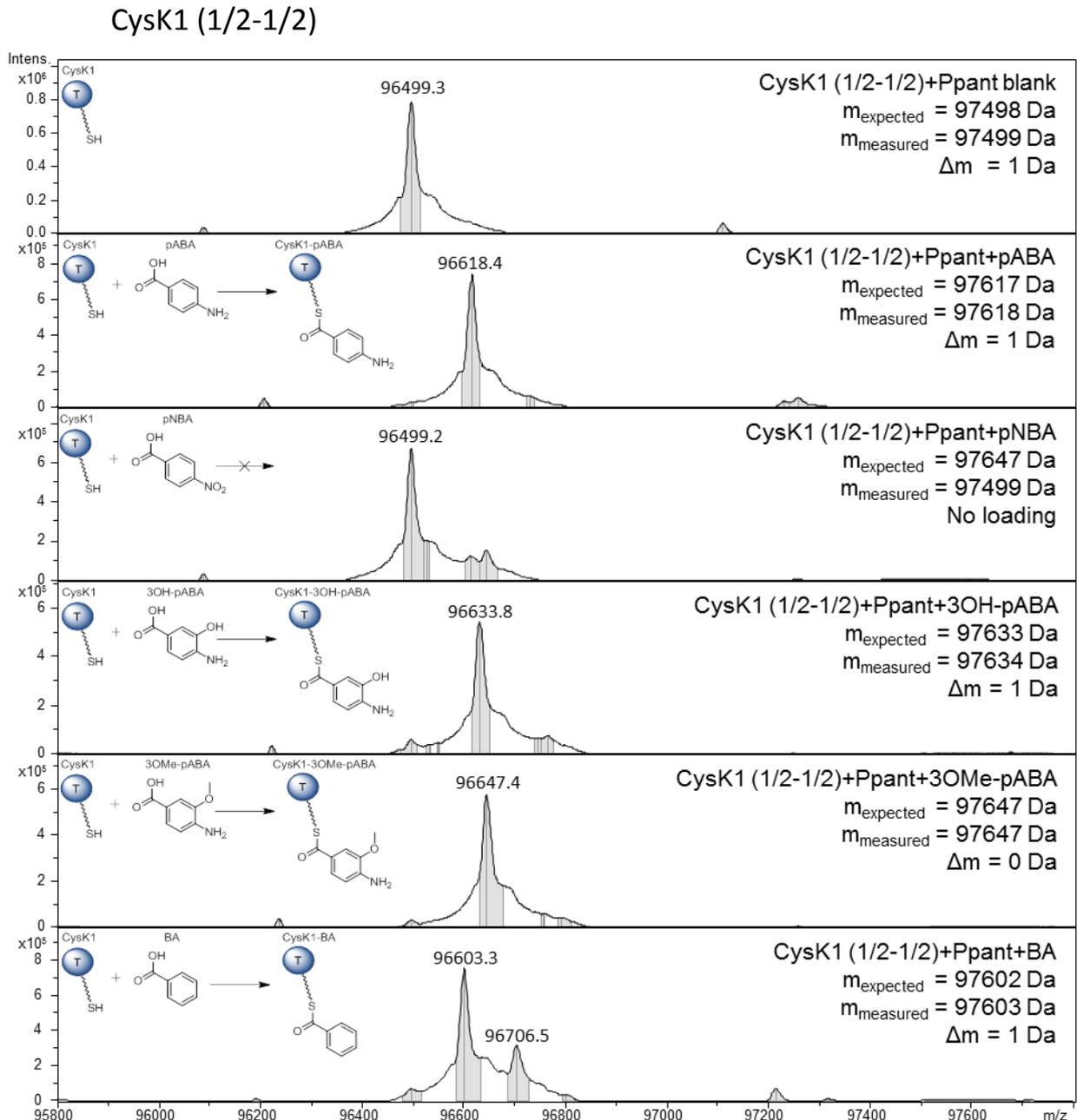


Figure 6.8: Protein MS analysis of adenylation and loading experiment for module CysK1 with half a linker on both the C and N-terminus in the presence of pABA, pNBA, 3-OHpABA, 3-OMe-pABA and BA. For the sake of simplicity only the T domain of Module CysK1 C-A-T was depicted.

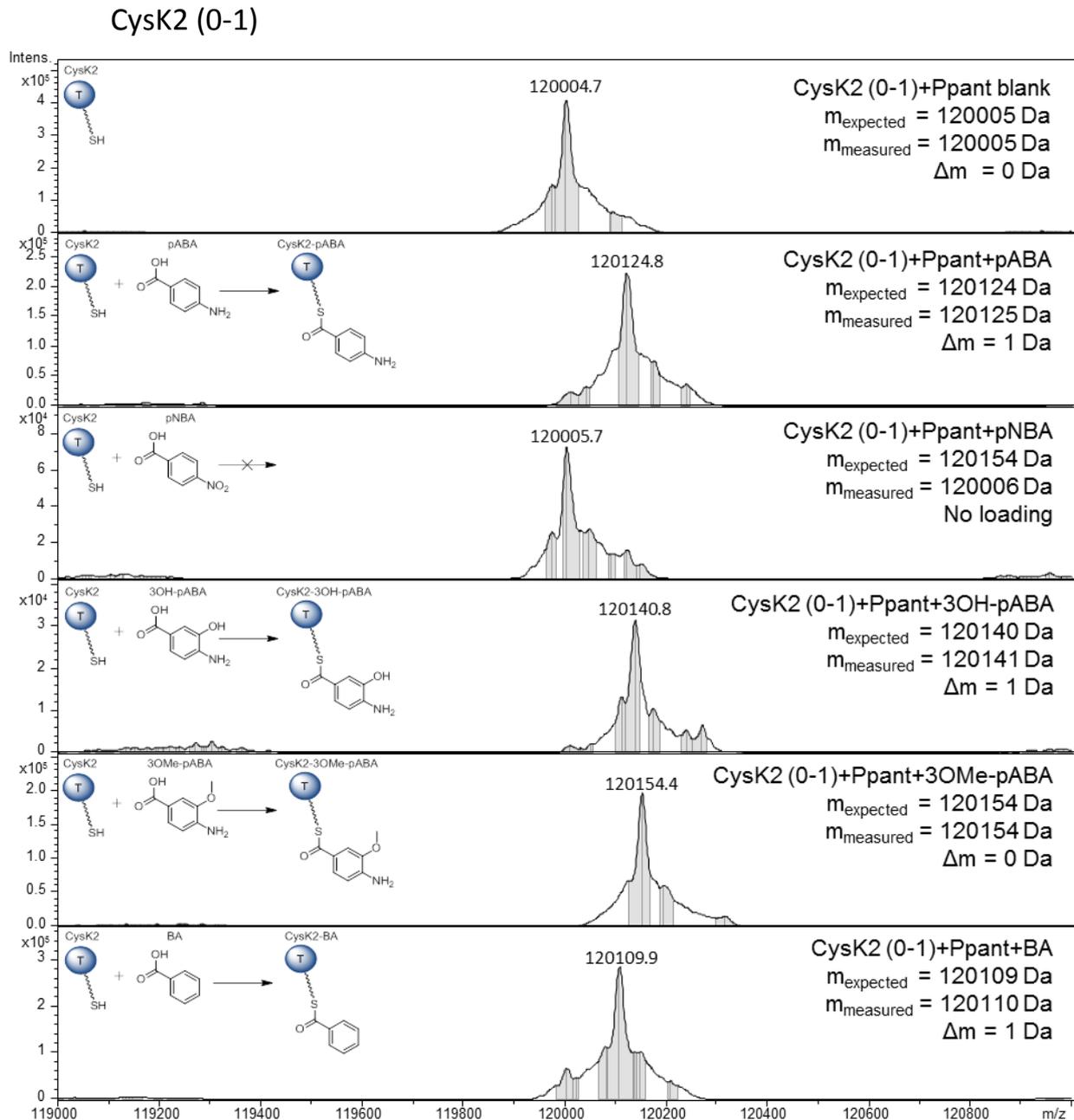


Figure 6.9: Protein MS analysis of adenylation and loading experiment for module CysK2 with a long C-terminal linker and no N-terminal linker in the presence of pABA, pNBA, 3-OHpABA, 3-OMe-pABA and BA. For the sake of simplicity only the T domain of Module CysK2 C-A-T was depicted.

CysK2 (1/2-1/2)

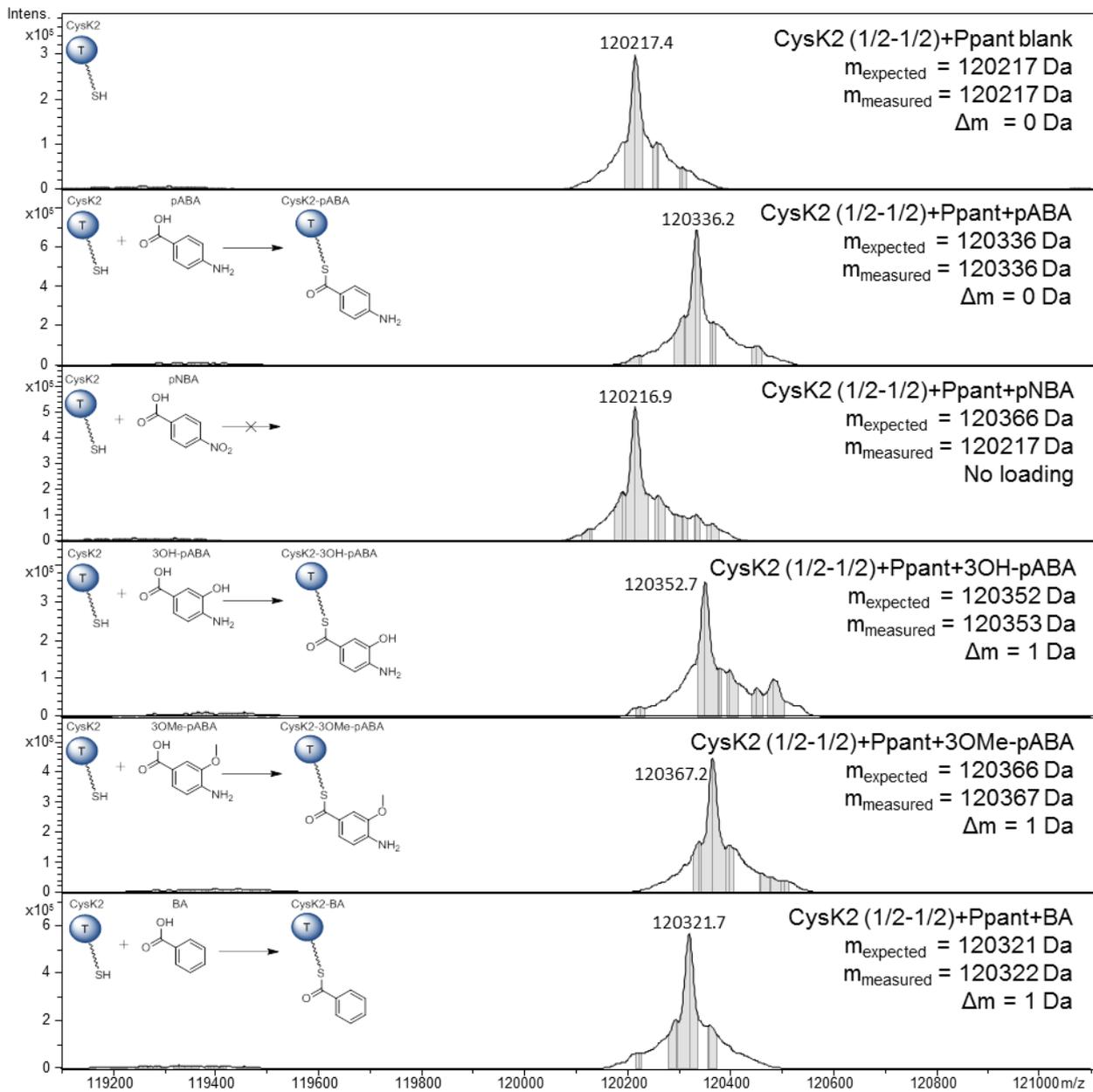


Figure 6.10: Protein MS analysis of adenylation and loading experiment for module CysK2 with half a linker on both the C and N-terminus in the presence of pABA, pNBA, 3-OHpABA, 3-OMe-pABA and BA. For the sake of simplicity only the T domain of Module CysK2 C-A-T was depicted.

## CysK4 (1/2-1/2)

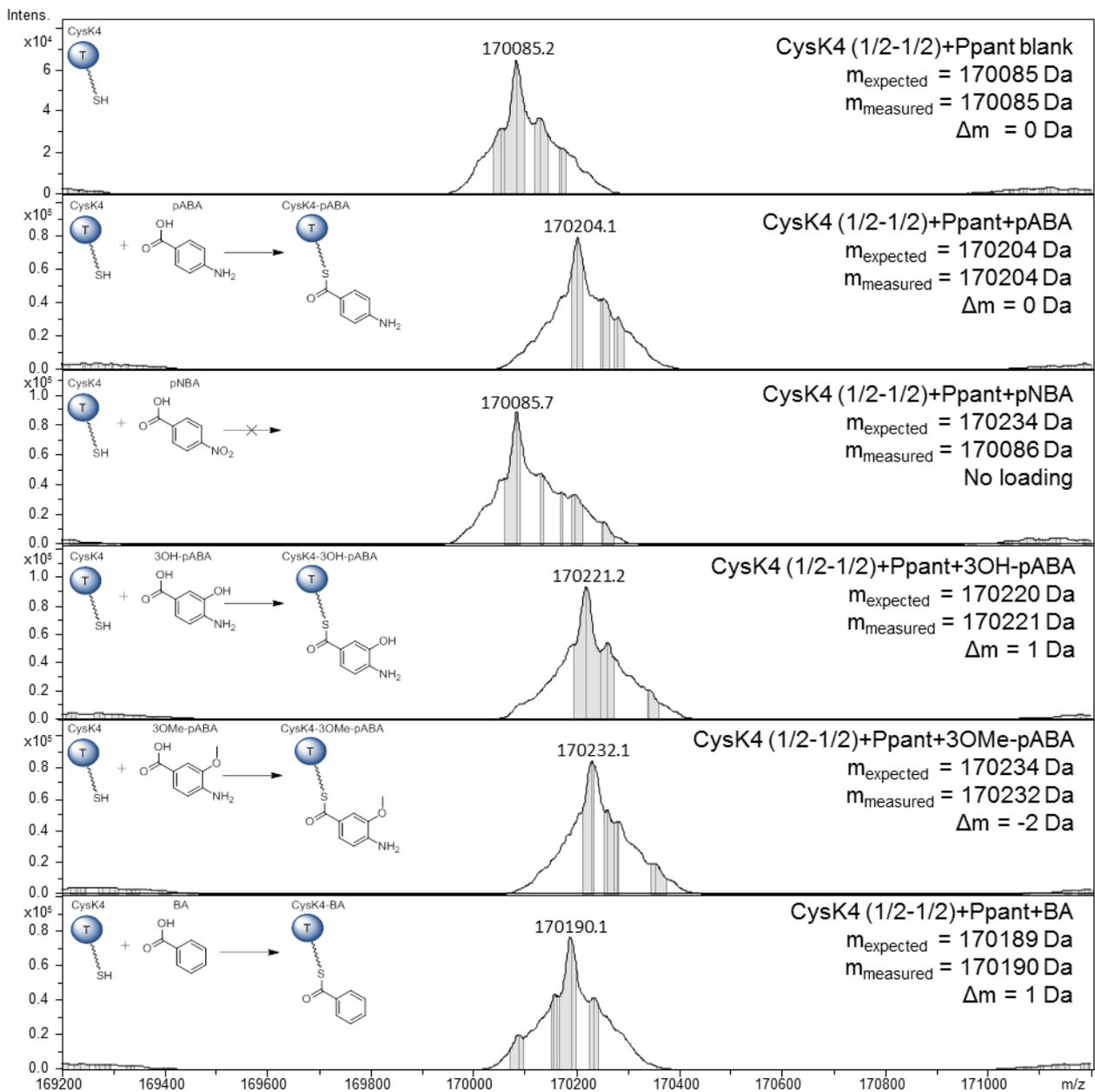


Figure 6.11: Protein MS analysis of adenylation and loading experiment for module CysK4 with half a linker on both the C and N-terminus in the presence of pABA, pNBA, 3-OHpABA, 3-OMe-pABA and BA. For the sake of simplicity only the T domain of Module CysK4 C-A-T-C was depicted.

CysG1 (0-1)

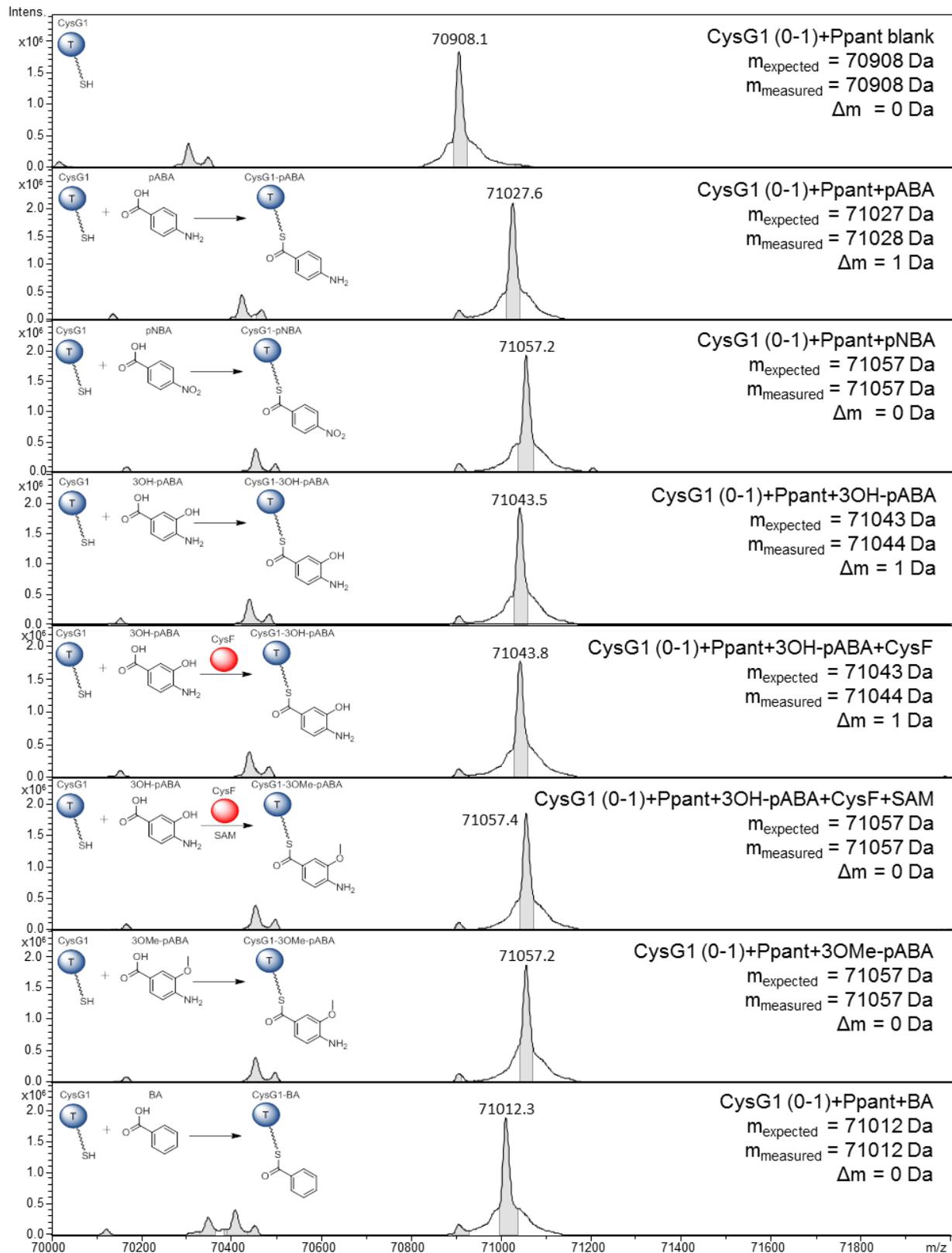


Figure 6.12: Protein MS analysis of adenylation and loading experiment for module CysG1 with a long C-terminal linker and no N-terminal linker in the presence of pABA, pNBA, 3-OHpABA, 3-OMe-pABA and BA. Additionally, the *in trans* methylation by CysF was confirmed. For the sake of simplicity only the T domain of Module CysG1 A-T was depicted.

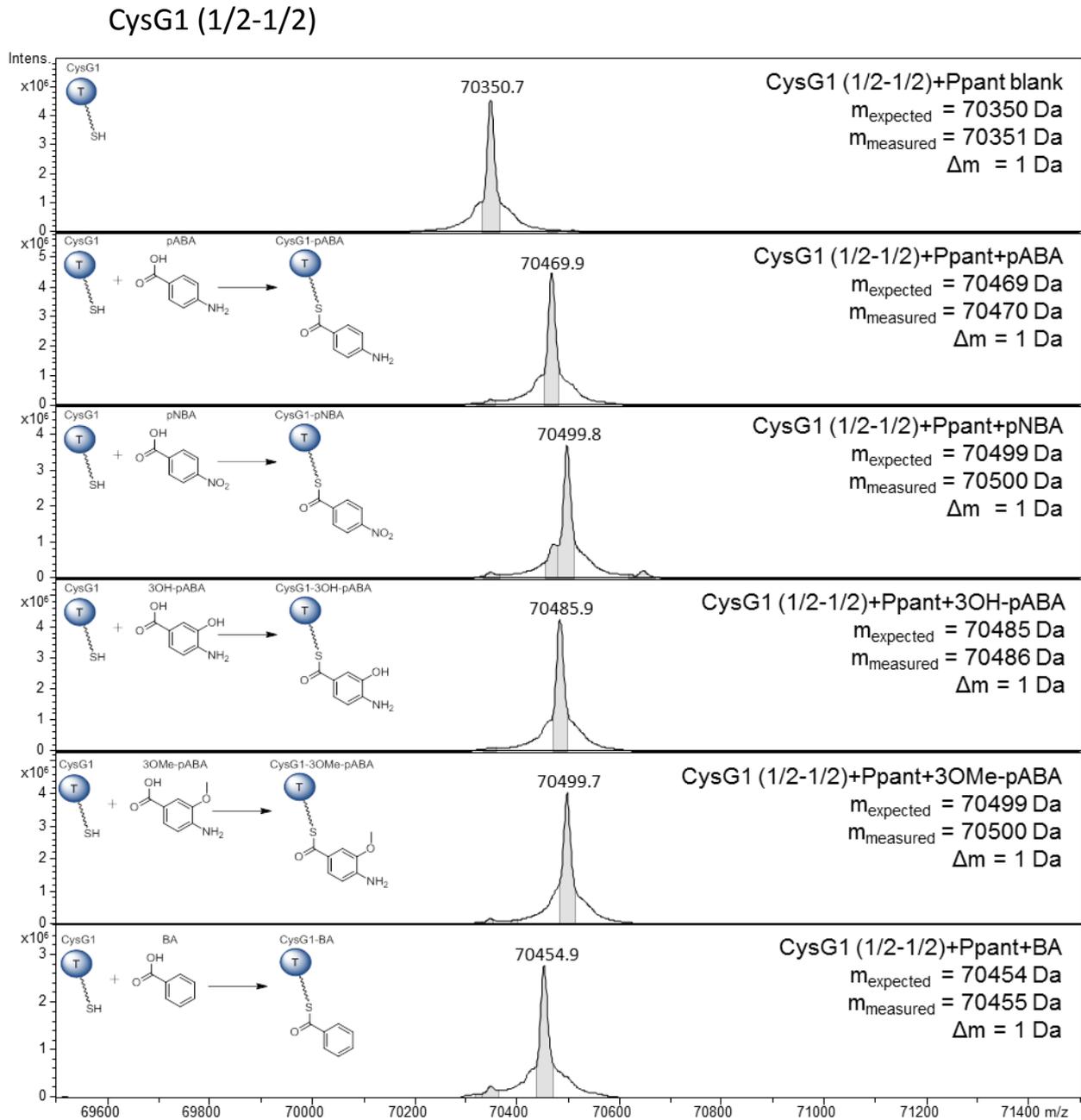


Figure 6.13: Protein MS analysis of adenylation and loading experiment for module CysG1 with half a linker on both the C and N-terminus in the presence of pABA, pNBA, 3-OHpABA, 3-OMe-pABA and BA. For the sake of simplicity only the T domain of Module CysG1 A-T was depicted.

CysG1 (0-1)

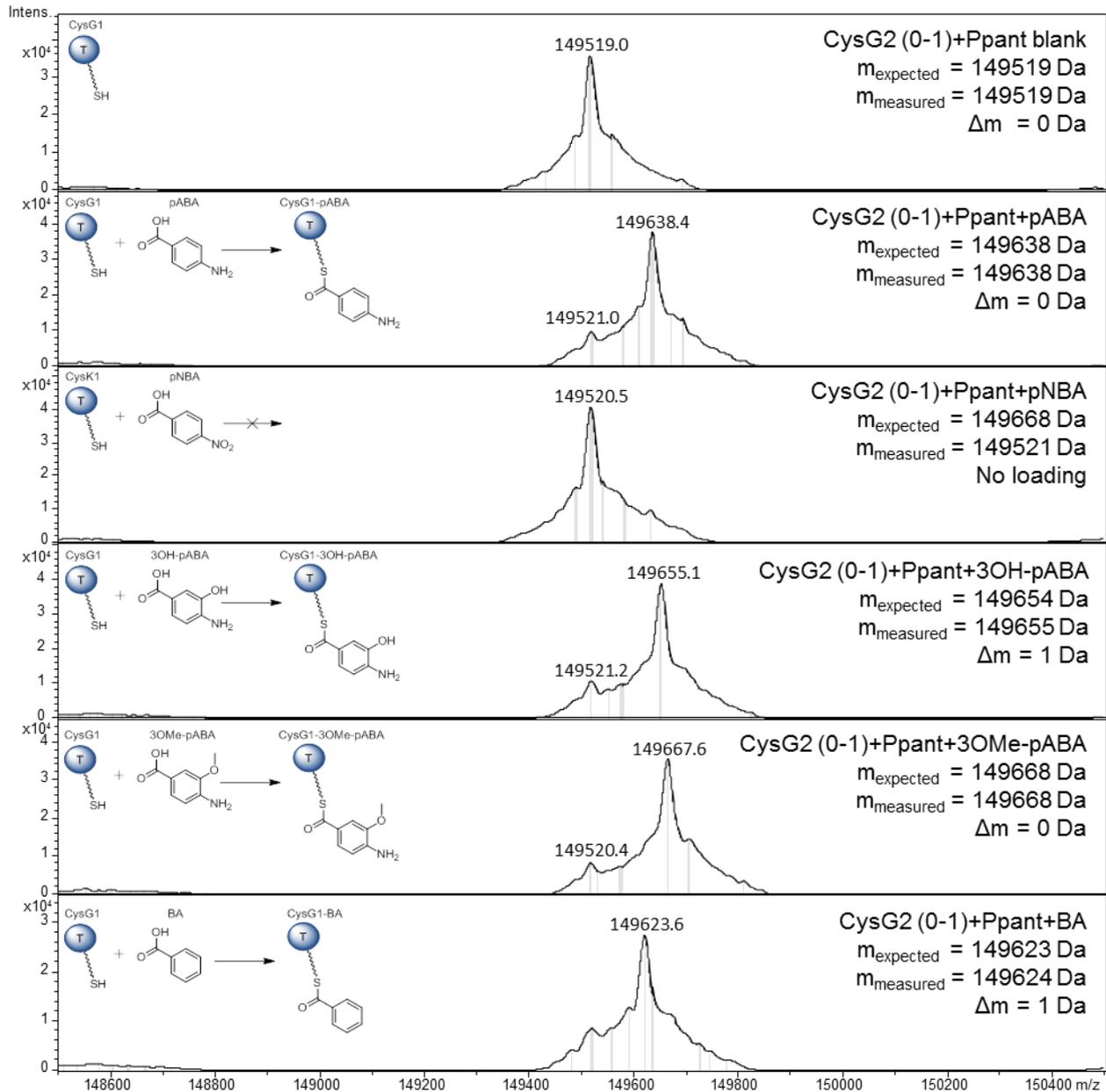


Figure 6.14: Protein MS analysis of adenylation and loading experiment for module CysG2 with no N-terminal linker in the presence of pABA, pNBA, 3-OHpABA, 3-OMe-pABA and BA. For the sake of simplicity only the T domain of Module CysG2 C-A-T-TE was depicted.

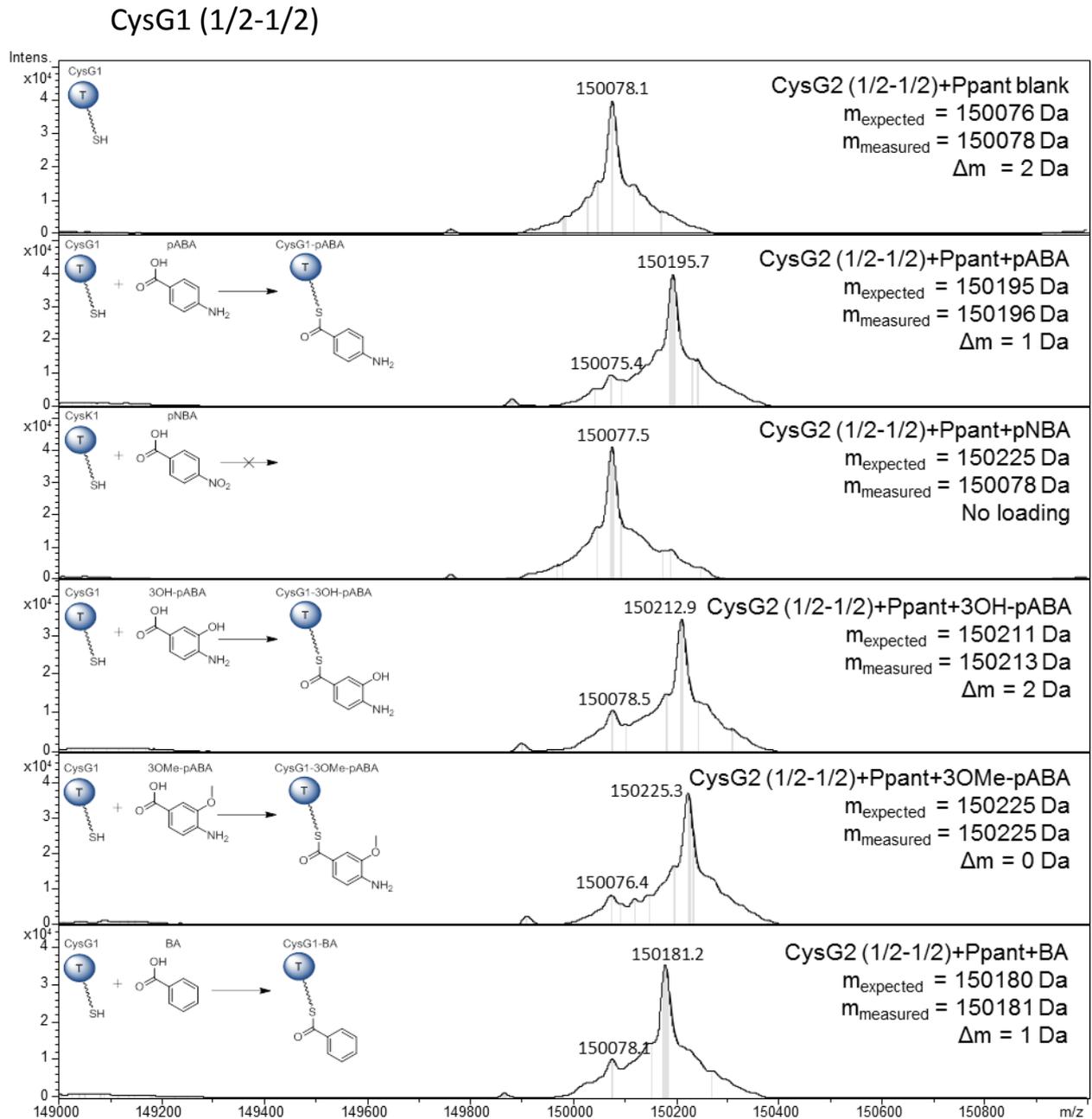


Figure 6.15: Protein MS analysis of adenylation and loading experiment for module CysG2 with half a linker on the N-terminus in the presence of pABA, pNBA, 3-OHpABA, 3-OMe-pABA and BA. For the sake of simplicity only the T domain of Module CysG2 C-A-T-TE was depicted.

### 3.6 Condensation reactions

After probing the activity of the adenylation domains *in vitro* another important point was to investigate the mechanic of the condensation domains. This biochemistry is far less well understood than the activation in the A domain, indeed there is no structural nor biochemical basis to explain the substrate specificity displayed by the condensation domains, thus *in silico* analysis of C domains is nowadays still impossible. In the case of Cystobactamides, however, it is noteworthy that the condensation domains in modules two and four share 98.9% identity, despite the fact that C2 condenses pNBA on pABA and C4 condenses the tripeptide pNBA-pABA-isoAsn on pABA.

Condensation reactions in NRPS biosynthesis follow the processivity rule which states that the reaction sequence matters and has to follow an order starting with the first module and ending with the last one.<sup>(17)</sup> Thus, the first condensation has to take place before the second where the condensation domain will accept only a dipeptide as a donor, and so on. In the case of the *in vitro* reconstitution of the Cystobactamide backbone biosynthesis, the amount of *in trans* tailoring is problematic, indeed the N-oxidation of pABA to pNBA on module one cannot be circumvented by loading pNBA directly on the module. Thus we had to resolve to SNAC esters as mimics of the loaded derivatives that can be encountered on module one and two in order to test the condensation between modules one, two and three.

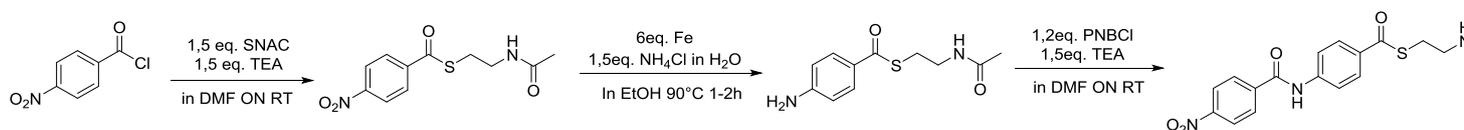


Figure 6.16: Synthesis of the pNBA-SNAC and of the pNBA-pABA-SNAC mimics

Synthesis of SNAC esters was done in a relatively simple fashion using para-nitrobenzoate chloride as starting material for the condensation with N-acetyl cysteamine, the resulting pNBA-SNAC thioester can be used both as mimic of the a loaded CysK1 module and as starting material for the synthesis of the pNBA-pABA-SNAC dipeptide CysK2 mimic through sequential reduction by metallic Fe and condensation with pNBA chloride which yielded the desired product. (Figure 6.16)

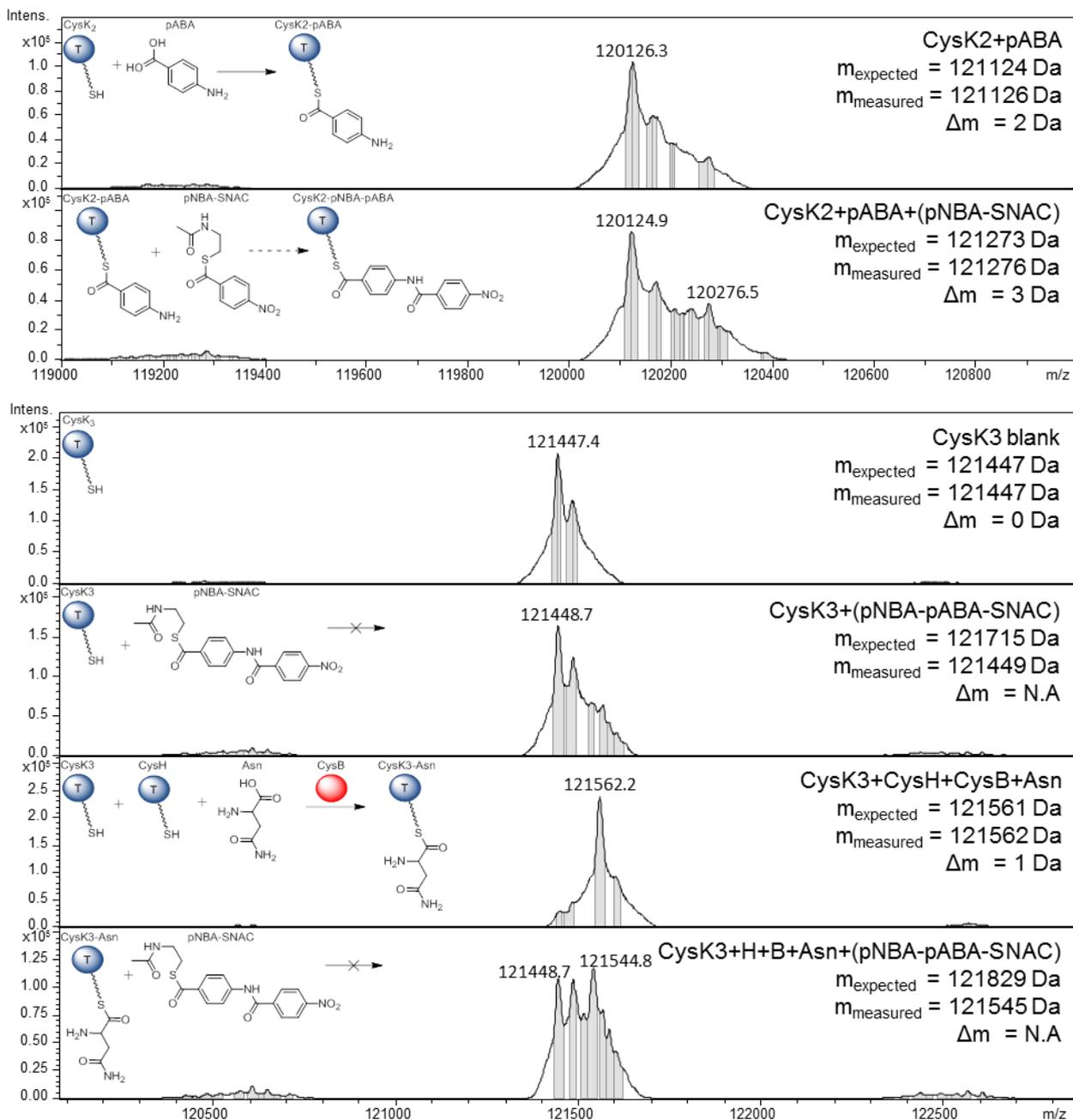


Figure 6.17: Protein MS of CysK2 and CysK3 incubated with their respective substrate in addition to donor SNAC thioester mimics. For the sake of simplicity only the T domain of Module CysK2, CysK3 (both C-A-T) and CysH (A-X-T) was depicted.

The SNAC esters were incubated in a number of conditions with their respective acceptor enzyme but no clear condensation could be observed. Protein MS analysis of the second module CysK2 after incubation with pABA and pNBA-SNAC could hint at a marginal condensation rate in specific conditions (30°C / pH 8,0) while incubation of pNBA-pABA-SNAC with CysH, CysB and CysK3 lead to a strong decrease in the quality of the signal with peaks for unloaded CysK3 showing again which could mean that the SNAC ester interferes with the rather sensitive asparagine shuttling process. (Cf. chapter 4) Furthermore while heterogeneity in the protein population regarding derivatives of the same large protein

with mass differences in the range expected here ( $\pm 114$  m/z for asparagine and  $\pm 268$  m/z for the pNBA-pABA dipeptide) should allow resolving of the spectra after deconvolution, this would still have an impact on the quality of the deconvoluted spectrum with broadened peaks having low mass precision.

## 4 Conclusion

We successfully managed to overexpress the Cystobactamide backbone assembly line as separate modules using a coexpression system with CysA. This emphasizes the role of MbtH type A domain activators as indispensable binding partners of NRPS modules, it is not yet clear however why this sequence is not integrated directly in the primary sequence of A domains or whether it plays a role in modifying the substrate specificity of the A domains or whether it has a function in regulating the activity of NRPS assembly lines. The type II thioesterase CysE proved to be a convenient and efficient tool to hydrolyze phosphopantetheine bound substrate and recover “clean” NRPS modules in holo-form. The NRPS modules proved to be active *in vitro* and could be used as the substrates for various *in trans* tailoring activity assays. The adenylation specificity originally assigned *in silico* was confirmed experimentally indicating that the terminal N-oxygenation takes place after loading of para-amino benzoic acid. However only a hint of *in vitro* condensation activity could be obtained by using SNAC ester and nothing was observed using full-sized modules, due the lack of literature regarding condensation domain biochemistry it is unclear whether specific reaction conditions, or maybe an unknown cofactor, would be needed for a full reconstitution of the backbone biosynthesis *in vitro*.

## 5 References

1. Patel, H. M., and Walsh, C. T. (2001) In vitro reconstitution of the *Pseudomonas aeruginosa* nonribosomal peptide synthesis of pyochelin: characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities, *Biochemistry* 40, 9023–9031.
2. Gaitatzis, N., Kunze, B., and Müller, R. (2001) In vitro reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases, *Proceedings of the National Academy of Sciences of the United States of America* 98, 11136–11141.
3. Lowry, B., Robbins, T., Weng, C.-H., O'Brien, R. V., Cane, D. E., and Khosla, C. (2013) In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase, *Journal of the American Chemical Society* 135, 16809–16812.
4. Finking, R., and Marahiel, M. A. (2004) Biosynthesis of nonribosomal peptides1, *Annual review of microbiology* 58, 453–488.
5. Bloudoff, K., and Schmeing, T. M. (2017) Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity, *Biochimica et biophysica acta* 1865, 1587–1604.
6. Belshaw, P. J. (1999) Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis, *Science* 284, 486–489.
7. Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011) NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity, *Nucleic acids research* 39, 7.
8. Challis, G. L., Ravel, J., and Townsend, C. A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains, *Chemistry & biology* 7, 211–224.
9. May, J. J., Kessler, N., Marahiel, M. A., and Stubbs, M. T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases, *Proceedings of the National Academy of Sciences of the United States of America* 99, 12120–12125.
10. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A., and Essen, L.-O. (2007) Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase, *Structure (London, England : 1993)* 15, 781–792.
11. Chen, W.-H., Li, K., Guntaka, N. S., and Bruner, S. D. (2016) Interdomain and Intermodule Organization in Epimerization Domain Containing Nonribosomal Peptide Synthetases, *ACS chemical biology* 11, 2293–2303.
12. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction, *Nature methods* 12, 7–8.
13. Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C., and Gulick, A. M. (2016) Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture, *The Journal of biological chemistry* 291, 22559–22571.

14. Herbst, D. A., Boll, B., Zocher, G., Stehle, T., and Heide, L. (2013) Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes, *The Journal of biological chemistry* 288, 1991–2003.
15. Buntin, K., Weissman, K. J., and Müller, R. (2010) An unusual thioesterase promotes isochromanone ring formation in ajudazol biosynthesis, *Chembiochem : a European journal of chemical biology* 11, 1137–1146.
16. Choi, Y. S., Zhang, H., Brunzelle, J. S., Nair, S. K., and Zhao, H. (2008) In vitro reconstitution and crystal structure of p-aminobenzoate N-oxygenase (AurF) involved in aureothin biosynthesis, *Proceedings of the National Academy of Sciences of the United States of America* 105, 6858–6863.
17. Stachelhaus, T., Mootz, H. D., Bergendahl, V., and Marahiel, M. A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain, *The Journal of biological chemistry* 273, 22773–22781.

## Chapter 7: Discussions

### 1 General summary of this work

The aim of this work was the elucidation of the biosynthesis of Cystobactamides using a biochemical approach with *in vitro* characterization of individualized biosynthetic protein with the ultimate goal of a complete *in vitro* reconstruction of the biosynthesis. After heterologous expression of most of the biosynthesis related genes in the Cystobactamide cluster, this goal was mainly achieved and we were able to shed light on the major part of the biosynthetic processes leading to Cystobactamides. The thoroughly studied primary metabolism proteins related to the biosynthesis of pABA (1) present in the cluster were voluntarily left out of this study since the *in silico* prediction regarding their activity is already very reliable and no discrepancy could be observed on this level with characterized pABA biosynthesis enzymes.

The “first” enzyme of the pathway is a pABA-CoA ligase (CysL) which branches pABA from the folate pathway into the Cystobactamide biosynthesis. It could be characterized and was shown to be prone to strong feedback inhibition by Cystobactamides probably mediated by a secondary domain. While feedback inhibition is a well-described process in primary metabolism (2) and has been suspected to exist in a number of secondary metabolism pathways it has only rarely been experimentally characterized in natural product biosynthesis.(3)

Even if methylation is an abundant tailoring modification of secondary metabolites (4) the methylation pattern of Cystobactamide is especially impressive since no less than three different methyltransferases were assigned to the cluster. CysF, a regular SAM-dependent methyltransferase, and CysS, a radical SAM C-methyltransferases, were fully characterized and shown to operate in sequence *in trans* to yield highly unusual isopropoxyl decorations on the two last pABA building blocks.

NRPS pathways are known for the incorporation of a wide variety of non proteinogenic amino acids (5) but they are often dispersed in a backbone of regular proteinogenic L configured amino acids, while Cystobactamides might seem to lack in variety of their building blocks since five out of six are derived from pABA, one has to take into account that pABA can be found only in two other natural products pathways (6,7) and that the last amino acid in an unprecedented highly modified isoasparagine moiety. We were able to characterize a new *in cis* aminomutase tailoring X domain responsible for the isomerisation of asparagine and an unusual aminotransferase performing a shuttling from the standalone NRPS module CysH where the tailoring is performed to the assembly line where the peptidic backbone is assembled.

Additionally, CysP which had previously been miss-assigned as tailoring enzyme could be shown to be a secondary Cystobactamide resistance protein along with the known

pentapeptide repeat protein CysO. CysP seems to work through sequestration of the antibiotic in a similar fashion as what is observed for the homologous bleomycin resistance protein.(8)

While the challenging heterologous expression of the complete assembly line was achieved through the use of further proteins of the cluster as tools for the expression and purification of the NRPS modules, only the adenylation and thiolation activities of the modules could be observed. No condensation could be detected *in vitro* which was assumed to be related to the unusually low reactivity displayed by the acid function of pABA. However, because the modules are now readily available, further studies might allow determining the conditions required by this biochemical process which could provide further insights in the biochemistry of condensation domains.

## 2 Cystobactamide biosynthesis: a canonical NRPS assembly line?

The biosynthesis of Cystobactamides is performed on a regular linear type A NRPS assembly line spread over two proteins CysK and CysG, featuring 6 modules as expected for the assembly of a hexapeptide. However, an additional standalone NRPS module, CysH, can be observed which is bending the collinearity rule (9) since two NRPS modules work hand in hand for the integration of a single amino acid. Indeed, CysH can mostly be seen as a module 3' the adenylation domain of which is making up for the inactive adenylation domain in module 3. The opportunity of activating and loading of the amino acid on a parallel platform still remains unclear but this phenomenon seems to be quite common when complex tailoring is required, especially in the case of oxidative reactions.(10) The use of extremely reactive superoxide radicals intermediates in  $\beta$ -hydroxylation and the risk of off-target reactions that they imply could explain the existence of these standalone NRPS platforms to avoid detrimental oxidations to happen to the assembly line.(11)

Additionally, Cystobactamides feature an unprecedented isoasparaginyl moiety originating in asparagine. The isomerization reaction leading to this residue is performed by a newly discovered dehydratase/mutase domain featured by CysH. This domain features a dual activity dehydratase/mutase depending on the hydroxylation state of the asparaginyl substrate. CysH is closely related to AlbIV from the Albicidin biosynthesis which displays the dehydratase activity since no oxygenase is present in this cluster.(12) Both reactions are hypothesized to be based on the same biochemistry with activation of either the amide oxygen or the hydroxyl oxygen for abstraction leading to a radically different reactivity depending on the bound substrate. It is noteworthy that while hydroxy-isoasparagine was encountered for the first time in Cystobactamide and  $\beta$ -cyanoalanine for the first time in Albicidin, a number of X domain homologues could be discovered in the NCBI database, meaning that these moieties might be more widespread than originally thought. The activity of Cystobactamide is, however, in discrepancy with tailoring since the major compounds carrying isoasparagine which are therefore sustaining a full tailoring seem less active than asparagine carrying Cystobactamides with incomplete tailoring.(13,14) It is however not

## 2 Cystobactamide biosynthesis: a canonical NRPS assembly line?

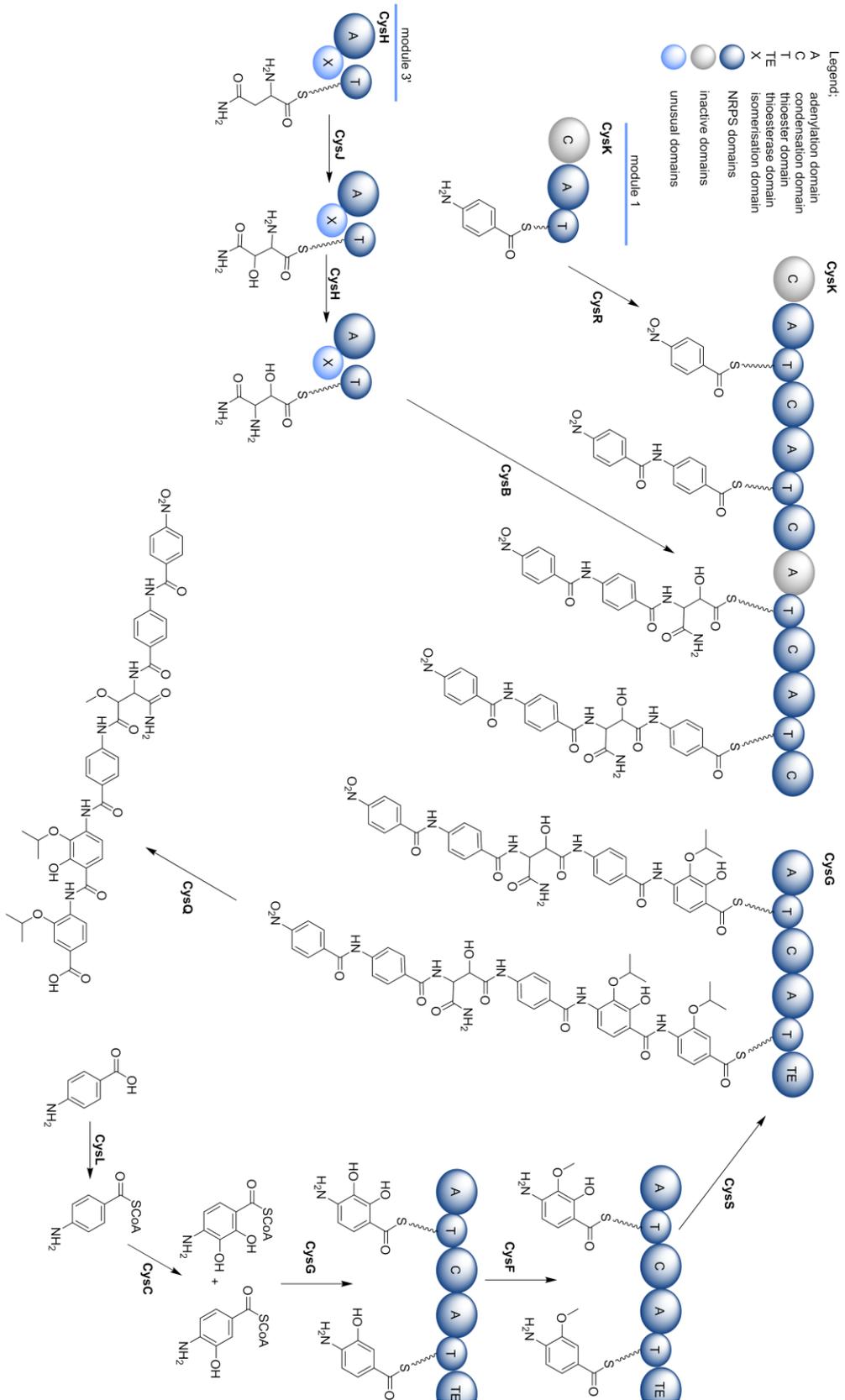


Figure 7.1: General scheme of the Cystobactamides biosynthesis

impossible that the antibiotic spectrum of both Cystobactamide types might be different.

When it comes to the linking of the stand-alone NRPS module CysH which the assembly line CysK, Cystobactamides do not follow the usual route since most of the time tailored amino acids are released from this parallel standalone A-T NRPS module by various mechanisms to create a pool of substrate that is independently activated on the actual assembly line.(10) However, in the case of Cystobactamide the adenylation on module 3 of the assembly line is impossible, as a result, the tailored hydroxy-isoasparagine has to be shuttled over an aminoacyl transferase, CysB. This phenomenon was only observed twice before, in the Syringomycin (15) and in the Coronatine (16) biosyntheses, while this process is more energy efficient, CysB being ATP independent, it results in the bypassing of the adenylation domain gatekeeping and has a very relaxed specificity which could explain why two series of Cystobactamides coexist with complete or incomplete tailoring of the asparagine linker.

Natural products are known for the variety of the tailoring reaction that can happen on a scaffold of simple building blocks, these reactions allow increasing of the chemical diversity of secondary metabolites. In the Cystobactamide biosynthesis the tailoring is remarkably diverse with no less than seven enzymes. While the tailoring reactions often take place on simple precursors that will later be integrated into the backbone or after the release of a complete natural product, in the case of Cystobactamides most of them are happening *in trans* on carrier protein bound substrates. It is likely that the activation of the aromatic ring of pABA by the thioesterification of the acid lowers the energy required for some of these reactions.

The Cystobactamide biosynthesis is remarkable for its extensive use of the rarely observed non proteinogenic pABA building blocks.(17) While NRPS are notorious for utilizing non-proteinogenic amino acids, these building blocks are still usually derived from the 20 proteinogenic amino acids and are hence still  $\alpha$ -amino acid or more rarely  $\beta$ -amino acids. On the other hand pABA is derived from the folate biosynthesis and is not strictly speaking an amino acid but rather an aminated aromatic acid, hence the delocalization of the electrons over the aromatic ring from the acid to the amine will have a significant impact on the reactivity of both chemical functions, which could interfere with the usual biochemical characteristics of NRPS systems.

The first significant difference with typical NRPS systems is the apparent pre-activation of pABA substrate as CoA in a similar fashion as what is observed for PKS biosynthesis.(18) However, the activity of AT domains is very different from A domains since they are serine protease homologues working through ping-pong mechanism (19) rather than an AMP ligase.(20) But since A domains have been shown to work as CoA ligases in absence of a carrier protein (21), the reversibility of this reaction could lead to a similar ping-pong mechanism where the serine intermediate is replaced by an adenylate intermediate. (Figure 7.2) While NRPS-PKS hybrids exist, the hybridization is happening on module level with assembly lines harboring both NRPS and PKS modules in sequence (22), while in this particular case hybridization on domain level with a C-AT-PCP architecture would seem

more efficient from a biochemical perspective, the evolution of such modules seems highly unlikely. The only possibility would arise with a type II trans-AT systems with a CP domain associated with standalone acyltransferase domain and condensation domain. Even if benzoic acids are not typically substrates for elongation in PKS or NRPS biosyntheses they can be found as starter units in natural products such as enterocin (23) and aureothin (24) were they are indeed activated by a CoA ligase prior to transfer on the first ACP by a dedicated AT domain. In the Albicidin assembly line, which is closely related to the Cystobactamide biosynthesis, a PKS starter module working in a similar fashion can be found activating para-hydroxybenzoic acid which then sustains a PKS elongation cycle to form para-coumaric acid.(12) It is possible that the CoA ligase is an evolutionary remnant present in the common ancestor of both pathways and while adenylation domains of the Cystobactamide and Albicidin assembly line do not show any remarkable difference with canonical A domains they might have adapted slightly into favoring more reversible reactions allowing for the postulated ping-pong mechanism to happen.

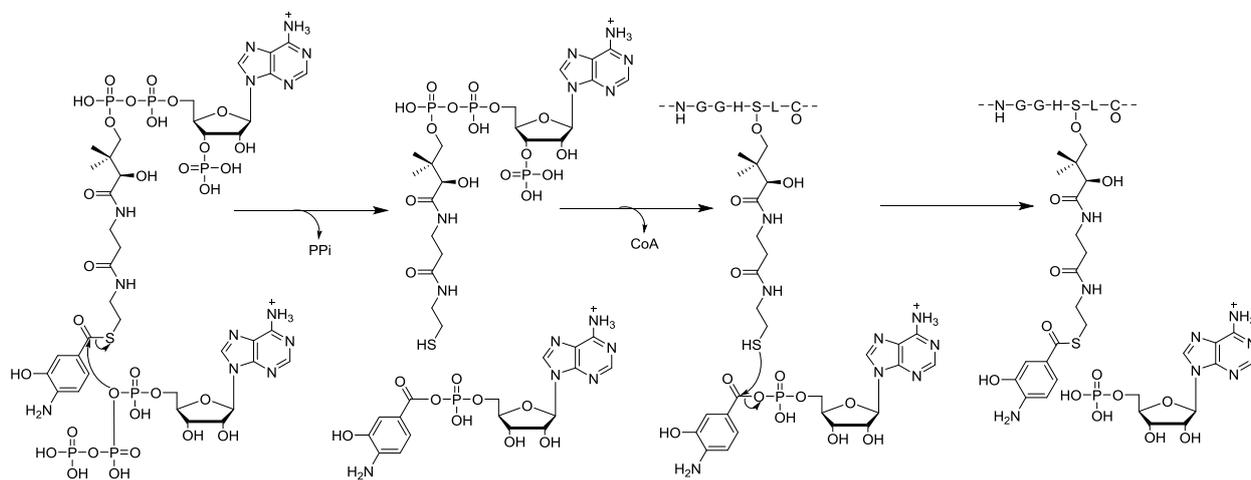


Figure 7.2: Hypothetical adenylation and thiolation activities in the Cystobactamide

The exact biochemistry leading to amide bond formation in condensation domains is still not very well understood. While a consensus HHxxxDG motif, where the second histidine would be considered the catalytic base activating the amine for nucleophilic attack, could be identified as necessary for catalysis (25), some examples can be found where the mutation of this position had no effect on the activity of the domain.(26,27) Since the activation of the acid as thioester would be sufficient for amide bond formation as long as the amine is in a non-ionized form allowing for nucleophilic attack, it is possible that the stabilization of both substrates in the active site in a position favoring the reaction could be sufficient for catalysis, this is however dependent on the inherent reactivity of the particular acid and amine involved in the reaction. It could explain the contradictory results of mutational studies regarding the importance of the second histidine as a catalytic residue. It is also noteworthy that condensation domains are not a coherent group of proteins and at least six different subgroups of C domains can be observed which catalyze slightly different

reactions.(28) While  ${}_L C_L$  and  ${}_D C_L$  are the two main subgroups and catalyze respectively the condensation of an L- or a D-amino acid donor on an L-amino acid acceptor, the starter C domain uses more variable donor substrates which are not always carrier protein bound. Epimerization domains form another subset of C domains which catalyze the epimerization of the  $\alpha$  carbon through a reversible removal of the  $\alpha$ -hydrogen leading to D configured amino acids. The two last subgroups present both a condensation domain activity and a secondary activity, these are the cyclisation domain which leads to thioazoline/oxazoline ring formation and the dual C/E domains performing both the epimerization of the donor amino acid and the subsequent condensation. While *in silico* sequence analysis of these domains reveal variability in certain conserved motifs, these data are not correlated with biochemical or structural analyses due to the difficulty in the expression and even more in the crystallization of these domains.

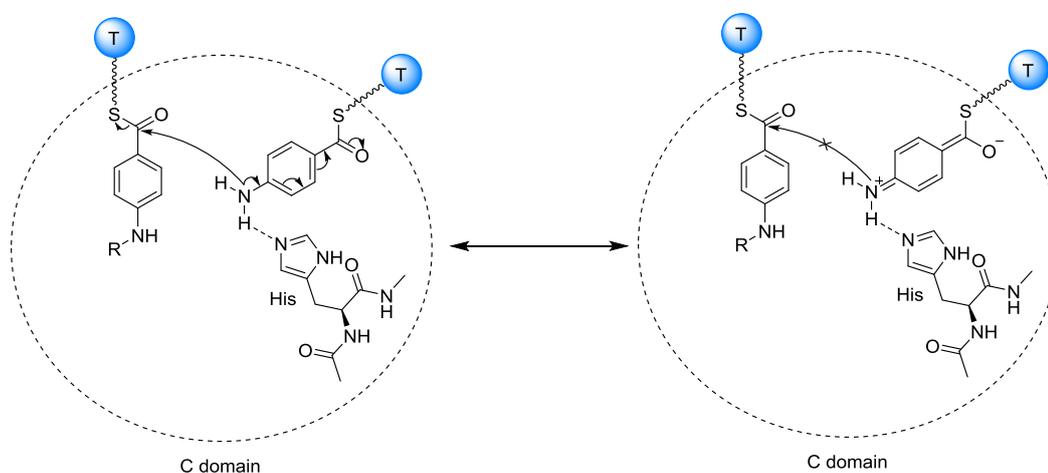


Figure 7.3: The condensation reaction involving pABA is possible in only one of the mesomeric states the molecule can take

As already mentioned the Cystobactamide assembly line makes extensive use of modified pABA moieties which are involved in every condensation step as donor substrate or acceptor substrate or even in both positions. It is surprising that no significant differences were observed *in silico* on the sequence of the condensation domains since their binding pockets have to be adapted to substrates with a much longer distance between the acid and the amine function compared to  $\alpha$  amino acids. Furthermore, the chemical synthesis of pABA SNAC esters showed that extremely harsh conditions were necessary for the thioester formation since an acyl chloride had to be heated at 60°C overnight to bring the reaction to completion. This suggests that the amine function of pABA is deactivated, probably due to the conjugation over the aromatic ring with the acyl in *para* position, implying in this specific case that catalysis probably cannot simply rely upon the proximity between the activated acyl thioester and the donor amine. It is also a possible explanation for the lack of observed *in vitro* condensation activity while the adenylation and thiolation activities could be observed. (Figure 7.3) A further troubling point is the very high sequence similarity displayed by modules 1, 2 and 4 (99,7 %), while this is understandable for the adenylation domains which

share the same substrate, it is surprising in regard of the condensation domains present in module 2 and 4 which catalyze respectively the condensation of pNBA on pABA and the condensation of the pNBA-pABA-IsoAsn tripeptide on pABA in the second case. This could imply that the substrate specificity is stringent only on the acceptor side which would be in discrepancy with previous studies (29), additionally while it is possible that the substrate positioning in the active site is mainly operated by stabilization of the phosphopantetheine arm rather than of the amino acid itself this is in contradiction with the low reactivity of the amine which would imply a stronger importance of acceptor and donor substrate positioning. It is however difficult to draw any conclusion given the lack of structural data about the binding of substrate aminoacids in the condensation domain.

## 3 General discussions

### 3.1 Challenges in the *in vitro* study of NRPS machinery

In order to understand a complex system such as NRPS biosynthesis, the most straightforward path is to take it apart and examine in detail all of its components to determine their individual function and the way they interact together. For this goal, the heterologous expression and *in vitro* analysis of these different components through biochemical and structural studies is a method of choice but a certain number of challenges have to be overcome in order to complete such a complex task. Indeed if the analogy with assembly line type machinery is fitting for NRPS systems they are nevertheless biological systems which are highly adapted to a specific environment and a myriad of factors can and will have an impact on them.

The first pitfall lies directly in the first step of such a study with the heterologous overexpression of the target enzymatic machinery, indeed since expression rates for secondary metabolite biosynthesis clusters are usually quite low in standard lab conditions (30), the first step to obtain enough material to study is to transfer the genes in a more controlled host to increase to expression levels. But while the tools for cloning, overexpression, and purification in these systems are nowadays quite well established making the heterologous host more suitable for expression in high yields (31,32), the proteins are nevertheless the fruit of a long evolution in their original producer which is bound to be the most suitable host for correct folding and stability. In the specific case of myxobacterial biosyntheses protein expression, the host of choice is usually *E.coli* which is easy to manipulate but quite limited when it comes to the overexpression of very large constructs such as NRPS proteins hundreds of kDa long.(33) Since historically a lot of natural products were isolated from *Streptomyces* the bulk of the biosynthesis research was performed on proteins originating from this family which prompted the development of *Streptomyces* type strains (34), however this kind of type strains adapted for protein overexpression are still lacking for myxobacteria. In this study, we chose to separate modules in order to overcome the size problem which allowed retaining sufficient yields. On the other hand while the

structure of NRPS modules is known, multimodular proteins are not as readily available and the examination of intermodular interactions at play in the assembly line still remains to be done. Furthermore the quaternary structure of NRPS assembly lines is still open for discussion, while recent cryoEM structural analysis of the two modules long DhbF revealed a flexible monomeric architecture (35), one has to take into account that the bacillibactin assembly line is an iterative type B NRPS and whether these results are transferable to linear type A NRPS is debatable. In the absence of a quaternary structure similar to type I PKS dimeric helix (36,37) which would order the modules in a linear fashion the flexibility of a monomeric assembly line should allow for carrier proteins from most of the assembly line to interact in any C domain and not only on the directly upstream and downstream positions.

A further unsuspected factor plays a critical role in the heterologous expression of NRPS modules which is the coexpression of the adenylation domain activator in the heterologous host.(38) While it is not clear whether this small protein (8 kDa) is a regulator of the adenylation activity, a chaperon-like protein or whether it might prevent degradation of the module by proteases, crystal structures show a tight interaction of MbtH type proteins with the adenylation domain.(39) Adenylation domain activators do not seem to present any catalytic activity but have been shown repeatedly to be necessary for the overexpression of NRPS modules and recently published crystal structures of full size modules all contain an associated MbtH protein.(40)

Post transcriptional modifications are critical in the function of proteins and are often a limitation of heterologous hosts, in the case of NRPS and PKS systems the phosphopantetheinylation reaction is indispensable for the activity since it provides the phosphopantetheine arm which will bring the substrate in the various active sites of adjacent domains. While two types of phosphopantetheine transferases are described in literature for *E.coli*, AcpS being related to the FAS and primary metabolism and EntD being involved in secondary metabolism (41), they never seemed to pose a problem for overexpression of NRPS or PKS in *apo* form in *E.coli* since EntD is not expressed under standard laboratory conditions in LB medium and AcpS seemed not to be able to phosphopantetheinylate carrier proteins from secondary metabolism assembly lines in most cases. But while this fact is valid for *Streptomyces* assembly line, myxobacterial carrier proteins seem to be a valid substrate for AcpS leading to a portion of the protein being purified in *holo* form during overexpression in *E.coli*. This discrepancy between CP from *Streptomyces* and *Myxobacteria* is quite troubling since nothing in the sequence supports such a difference. However, this is further proof that our knowledge of assembly line type biosynthesis is still quite shallow even on domains seemingly as simple as the carrier proteins.

While biochemical investigations provided insight into the mode of action of individual NRPS domains as well as multimodular assembly lines, structural characterization of complex structures of various domains is still missing. To this date no *in cis* tailoring domain has been crystallized and the comprehension of the biochemistry of condensation domains would benefit greatly of complex structures of C domains harboring their substrate in acceptor and donor position. Furthermore, the crystallization of multimodular NRPS

assembly lines would provide great insights into the intermodular interactions which are to this date still understudied. However the multidomain nature of the assembly line architecture and the inherent flexibility they present proved to be a great challenge not only in terms of overexpression but also especially in terms of crystallization.

### 3.2 Limits of NRPS engineering and *de novo* NRPS biosynthesis

The ultimate purpose of biosynthesis studies as a research field lies in the promise that a sufficient knowledge of the biochemistry, structure and dynamics of biosynthesis enzymes would allow the reengineering of assembly line type biosynthesis systems and *in fine* the *de novo* biosynthesis of virtually any organic compound.(42)

It is noteworthy that most biosynthesis engineering studies use molecular biological approaches because they are often target driven and try to obtain pharmacologically relevant modifications of the final product. Thus, they tend to use a heterologous producer expressing the whole biosynthetic gene cluster and to modify specific genes either from the assembly line or related to tailoring. However, the multiplicity of enzymatic activities involved in a natural product biosynthesis makes it difficult to identify the weak link that led to the abolishment of production. Implying that in the case of a negative outcome it is impossible to identify whether the modification was unsuccessful or whether a downstream process was impacted by the substrate modification, *i.e.* the subsequent C domain blocked the new modified substrate. While biochemistry of most of the reactions involved in NRPS biosynthesis are known, with the notable exception of the condensation reactions, the operational mode of substrate recognition is still mostly unidentified. The lack of knowledge about the critical substrate-protein and protein-protein interactions taking place in the assembly line implies that biosynthetic processes in NRPS assembly lines still have to be investigated further *in vitro* in the frame of combinatorial biosynthesis and NRPS biosynthesis engineering studies.

Natural product biosynthesis engineering provides an alternative to synthetic chemistry for the production of potentially pharmaceutically active derivatives of natural compounds. Indeed even if the toolbox is not as developed as organic chemistry and therefore does not allow as much diversity in the derivatives production, the modification of biosynthesis on genomic level allows producing new natural products derivatives in a fast and elegant fashion.(43) Early works in the field are based on precursor directed biosynthesis since this technique requires only limited knowledge about the biosynthesis. It consists in the feeding of homologues of specific precursor molecules in the biosynthesis in the hope that they can be incorporated if the biosynthetic processes are not to substrate specific. This method was successfully applied in several cases (44–46) but this technique has a limitation since the biosynthesis of the original precursor is still active and leads to a competition between the natural substrate and the artificial substrate that was fed to the culture. To tackle this issue mutasynthesis was developed, this technique is an improvement of precursor directed biosynthesis in which additional genetic engineering is performed leading to the

abolishment of the biosynthesis of the natural precursor in order to increase the efficiency of the feeding of the modified precursor.(47,48) (Figure 7.4)

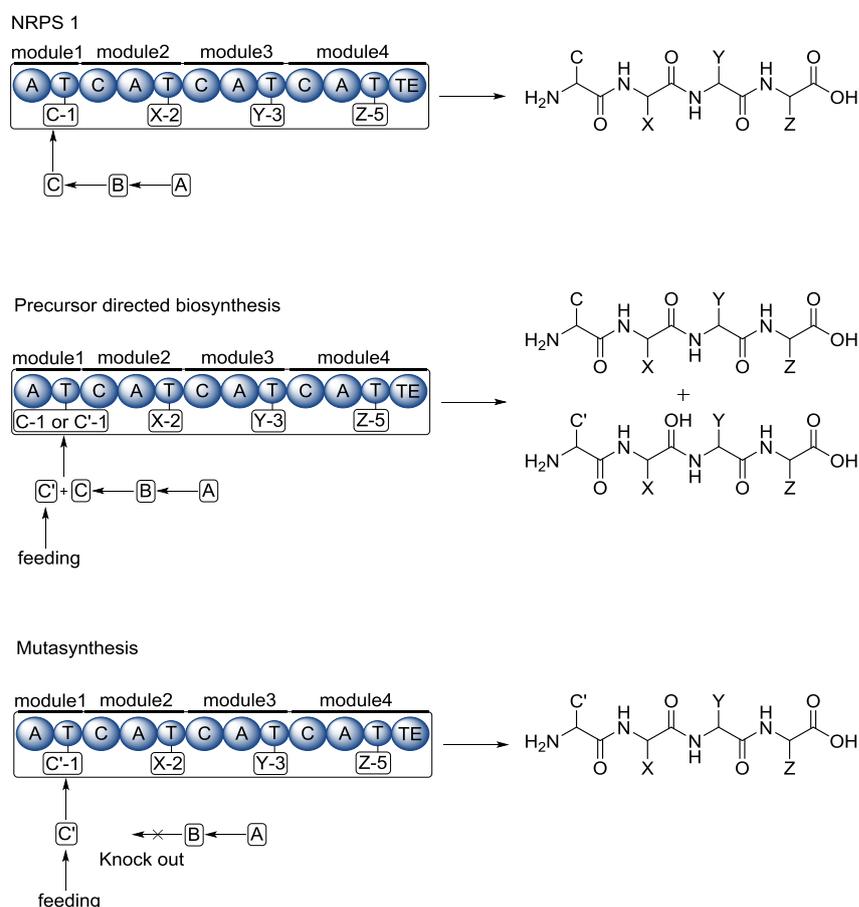


Figure 7.4: Schematic representation of the modification of a theoretical NRPS biosynthesis by precursor directed biosynthesis and mutasynthesis

Rational engineering can only be used once a system is known with a sufficient amount of details, in this case, the ultimate test of knowledge is to try and redesign it or even rebuild it. The biosynthesis of NRPS and PKS can be subdivided between the assembly line type biosynthesis of the backbone and the tailoring reaction.

An important aspect of biosynthesis engineering is the modification of the tailoring affecting a natural product in order to generate new decoration patterns on an unmodified backbone. This has triggered the search for promiscuous tailoring enzymes that could be used for derivatization of other natural products than their original substrate.(43) It is however notable that in the case of tailoring enzymes exerting their activity on the completed backbone after its release from the assembly line, only enzymes from the biosynthesis of very closely related compounds can be transferred from one system to another.(49) But tailoring enzymes acting *in trans* on a specific PCP loaded amino acid should be readily transferable from one assembly line to another using the same amino acid as building block. However, tailoring enzymes have to display a certain level of specificity for the NRPS module rather

than only for their substrate since assembly line featuring multiple occurrences of the same amino acid can display differential tailoring for them. It would thus be necessary to identify the binding surface that allows recognition between the tailoring enzyme and the NRPS module. An interesting case of tailoring engineering can be found with the structurally related compounds Ramoplanin and Enduracidin (50), in particular, their halogenases triggered a lot of interest since it is well known that halogenation of a compound has drastic effects on its solubility and pharmacokinetics in general. Ramoplanin and Enduracidin feature different chlorination pattern with a dichlorination on hydroxyphenylglycine 13 for Enduracidin and a single chlorination on hydroxyphenylglycine 17 for Ramoplanin. Upon knockout of the Enduracidin halogenase and complementation with the Ramoplanin homologue a monochlorinated Enduracidin derivative can be obtained. More importantly the chlorination occurs on hydroxyphenylglycine 13 rather than 17; this implies that while the activity of the tailoring enzyme is retained (mono chlorination versus dichlorination) the substrate specificity is carried mainly by the NRPS module rather than by the loaded amino acid itself. This partial decoupling of the catalytical activity from the substrate recognition is both an exciting opportunity in terms of biosynthesis engineering but it underlines the necessity of further study of assembly line-tailoring protein interactions.

In the case of NRPS assembly line engineering, the adenylation domain was originally supposed to be the sole gatekeeper since it is responsible for the selection of the amino acid to be integrated in the peptidic chain and its modification was initially thought to be sufficient for the re-engineering of the NRPS assembly line.(43) Furthermore, the importance of the Stachelhaus code amino acids lining the active site was identified quite early, latter the ten residue code was extended to thirty four within 8 Å of the binding pocket.(28) However even if the modification of these residues sometimes successfully led to the production of new derivatives it also led to decreased yields (51), it is likely that the modification of this active site “shell” prompts rearrangements within the hydrophobic core of the protein and will lead to a need for mutation of even more residues in order to obtain domain with satisfactory kinetic parameters. Other strategies focused on the replacement of the entire adenylation domain or of the larger core domain (52,53) but were not much more successful either, probably because interdomain and intermodular interactions rely on a few important surface residues that cannot be modified. Later on, the importance of the condensation domain as a gatekeeper was acknowledged (54) but contrary to the adenylation domain the residues responsible for substrate binding were never identified in condensation domains due to the already discussed complexity of structural studies of quaternary complexes involving two phosphopantetheine bound substrates.(25)

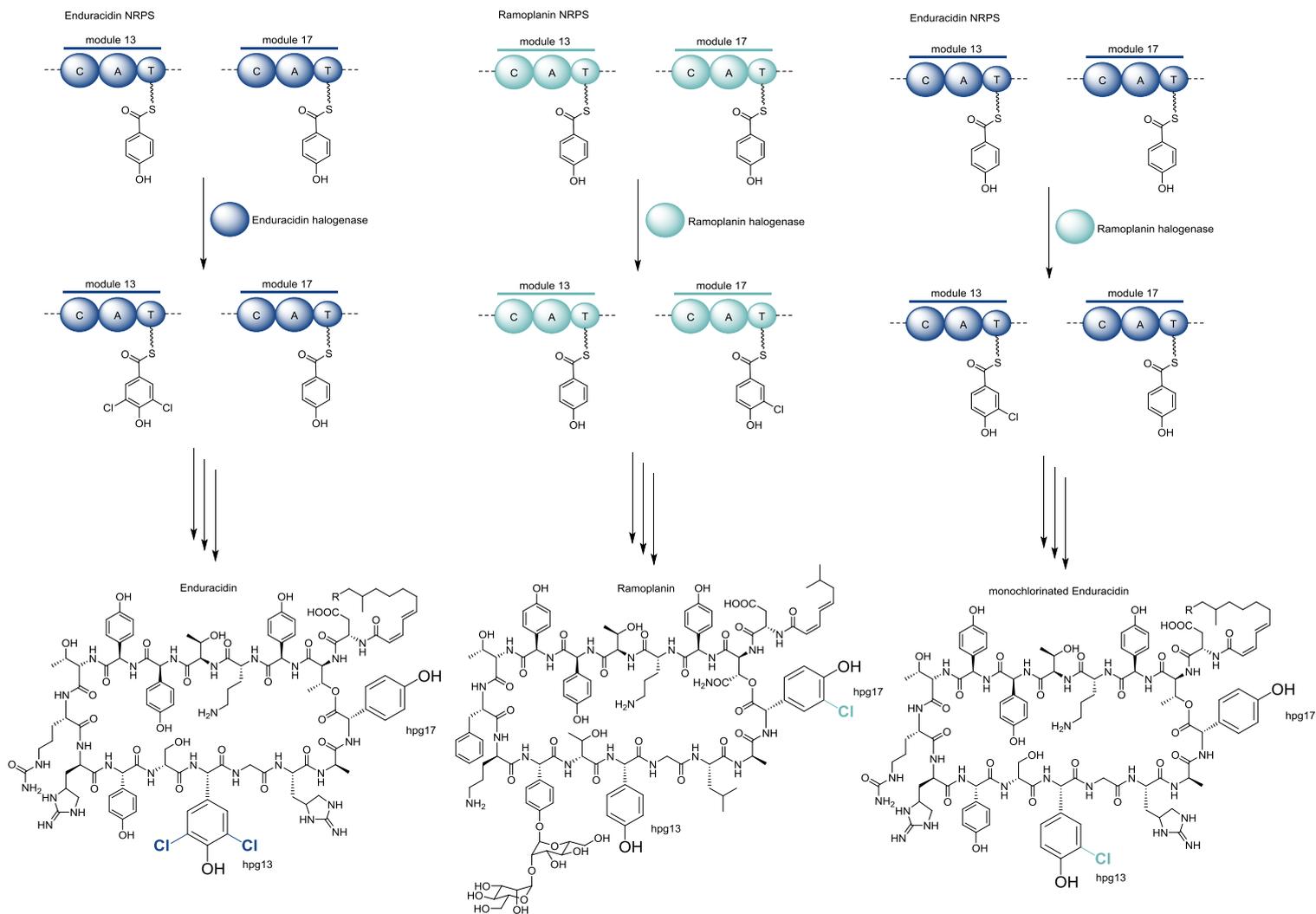


Figure 7.5: Modification of the halogenation pattern of Enduracidin by replacing the original hydroxyphenylglycine (hpg) halogenase by a homologue from the related Ramoplanin cluster

Modifications of single domains or full modules have occasionally been successful, for example in the case of Daptomycin where gene fragments could readily be swapped with others originating from the closely related clusters encoding CDA and A54145 biosyntheses (55,56,53) but this proof of concept has failed to be extended to other natural compound biosyntheses. In the case of Daptomycin the swapping of C-A-T modules was performed with the integration of module 11 from A54145 (specific for asparagine) in place of modules 8 and 11 from Daptomycin (respectively specific for alanine and serine) allowing the isolation of modified Daptomycin derivatives harboring D-asparagine in these positions. Additionally the complete DptD gene, coding for two last modules of the assembly line, could be replaced by CdaPS3 and LptD from CDA and A54145 yielding further modified Daptomycin derivatives. (Figure 7.6) However these clusters are very closely related and share an almost identical architecture which translates in a very similar protein structure allowing to readily swap those modules between biosyntheses. This approach has however been very difficult to replicate when using sequences from non-related clusters. This lack of reproducibility cast doubt on

---

the feasibility of *de novo* cluster engineering or even targeted modification of specific clusters having no related homologues.

Recently the use of A-T-C exchange unit (XU) rather than the traditional C-A-T modules was performed successfully.<sup>(57)</sup> This idea was originally envisioned taking into account the unusual length (32 amino acids) and structure of the C-A linker which would make a better cutting point in the assembly line rather than the relatively short T-C linker. The cutting point in the linker was identified after the  $\alpha$ -helix at the end of the 22 N-terminal amino acids which are C domain interaction partners and before the 10 C-terminal which are interacting with the A domain. It led to a change in the module paradigm which was originally based on the activation of the amino acid as a centerpiece and the tailoring it might receive prior to its condensation with the growing chain, toward a model based on the interaction between proteins and condensation activity between upstream and downstream partners as a focal point. Thus, the exchange unit is not only based on the amino acid it activates but also on the condensation specificity between this amino acid and its downstream partner. (Figure 7.7) This new paradigm led to successful engineering of clusters and even generation of new biosyntheses, with good overall yields for modified biosyntheses. The yields were however shown to be highly dependent on the conservation of the key residues making the C-A interface between XU pinpointing the critical importance of the mostly unknown processes of protein-protein interactions within the assembly line. While this approach showed for the first time a reliable *de novo* biosynthesis strategy, it is fairly limited regarding the monomers that can be used due to the dual specificity that has to be accommodated. Hence even with the huge amount of NRPS sequences available only a small portion of the possible amino acids can be used with such an approach. It is however noteworthy that the XU strategy was performed using clusters originating only from *Photorhabdus* and *Xenorhabdus* species and has yet to be replicated with sequences from actinobacteria or myxobacteria.

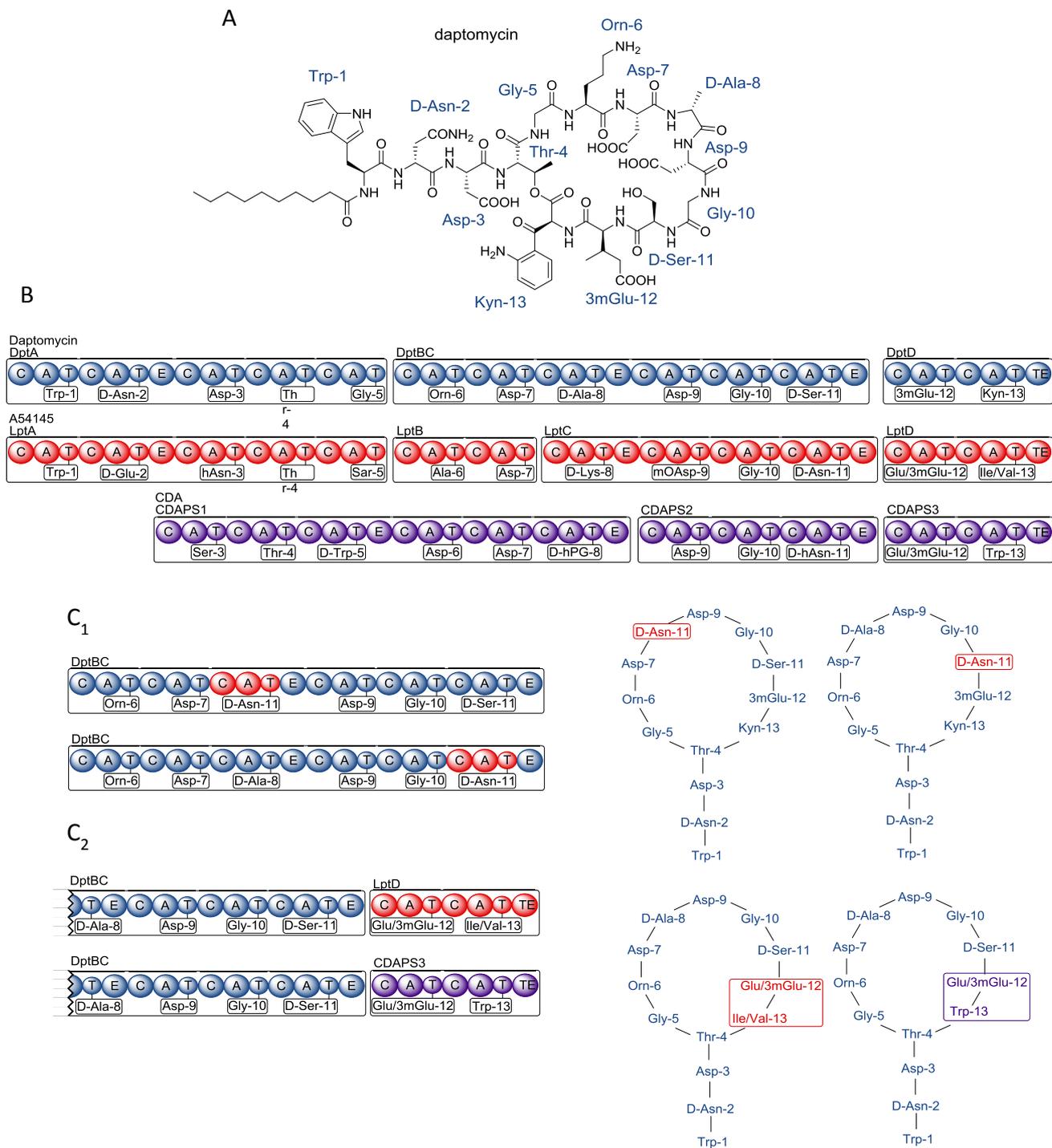


Figure 7.6: **A**: Structure of Daptomycin; **B**: Schematic representation of the related Daptomycin (blue), A54145 (red) and CDA (purple) assembly lines; **C**: Modifications of the Daptomycin gene cluster with integration of module 11 coding for D-Asn from the A54145 assembly line in position 8 and 11 of the Daptomycin assembly line (**C<sub>1</sub>**) and replacement of the DptD gene by its equivalent LptD (red) from the A54145 cluster and CDAPS3 (purple) from CDA cluster (**C<sub>2</sub>**)

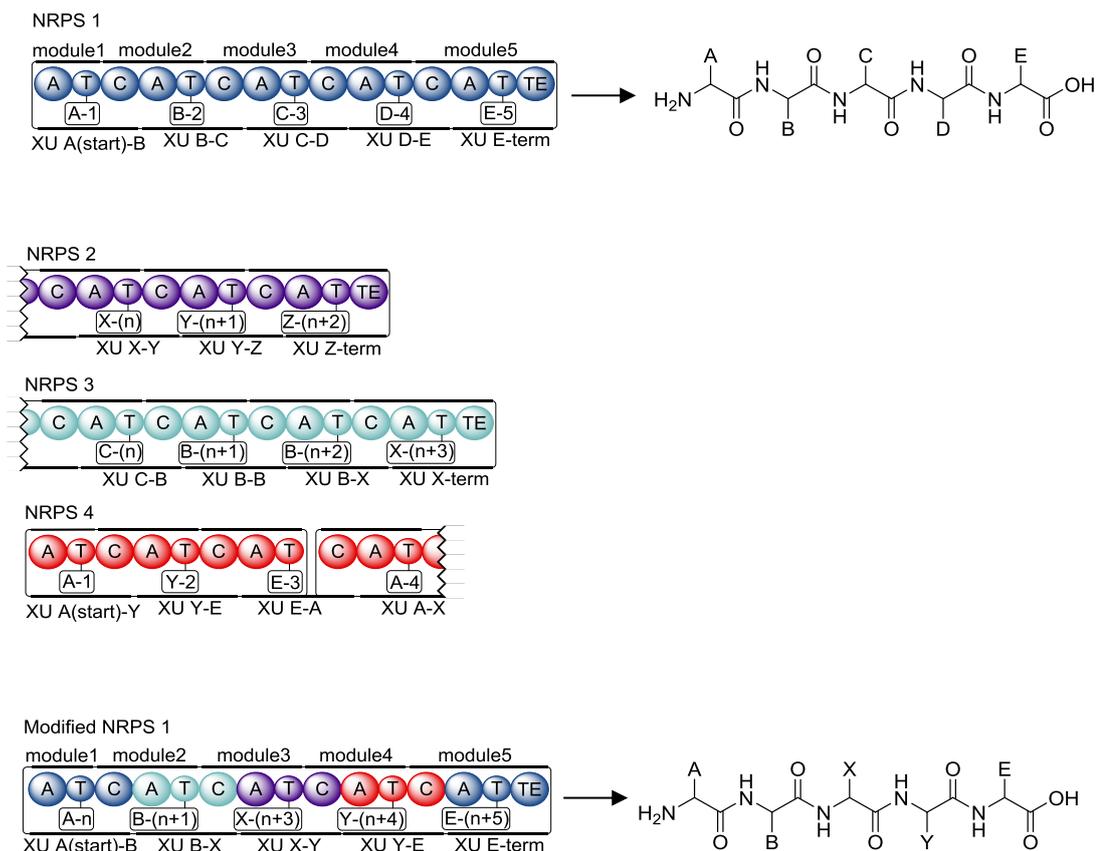


Figure 7.7: Engineering of a theoretical NRPS cluster based on the A-T-C exchange units (XU) strategy, the NRPS 1 cluster producing the A-B-C-D-E peptide can be modified to produce the A-B-X-Y-E peptide by integration of three XU belonging to various clusters from a database but the dual specificity of the XU has to be respected. To accommodate this, the integration of an X amino acid in position three implies both the exchange of the XU3 to replace the adenylation domain of module 3 and exchange of the XU2 to fit the specificity of the condensation domain. And the integration of a Y amino acid in position four implies the selection of an XU4 not only selective for Y but also for the subsequent E amino acid.

## 4 Conclusion and Outlook

The primary objective of this work has been to unravel the complex biosynthetic mechanisms that underly the production of the recently isolated myxobacterial compound family Cystobactamides. In the frame of biosynthesis studies, the focus has been put on the characterization of the unusual biochemical mechanisms involved in the biosynthesis such as the isomerization and shuttling of the isoasparagine moiety and the complex methylation pattern. Additionally, the presence a feedback inhibition mechanism has been demonstrated in the Cystobactamide biosynthesis, a process that, while widely accepted to exist in secondary metabolism, was almost never evidenced. While the Cystobactamides biosynthesis features a few fascinating examples of unusual processes it also provides a perfect framework for studying the more classic processes involved in biosynthesis because of its important use of non- $\alpha$  amino acids which provided insights in the adaptation an NRPS assembly line can show to accommodate unusual building blocks. Furthermore, the heterologous expression platform for NRPS modules that was developed was used to study the extensive *in trans*

tailoring at play in the Cystobactamide biosynthesis on the natural substrate rather than on a SNAC ester mimic allowing the investigation the protein-protein interactions involved in such reactions.

While the complex process resulting in isomerization or dehydration of asparagine performed by a newly discovered NRPS domain has been identified for the first time in this work, further investigations through structural and mutational analysis of the X domain could allow understanding its unique biochemistry. Furthermore, while the expression of the NRPS modules was achieved in an active form, determination of the conditions allowing the condensation domains to operate *in vitro* would provide insights in this still unelucidated mechanism and yield an interesting *in vitro* biosynthesis platform allowing a biochemical investigation of each step. Completion of this work further deepened the knowledge about biosynthesis systems by experimentally assigning and confirming the function of most of the tailoring enzymes in the cluster. Since the beginning of the post-genomic era, the improvement of genomic databases is critical to increase the precision of *in silico* function assignment of newly discovered biosynthesis genes. The fundamental biochemical research performed in this thesis also laid the foundations for further synthetic biology work allowing for a more target driven approach of biosynthesis engineering with respect of pharmacological aspects in order to obtain, through fermentation, Cystobactamide antibiotic derivative with improved antibacterial spectrum and pharmacokinetics having the potential to reach the market.

## 5 References

1. Herrmann, K. M. (1995) The shikimate pathway as an entry to aromatic secondary metabolism, *Plant physiology* 107, 7–12.
2. Gerhart, J. C., and Pardee, A. B. (1962) The enzymology of control by feedback inhibition, *The Journal of biological chemistry* 237, 891–896.
3. McFarlane, J., Madyastha, K. M., and Coscia, C. J. (1975) Regulation of secondary metabolism in higher plants. Effect of alkaloids on a cytochrome P-450 dependent monooxygenase, *Biochemical and Biophysical Research Communications* 66, 1263–1269.
4. Liscombe, D. K., Louie, G. V., and Noel, J. P. (2012) Architectures, mechanisms and molecular evolution of natural product methyltransferases, *Natural product reports* 29, 1238–1250.
5. Caboche, S., Pupin, M., Leclère, V., Fontaine, A., Jacques, P., and Kucherov, G. (2008) NORINE: a database of nonribosomal peptides, *Nucleic acids research* 36, 31.
6. Gil, J. A., and Campelo-Diez, A. B. (2003) Candicidin biosynthesis in *Streptomyces griseus*, *Applied microbiology and biotechnology* 60, 633–642.
7. Thongkongkaew, T., Ding, W., Bratovanov, E., Oueis, E., Garcı A-Altare, M. A., Zaburannyi, N., Harmrolfs, K., Zhang, Y., Scherlach, K., Müller, R., and Hertweck, C. (2018) Two Types of Threonine-Tagged Lipopeptides Synergize in Host Colonization by Pathogenic Burkholderia Species, *ACS chemical biology* 13, 1370–1379.
8. Gatignol, A., Durand, H., and Tiraby, G. (1988) Bleomycin resistance conferred by a drug-binding protein, *FEBS Letters* 230, 171–175.
9. Keating, T. A., and Walsh, C. T. (1999) Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis, *Current opinion in chemical biology* 3, 598–606.
10. Chen, H., Thomas, M. G., O'Connor, S. E., Hubbard, B. K., Burkart, M. D., and Walsh, C. T. (2001) Aminoacyl- S -Enzyme Intermediates in  $\beta$ -Hydroxylations and  $\alpha,\beta$ -Desaturations of Amino Acids in Peptide Antibiotics †, *Biochemistry* 40, 11651–11659.
11. Fridovich, I. (1986) Biological effects of the superoxide radical, *Archives of biochemistry and biophysics* 247, 1–11.
12. Cociancich, S., Pesic, A., Petras, D., Uhlmann, S., Kretz, J., Schubert, V., Vieweg, L., Duplan, S., Marguerettaz, M., Noëll, J., Pieretti, I., Hügelland, M., Kemper, S., Mainz, A., Rott, P., Royer, M., and Süßmuth, R. D. (2015) The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine, *Nature chemical biology* 11, 195–197.
13. Baumann, S., Herrmann, J., Raju, R., Steinmetz, H., Mohr, K. I., Hüttel, S., Harmrolfs, K., Stadler, M., and Müller, R. (2014) Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity, *Angewandte Chemie (International ed. in English)* 53, 14605–14609.
14. Hüttel, S., Testolin, G., Herrmann, J., Planke, T., Gille, F., Moreno, M., Stadler, M., Brönstrup, M., Kirschning, A., and Müller, R. (2017) Discovery and Total Synthesis of

Natural Cystobactamid Derivatives with Superior Activity against Gram-Negative Pathogens, *Angewandte Chemie (International ed. in English)* 56, 12760–12764.

15. Singh, G. M., Vaillancourt, F. H., Yin, J., and Walsh, C. T. (2007) Characterization of SyrC, an aminoacyltransferase shuttling threonyl and chlorothreonyl residues in the syringomycin biosynthetic assembly line, *Chemistry & biology* 14, 31–40.

16. Strieter, E. R., Vaillancourt, F. H., and Walsh, C. T. (2007) CmaE: a transferase shuttling aminoacyl groups between carrier protein domains in the coronamic acid biosynthetic pathway, *Biochemistry* 46, 7549–7557.

17. Walsh, C. T., Haynes, S. W., and Ames, B. D. (2012) Aminobenzoates as building blocks for natural product assembly lines, *Natural product reports* 29, 37–59.

18. Hertweck, C. (2009) The biosynthetic logic of polyketide diversity, *Angewandte Chemie (International ed. in English)* 48, 4688–4716.

19. Li, Y., Zhang, W., Zhang, H., Tian, W., Wu, L., Wang, S., Zheng, M., Zhang, J., Sun, C., Deng, Z., Sun, Y., Qu, X., and Zhou, J. (2018) Structural Basis of a Broadly Selective Acyltransferase from the Polyketide Synthase of Splenocin, *Angewandte Chemie (International ed. in English)* 57, 5823–5827.

20. Yonus, H., Neumann, P., Zimmermann, S., May, J. J., Marahiel, M. A., and Stubbs, M. T. (2008) Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains, *The Journal of biological chemistry* 283, 32484–32491.

21. Linne, U., Schäfer, A., Stubbs, M. T., and Marahiel, M. A. (2007) Aminoacyl-coenzyme A synthesis catalyzed by adenylation domains, *FEBS Letters* 581, 905–910.

22. Du, L., Sánchez, C., and Shen, B. (2001) Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules, *Metabolic engineering* 3, 78–95.

23. Piel, J., Hertweck, C., Shipley, P. R., Hunt, D. M., Newman, M. S., and Moore, B. S. (2000) Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate ‘*Streptomyces maritimus*’. Evidence for the derailment of an aromatic polyketide synthase, *Chemistry & biology* 7, 943–955.

24. Müller, M., He, J., and Hertweck, C. (2006) Dissection of the late steps in aureothin biosynthesis, *Chembiochem : a European journal of chemical biology* 7, 37–39.

25. Bloudoff, K., and Schmeing, T. M. (2017) Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity, *Biochimica et biophysica acta. Proteins and proteomics* 1865, 1587–1604.

26. Keating, T. A., Marshall, C. G., Walsh, C. T., and Keating, A. E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains, *Nature structural biology* 9, 522–526.

27. Marshall, C. G., Hillson, N. J., and Walsh, C. T. (2002) Catalytic Mapping of the Vibriobactin Biosynthetic Enzyme VibF †, *Biochemistry* 41, 244–250.

28. Rausch, C., Hoof, I., Weber, T., Wohlleben, W., and Huson, D. H. (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution, *BMC evolutionary biology* 7, 78.
29. Stachelhaus, T., Mootz, H. D., Bergendahl, V., and Marahiel, M. A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain, *The Journal of biological chemistry* 273, 22773–22781.
30. van Wezel, G. P., and McDowall, K. J. (2011) The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances, *Natural product reports* 28, 1311–1333.
31. Hannig, G., and Makrides, S. C. (1998) Strategies for optimizing heterologous protein expression in *Escherichia coli*, *Trends in biotechnology* 16, 54–60.
32. Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges, *Frontiers in microbiology* 5, 172.
33. Chen, R. (2012) Bacterial expression systems for recombinant protein production: *E. coli* and beyond, *biotechnology advances* 30, 1102–1107.
34. Anné, J., Maldonado, B., van Impe, J., van Mellaert, L., and Bernaerts, K. (2012) Recombinant protein production and streptomycetes, *Journal of biotechnology* 158, 159–167.
35. Tarry, M. J., Haque, A. S., Bui, K. H., and Schmeing, T. M. (2017) X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture, *Structure (London, England : 1993)* 25, 783.
36. Edwards, A. L., Matsui, T., Weiss, T. M., and Khosla, C. (2014) Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase, *Journal of molecular biology* 426, 2229–2245.
37. Smith, J. L., Skinotis, G., and Sherman, D. H. (2015) Architecture of the polyketide synthase module: surprises from electron cryo-microscopy, *Current opinion in structural biology* 31, 9–19.
38. Herbst, D. A., Boll, B., Zocher, G., Stehle, T., and Heide, L. (2013) Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes, *The Journal of biological chemistry* 288, 1991–2003.
39. Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C., and Gulick, A. M. (2016) Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture, *The Journal of biological chemistry* 291, 22559–22571.
40. Drake, E. J., Miller, B. R., Shi, C., Tarrasch, J. T., Sundlov, J. A., Allen, C. L., Skinotis, G., Aldrich, C. C., and Gulick, A. M. (2016) Structures of two distinct conformations of holo-non-ribosomal peptide synthetases, *Nature* 529, 235–238.
41. Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P., and Burkart, M. D. (2014) The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life, *Natural product reports* 31, 61–108.
42. Winn, M., Fyans, J. K., Zhuo, Y., and Micklefield, J. (2016) Recent advances in engineering nonribosomal peptide assembly lines, *Natural product reports* 33, 317–347.

43. Winn, M., Fyans, J. K., Zhuo, Y., and Micklefield, J. (2016) Recent advances in engineering nonribosomal peptide assembly lines, *Natural product reports* 33, 317–347.
44. Traber, R., Hofmann, H., and Kobel, H. (1989) Cyclosporins. New analogues by precursor directed biosynthesis, *J. Antibiot.* 42, 591–597.
45. Krause, M., Lindemann, A., Glinski, M., Hornbogen, T., Bonse, G., Jeschke, P., Thielking, G., Gau, W., Kleinkauf, H., and Zocher, R. (2001) Directed biosynthesis of new enniatins, *The Journal of antibiotics* 54, 797–804.
46. Xu, Y., Zhan, J., Wijeratne, E. M. K., Burns, A. M., Gunatilaka, A. A. L., and Molnár, I. (2007) Cytotoxic and Antihaptotactic beauvericin analogues from precursor-directed biosynthesis with the insect pathogen *Beauveria bassiana* ATCC 7159, *Journal of natural products* 70, 1467–1471.
47. Weissman, K. J. (2007) Mutasynthesis - uniting chemistry and genetics for drug discovery, *Trends in biotechnology* 25, 139–142.
48. Kirschning, A., Taft, F., and Knobloch, T. (2007) Total synthesis approaches to natural product derivatives based on the combination of chemical synthesis and metabolic engineering, *Organic & biomolecular chemistry* 5, 3245–3259.
49. Deb Roy, A., Grüschow, S., Cairns, N., and Goss, R. J. M. (2010) Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs, *Journal of the American Chemical Society* 132, 12243–12245.
50. Yin, X., Chen, Y., Zhang, L., Wang, Y., and Zabriskie, T. M. (2010) Enduracidin analogues with altered halogenation patterns produced by genetically engineered strains of *Streptomyces fungicidicus*, *Journal of natural products* 73, 583–589.
51. Thirlway, J., Lewis, R., Nunns, L., Al Nakeeb, M., Styles, M., Struck, A.-W., Smith, C. P., and Micklefield, J. (2012) Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity, *Angewandte Chemie (International ed. in English)* 51, 7181–7184.
52. Kries, H., Niquille, D. L., and Hilvert, D. (2015) A subdomain swap strategy for reengineering nonribosomal peptides, *Chemistry & biology* 22, 640–648.
53. Nguyen, K. T., Ritz, D., Gu, J.-Q., Alexander, D., Chu, M., Miao, V., Brian, P., and Baltz, R. H. (2006) Combinatorial biosynthesis of novel antibiotics related to daptomycin, *Proceedings of the National Academy of Sciences of the United States of America* 103, 17462–17467.
54. Belshaw, P. J. (1999) Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis, *Science* 284, 486–489.
55. Baltz, R. H., Brian, P., Miao, V., and Wrigley, S. K. (2006) Combinatorial biosynthesis of lipopeptide antibiotics in *Streptomyces roseosporus*, *Journal of industrial microbiology & biotechnology* 33, 66–74.
56. Miao, V., Coëffet-Le Gal, M.-F., Nguyen, K., Brian, P., Penn, J., Whiting, A., Steele, J., Kau, D., Martin, S., Ford, R., Gibson, T., Bouchard, M., Wrigley, S. K., and Baltz, R. H. (2006) Genetic engineering in *Streptomyces roseosporus* to produce hybrid lipopeptide antibiotics, *Chemistry & biology* 13, 269–276.

57. Bozhüyük, K. A. J., Fleischhacker, F., Linck, A., Wesche, F., Tietze, A., Niesert, C.-P., and Bode, H. B. (2018) De novo design and engineering of non-ribosomal peptide synthetases, *Nature chemistry* 10, 275–281.