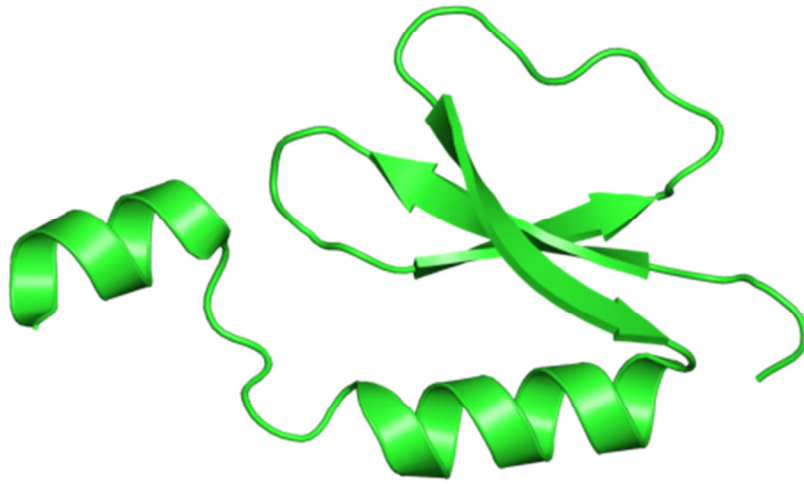# STRUCTURAL AND FUNCTIONAL ANALYSES OF THE MBTH-HOMOLOGS

Simone Pander

Supervisor: Reto Zwahlen

Department of Molecular Microbiology

19 November 2013

On the front-page the 3D structure of the MbtH-protein is shown, which is the PA4212 protein [6].

TABLE OF CONTENTS

ABSTRACT

Primary and secondary metabolites are produced by microorganisms to survive in their environment, where primary metabolites are essential and mainly used for growth, the secondary metabolites are non-essential and used for a broader range of purposes. The final products can reach from antibiotics, cytostatics, fatty acids to several other compounds [1]. A small group of secondary metabolites are produced by non-ribosomal peptide synthases (NRPSs), which produce non-ribosomal proteins (NRPs). They contain several domains of which the adenylation domain, defines the specificity of the module. Adenylation domains need sometimes the assistance of a MbtH-like domain/ protein to function properly, which gives the MbtH-like protein its function as a helper protein. MbtH-like proteins are mainly produced by *Actinobacteria*, but they are also found in several other bacteria, generally prokaryotes [4]. The predicted structure of the MbtH-like proteins reveals the importance of two conserved tryptophan residues, which form a small pocket in the MbtH-like domain [9]. The precise function of the MbtH-like proteins is still unknown, though several depletion and reconstitution assays show, that the MbtH-like proteins help with the activation, solubilisation and overexpression of NRPSs. Also MbtH-like proteins from several species can replace each other, which means that the protein is likely a chaperone, facilitator or stabilization factor. But some NRPSs have also activity without MbtH-like proteins or can be replaced by the RubC1a domain, which is a fusion of the A-domain and a MbtH-like protein and is seen as the ancestral version of the MbtH-like protein. Because of these possibilities it is likely that MbtH-like proteins have a specific function, which could by exploited for biological applications, like strain-improvement and during the optimization of antibiotic production by secondary metabolite biosynthesis.

INTRODUCTION

Microorganisms produce primary and secondary metabolites. Where primary metabolites are essential for growth and reproduction, the secondary metabolites are produced for different purposes. The final products consist of antibiotics, cytostatics, fatty acids, immunosuppressants and many more. The secondary metabolites are produced during the stationary phase of bacterial growth and used for example as protection for the cell, but for some secondary metabolites the function is still unknown. Different growth conditions can form different secondary metabolites, due to for instance, the medium composition. Secondary metabolites can be classified due to their producing enzymes, or according to their biological application. If they are classified by their producing enzyme the following classes can be formed: polyketides, terpenoids and steroids, phenylpropanoid compounds, alkaloids, carbohydrates and amino acids or peptides [1]. The production of metabolites is catalysed by for example multimodular enzymes. These enzymes consist of several domains, which together form large (circular) compounds. Three classes of multimodular enzymes can be distinguished: Polyketide synthase (PKS), Fatty acid synthase (FAS) and non-ribosomal peptide synthase (NRPS). FAS, PKS and NRPS enzymes produce their characteristic low molecular weight metabolites by the loading and condensation of smaller monomeric units. The following properties distinguish these enzymes: FAS produce mainly primary metabolites, where PKS produce primary and secondary metabolites and NRPSs only produce secondary metabolites. In FAS and PKS systems the small monomer units are fatty acids and ketides. NRPS enzymes introduce functional group diversity in their products by using all natural amino acids as well as unnatural amino and aryl aid substrates and additional diversity is achieved by a variety of domains: epimerization, halogenation, hydroxylation or tailoring enzymes, which alter the products in their L/D-conformation. This can be performed by for instance tailoring enzymes, which alter the compound after its release. This results in a different biosynthesis of the products. The FAS and PKS use the acyl monomer units by the action of acyltransferase (AT) domains, which uses serine for a nucleophilic attack of the CoA monomer. This intermediate is then transferred to an acyl carrier protein (ACP) domain. The growing acyl chain is condensed by the ketosynthase (KS) domain and the product is reduced by ketoreductase (KR), dehydratase (DH) and enoylreductase (ER). This results in a saturated hydrocarbon using FAS and unsaturated hydrocarbon by PKS, due to inactive reductase domains. [2]

NRPSs are a class of enzymes which can produce natural non-ribosomal secondary metabolites or non-ribosomal peptides (NRPs). NRPs are closely related to polyketide and fatty acid synthase products. Also their productive enzymes are closely related. During assembly of the NRP, each building block is activated and then incorporated into the growing molecule. This is catalysed by a set of different domains, starting with the adenylation domain (A). The A-domain recognizes, activates and transfers the substrate to a peptidyl carrier protein (PCP). The condensation (C) domain catalyses the peptide bond formation between two PCP-linked substrates. The epimerisation domain (E) is required for the change of an L-amino acid to the D-amino acid conformation. The thioesterase (TE) domain finally releases the substrate from the enzyme and hereby enables another turnover. [3] See figure 1. All NRPSs are arranged in gene clusters. It turns out that NRPS can have a helper protein in their gene cluster, where PKS en FAS don't need a helper protein.

***Figure* 1** An overview of three NRPS domains in the ACVS. The same domain arrangement can be found in other NRPSs. . .

The gene cluster of NRPSs generally contains the modules in order of the monomer composition of the final product [3] and the gene cluster contains a small unknown protein, which was first discovered in *Mycobacterium Tuburculosis* and named MbtH protein, which is the helper protein [4]. It turns out that a few other proteins were found, which have similarities in their genome sequence and all these proteins are called MbtH-like proteins/MbtH homologs. This thesis contains many aspects of MbtH-like proteins. In the first section it is described which organisms produce the MbtH-like proteins and what kind of conserved motifs they contain. After this, the structural analysis of the MbtH homolog will be discussed. The analysis was done with a MbtH homolog, using the integral MbtH domain attached to the adenylating enzyme and some analysis without the NRPS or adenylating enzyme. In the third section it is discussed what the possible functions of the MbtH-like proteins could be, based on depletion and reconstitution assays. The last section will contain an overview of the biological applications, if it is known what kind of role the MbtH-like proteins play during the synthesis of secondary metabolites.

PREVALENCE OF MBTH HOMOLOGS

Taxonomic group	Sequenced genomes	DptG homologs	DptG homologs/ genome
Actinobacteria	168	188	1.1
Streptomyces sp.	25	65	2.6
Mycobacterium sp.	17	35	2.1
All others	126	88	0.7
Bacteroidites/Chlorobi	128	1	<0.01
Chlamydiae	9	0	<0.2
Cyanobacteria	49	7	0.14
Firmacutes	346	31	0.09
Bacillales	73	31	0.4
Clostridia	138	0	<0.01
Lactobacillales	72	0	<0.01
Molecutes	29	0	<0.03
Others	34	0	<0.03
Proteobacteria	627	188	0.3
Alpha	180	16	0.09
Beta	99	56	0.6
Gamma	247	101	0.4
Delta	45	6	0.1
Epsilon	34	0	<0.03
Spirochaetales	22	0	<0.05

Table 1: The organisms that contain at least one MbtH homolog.

At this moment the amount MbtH-like proteins found in organisms is growing rapidly. So far over 500 MbtH-like proteins were discovered [4] and taxonomy trees formed. Almost all of the MbtH like proteins exist in bacteria. A BLAST search was carried out by R. Baltz to determine the taxonomy of the MbtH-like proteins. The protein DptG (MbtH like protein, associated with the daptomycin gene cluster) was used for the search and most proteins were found in bacteria. Table 1 shows the distribution of the MbtH homologs in the subgroups of bacteria. It shows that most of the MbtH homologs are found in *Actinobacteria*, *Firmicutes* and *Proteobacteria* 188 DptG homologs are encoded in 168 sequenced *Actinobacteria*; 65 of them are encoded in 25 species of *Streptomyces*. All 31 DptG homologs in the *Firmicutes* were found in different *Bacillus* species. This means that most of the MbtH-homologs

were found in free-living soil microbes. These species contain a relative large genome and are capable of producing secondary metabolites. The products contain 5 to 13 amino acids and include D- and L-conformations, along with other modifications. In most cases they are cyclized, glycosylated, lipidated or a combination of these three. The MbtH-like proteins consist of 62-80 amino acids and their genes are located in the gene cluster of secondary metabolites, along with other required enzymes for the formation of secondary metabolites. One MbtH homolog can support multiple NRPS gene clusters. There is one exception, the teicoplanin gene cluster, which contains two MbtH homologs. All but one *Streptomyces* genome contains at least one MbtH homolog and the average is three. The genome size of *Streptomyces* range from 6.6 and 11.0 mega bases, but there is no correlation between the number of MbtH-like proteins and the genome size. The highest number of MbtH homologs is observed in *S. clavuligerus* and *S. roseosporus,* which have genome sizes between 7.8 and 8.6 mega bases. The MbtH homologs are located in the gene cluster, so they are up- or downstream of the NRPS. In a few cases the MbtH-like protein is fused to the adenylation domain, which is separated by a flexible amino acid linker. An alignment of the found MbtH-like proteins reveals a signature sequence for the proteins. It can be presented as: NXEXQXSXWP-X5-PXGW-X13-L-X7-WTDXRP. These amino acids are potentially functional in the NRPS pathway. Especially the three tryptophan residues (W25, W35 and W55) are conserved in all known MbtH-like proteins. The MbtH homologs that are fused to the NRPS are missing the W55, which indicates that the W55 may lie beyond the point of fusion, or it might imply that the W55 isn't required for fusion proteins, which means that it stabilizes the adenylating enzyme in another way. Or different sets of different residues take over the function of W55. [4]. Even though the conserved regions are known, it is still unknown how they interact with the NRPS modules and alter their functions.

STRUCTURAL ANALYSIS AND INTERACTIONS OF THE MBTH HOMOLOG

Crystal structures of the MbtH domain attached to the NRPS domain is difficult to retrieve because of the flexibility of the NRPS domains. After the initial adenylation step a conformational change occurs in which the C-terminal part rotates by ~140° in order to facilitate a second conformation for thioester bond formation. The flexible part of the adenylating enzymes is the C-terminal part of the NRPS, called Asub and it consists of 120 amino acids. The remaining N-terminal part is called the Acore domain and is a rigid structure, which lies next to the active site. The active site of the A domain lies in between the A sub and A core domain and which doesn't undergo the conformational change. Since the conformational change gives two confirmations (open and closed) the crystal structure of the open form is difficult to achieve. The same is seen in NRPS enzymes which structures are also only known in the closed confirmation, which leads to the poor understanding of how the conformational change works. So crystal structures of MbtH-like proteins attached to the NRPS domains are difficult to achieve as well. Though three structures of different MbtH-like proteins are determined, two structures of the MbtH-protein only and the first crystal structure of an MbtH-like domain attached to the adenylating enzyme, which was discovered in 2013 and is shown last. The first crystal structure of the MbtH protein is of PA4212, which is an independent MbtH-like protein from the Pyoverdine cluster of *P. aeruginosa*. The second structure was obtained using nuclear magnetic resonance (NMR) and Rv2377c is the first identified member of the MbtH-like proteins, derived from Mycobacterium tuberculosis. The final structure is obtained from SlgN1, a 3-methylaspartate adenylating enzyme, which contains an MbtH-domain at its N-terminal side. This represents the first structure of an MbtH-like domain attached to the adenylating enzyme.

The overall structure of MbtH-like proteins

The PA2412 protein dimensions are ~25 x 43 x 13 Å and it contains a central three-stranded antiparallel β-sheet, followed by two α-helixes. The first α-helix is packed to the sheet, where the second helix is turning to the C-terminus [6]. But the second helix wasn't seen in the other two publications (SlgN1 and Rv2377c) of the structures of the MbtH-like protein. The β-sheet bundle is organised as β2(S23-P26)- β1(F13-N18)- β3(W35-H39)[7] and also contains the conserved tryptophan residues. The conserved tryptophan residues are thought to play an important role in the function of the MbtH-like proteins. The first tryptophan residue (W25) can be found in the second barrel of the antiparallel β-sheet, where it lies in the conserved region: SXWP and the second tryptophan (W35) can be found structurally near the first residue in the second conserved region, embedded in the PXGW motif. Both tryptophan residues lie on the surface of the MbtH-protein, which allows easy interactions with residues of the opposite adenylation domain.

Figure 2: Overall structure of an MbtH-like protein as it is observed with PA2412. It shows the three stranded β-sheet and the two α-helixes at the C-terminal end.

The second tryptophan (W35) is located in the loop between β2 and β3 where a small pocket is formed by the indole side chains bordered on one side by P32 and on the other side by S23. The base of the pocket is formed by either V15 or L15. The third tryptophan is surrounded by the following motif: WTDXRP, which is also conserved in many MbtH-like proteins. Where the first tryptophan residue is important for the interaction with the A domain, the third tryptophan residue is important for the internal structure and the arrangement of the protein. The tryptophan (W55) is followed by an asparagine (D57) residue, which forms an ionic interaction with the following arginine (R59). The last β-sheet contains a conspicuous twist and one β-bulge, which results in a lower free energy than the first two β-strands (V38) [8]. After the β-sheet, a α-helix is found, which is stabilised by the hydrophobic core of the side chains in the β-sheet (V15, L25, V39, Y51 and V52). A second α-helix is seen in the structure of PA2412, which is initiated with a proline residue (P60), though the same region has been determined to be disordered and flexible within the structure of Rv2377c and SlgN1. Figure 2 shows clearly the divergence between the PA2412 (2PST) and Rv2377c

Figure 2: Comparison of the predicted structure of PA2412 (2PST) and Rv2377c (2GPF and 2KHR) [6].

(2GPF and 2KHR) at the C-terminal end. The structure of the MbtH-like proteins is temperature dependent, but heating experiments of Rv2377c showed that the structure is randomized at high
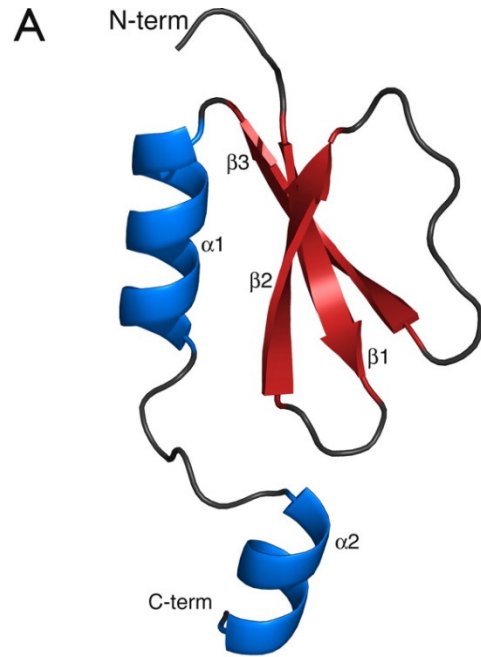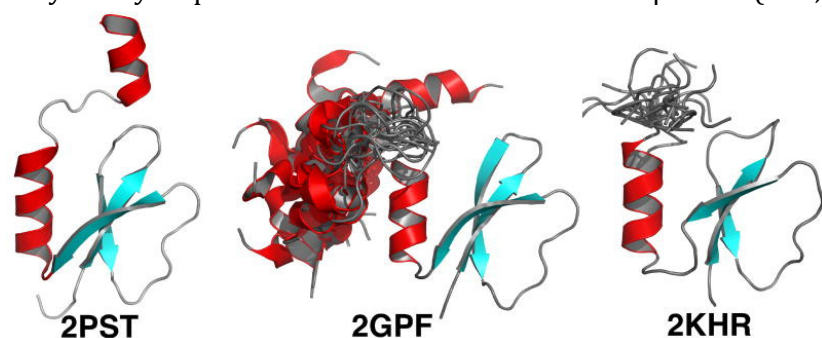
temperature, but re-folded after cooling down, which means that all the interactions can be reformed [7].

Adenylating enzyme attached to an MbtH-domain.

Since a few structures of MbtH-like proteins are available, some structural importance can be determined. Only one structure shines light on the interactions of the MbtH-like protein to the adenylating domain. The SlgN1 is involved in the biosynthesis of the antibiotic Streptolydigin and the first MbtH-domain, which is crystalized together with the adenylating domain of an adenylating enzyme. NRPSs contain adenylating domains as mentioned earlier, this domains consist of an A-core and an A-sub domain. How the conformational changes occur is unknown but it is thought to be that the A-sub domain is a smaller flexible and rotational domain, where the A-core domain is the domain which doesn't move. The protein used for crystallisation consists of the MbtH-like domain attached to the A-core domain and is called the SlgN1ΔAsubc protein [9]. The

Figure 2. Overall structure of SlgN1 in which the MbtH-like domain is coloured dark grey. The C-terminus is attached to the SlgN1 by a few peptide bonds. The MbtH-like domain interacts with α-helix 11.

SlgN1ΔAsubc protein belongs to the Acyl-CoA, NRPS A domain, Luciferase family (ANL) [10] and it contains 5 β-sheets, 11 α-helixes, two 3_{10} helixes and three disordered regions. The MbtH-like domain is connected to the A-domain by a flexible linker, which consists of 10 amino acids (see figure 2). The MbtH-like domain has the same typical properties as the proteins described above. Also this structure doesn't show the second α-helix, like PA2412. The MbtH-like protein is attached to the A-core domain and an interface is created. A-Helix 11 and the β-strands β19 and β14 of the A-core domain are involved in the interface. The main interactions of the MbtH-like domain are found in the β2, the following loop towards β3 and the α-helix. The interactions exist of Hydrogen bonds, salt bridges and nonpolar interactions, which are facilitated by thirty interactions of the MbtH-like domain and 36 of the A-core domain. This means that more than 40% of the MbtH-like domain is used for the interactions with the A-core domain, which yields in a high energy gain upon complex formation. The interface is stabilized by 5 hydrogen bonds and two salt-bridges. The main interactions of the MbtH-like domain are facilitated by the conserved tryptophan residues. The first two conserved tryptophan residues form a small pocket with the indole rings. This pocket is a negative imprint of Ala-433 of the A-core domain and the MbtH-like domain tightly fits in the flat surface ring of W35, S23, W25 and P32 (see attachment 1), which surrounds the Ala-433. The MbtH-like protein doesn't make any interactions with the active centre of the A-core domain or its residues, though indirectly Tyr-420 stabilizes Glu-360 of the A-core domain, which participates in the coordination of the catalytic magnesium. Tyr-420 is also part of the interface between the MbtH-like and A-core domain.
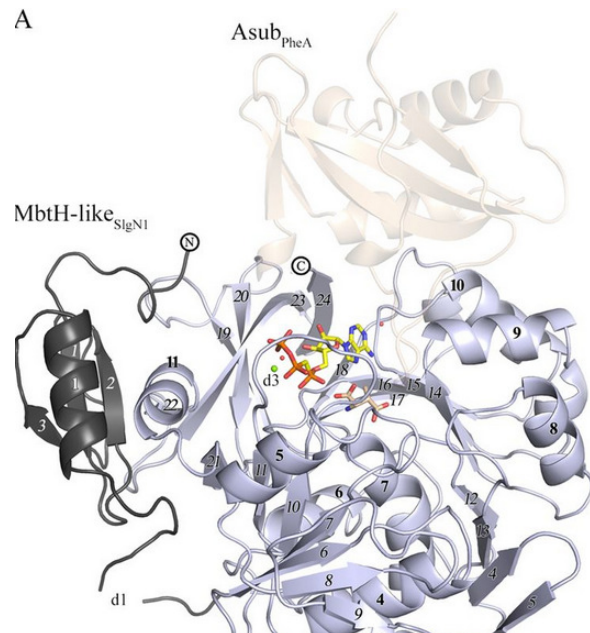
DEPLETION AND RECONSTITUTION ASSAY REVEAL CERTAIN FUNCTIONS OF THE MbtH-LIKE PROTEIN

NRPS proteins are soluble co-expressed with MbtH-like proteins

Mycobacterium tuberculosis is a pathogenic bacteria and a more efficient treatment for tuberculosis is preferred. The NRPS components MbtB, MbtE and MbtF are determined as virulence factors, which could be a useful drug target. During the purification of the NRPS components a small ~72 amino acid protein was co-purified, because of its co-expression and high affinity to NRPS components. The protein is called MbtH, encoded by a gene in the mbt-1 biosynthetic gene cluster. [12] The MbtH protein influences the solubility and activity of some NRPSs. MbtB is an NRPS domain and its solubility is increased two fold in the presence of MbtH. The same is observed for MbtE and MbtF which are soluble with MbtH co-expression. If MbtH isn't co-expressed then MbtE and MbtF aren't soluble. The mechanism used for the increase of solubility is still unknown, but a theory exists in which the hydrophobic side of the MbtH-like protein interacts with the hydrophobic side of the A-domain. This leaves a bigger complex, which automatically exhibits a higher solubility due to the coverage of the most hydrophobic parts. The theory is also supported by the data of the SlgN1 [11] [9].

YbdZ is an MbtH-like protein in *E. coli*, which plays a role in the ENT siderophore biosynthetic pathway. YbdZ co-expresses and co-purifies with EntF, 0.6 mole of YbdZ is used for each mole of EntF. Approximately the same amount is seen in the biosynthesis of the anti-tuberculosis drugs Capreomycin (CMN) and Viomycin (VIO). CmnO and VioO are both A-domains which require MbtH-like proteins. CmnN (MbtH-like protein) co-purifies with 0.42 mole for each mole of CmnO and the same results are shown for VioN (MbtH-like protein), which co-purifies with VioO [13]. The same results are shown in another study where the first NRPS in the assembly line of Glidobactins (GlbF) is co-expressed with the MbtH-like protein (GlbE) in *E. coli*. GlbE is located upstream of the GlbF-gene and the co-expression of the MbtH-like and adenylation domain resulted in a soluble and overexpressed complex, which consisted of 1.71:1 GlbE : GlbF. GlbE purification and overexpression isn't successful without the presence of GlbF[14]. Baltz already concluded that the stoichiometric ratio between the A domain and the MbtH-like protein is considered to be 1:2 in all cases [4].

Activity of NRPSs influenced by the addition of MbtH-like proteins

As mentioned before, YbdZ is a MbtH-like protein in *E. coli*, which results in a higher turnover rate for the substrate if it is added to the NRPS EntF. Affinity measurements show a 15-fold higher affinity for the substrate with the addition of YbdZ [13]. The same is seen in the Vicibactin biosynthesis, which also contains an NRPS together with a MbtH-like protein. The NRPS VbsS is

co-expressed together with the MbtH-like protein VbsG. Activity of the purified VbsS was measured with a pyrophosphate exchange assay and didn't show any activity. When both proteins are purified and used for determination of the activity it shows a high activity [15]. The same is seen in the biosynthetic gene cluster of the Pacidamycin gene, which produces the NRPS PacL. It is critical that the MbtH-homolog PacJ is added to overexpress and purify the NRPS [16].

MbtH-like proteins can interact with multiple NRPS gene clusters

It has already been proven that MbtH-like proteins can influence the activity and expression of their native NRPSs [11,13]. But the question remains if MbtH-like proteins can influence multiple genes or if one NRPS can be activated by multiple MbtH-like proteins. The only *Streptomyces* species that contains two MbtH-like genes in their gene cluster is *Streptomyces coelicolor* [4]. The first one is CchK, which is encoded by a gene in the gene cluster of Coelichelin, a peptide siderophore assembled by an NRPS. The second MbtH-like protein is CdaX, which is located in the gene cluster of an NRPS mediated assembly of the Calcium-dependent antibiotic (CDA). It turns out that either one of these MbtH-like proteins can activate both NRPSs without the presence of the other one. The other important outcome is that if both MbtH-like proteins are depleted, both NRPSs are partially inactivated, but if one of the two MbtH-like protein is added to the double knock-out strain, the activity of the NRPS protein is dramatically increased. It is important to exclude that the CchK and CdaX proteins aren't involved on a transcriptional level of the NRPSs. It turned out that both NRPS genes were transcribed in the double MbtH-like genes knock-out strains. This indicates that MbtH-like proteins have no influence on the transcription or translation of the genes [17]. This experiment shows that MbtH-like proteins can complement each other if they are derived from the same strain. But the question remains if the experiment can be performed with other MbtH-like proteins and is it also possible that MbtH-like proteins complement each other if they are derived from different species? Studies of the PacL NRPS show that it is necessary for the overexpression to add an MbtH-like protein. To evaluate this, the PacL strain undergoes NRPS expression and purification with multiple MbtH-like proteins: PacJ, VbsG, GlbE, Ybdz and Kutzneride (KtzJ). The last MtbH-like protein is found in the *Actinomycete*-derived cyclic depsipeptide biosynthetic gene cluster [18]. PacJ is the native MbtH-like protein of PacL and PacL is completely dependent on PacJ. But the unnatural MbtH-like proteins for PacL show that KtzJ and GlbE could actually stimulate PacL adenylation activity. While the other unnatural MbtH-like proteins (VbsG and Ybdz) are less effective. In the genome sequence, no explanation can be found for this, because all the unnatural MbtH-like proteins share an identity of around 30% to PacL, but differences within the 30% could probably explain the difference in efficiency. Even though, not all the MbtH-like proteins could activate PacL, it corroborates the idea that MbtH-like proteins can complement each other and that they can cross-talk between different NRPS clusters [19]. The same is observed with the NRPS gene cluster of Clorobiocin, which also contains a MbtH-like protein (CloY). It is shown that the inactivation of CloY results in a complete abolition of antibiotic production [20], but the result is not repeatable, which might indicate, that another MbtH-like protein replaces the functional activity of CloY. CchK and CdaX have 68% and 62% similarity to CloY and all three proteins are found in *Streptomyces coelicolor.* A triple knock-out was made, which resulted in a reduction of up to 99% of the Clorobiocin production. If CloY, CchK and CdaX were added to the triple knock-out strain, the Clorobiocin production increased 10-fold. Another protein could possibly fully restore the Clorobiocin production.. CouY has 81% similarity to CloY and is derived from the Aminocoumarin gene cluster. After introducing the CouY into the triple knock-out strain the Clorobiocin production is fully restored, which means that CouY can replace CloY in the Clorobiocin biosynthesis [21]. The Tyrosine-adenylating enzymes CloY, NovH, SimH and Pcza361.1 are involved in the biosynthesis of Aminocoumarins. The last three proteins are involved in the biosynthesis of Simoncyclinon D8 and Vancomicyn, which both require an MbtH-like protein for the formation of a heterotetrameric complex. All these NRPSs were cloned into *E. coli* using a strain with and without its MbtH-like protein YbdZ. All NRPSs show some activity in

the *E. coli* strain with *ybdZ* and didn't have any activity in the *E. coli* Δ*ybdZ*, which indicates that the MbtH-like protein YbdZ is actually activating all tested NRPSs. The same is seen during an assay in which all NRPSs were analysed with the four MbtH-like proteins (CloY, CdaX, SimY and Orf1van). It turned out that all MbtH-like domains could activate the NRPSs. Only NovH did also show some activity without any MbtH-like protein, which could be an MbtH-independent NRPS, but some NRPSs show still some activity without their MbtH homolog. So the MbtH independency of the protein is doubtful [22].

MbtH-independent activity of NRPS.

An unknown gene (orf1) with unknown function is found in the biosynthetic gene cluster of Balhimycin in *Amycolatopsis balhimycina*. This gene was identified as an MbtH-homolog, because it is around the same size and has a sequence identity of 68% to the MbtH protein and has a 100% identity with the conserved motif, which is mentioned before. The *orf1* is found downstream of the NRPS module 7. The knock out gene shows the same activity as a the wild type strain. This indicates that the NRPS is independent of Orf1 [23] and proves that independent MbtH activity of NRPS exist, which is also seen in all eukaryotic organisms that also uses NRPSs to produce secondary metabolites. An example is the delta-(L-alpha-amino-Adipyl)-L-Cysteinyl-D-Valine synthetase (ACVS) in *P. Chrysogenum* [28]. The same independence is also seen with NovH, which can also produce nonribosomal peptides without the addition of an MbtH-like protein. A comparison between CloH and NovH reveals that all non-conserved amino acids, except one, are located outside the active site. The only one which is different, is located at the active site of the NRPS, at position 383. NovH contains at that position a Methionine where CloH contains a Leucine. The introduction of the Methionine in CloH at position 383 leads to activity of CloH without any MbtH-homologs. This leads to the conclusion, that the mutation at position 383 induces MbtH-independence. The activity of CloH-L383M is 33% of the total activity of NovH in the absence of a MbtH-like protein and the activity is 36% of the CloH together with CloY. The NRPS isn't fully independent of MbtH-like proteins, but it still reveals the possibility of MbtH independent activity [22].

A domain can functionally replace MbtH-like domains

In the Rubradirin biosynthesis a Tyrosine adenylating enzyme was found, which catalyses two different key-reactions in the synthesis of the aminocoumarin Rubradirin. In figure 4, the composition of RubC1is shown, which contains the aminocoumarin acyl ligase, the Tyrosine-adenylase and the peptidyl-carrier domains [24]. The

Figure 3. Domain organization of the RubC1 and its truncated constructs RubC1a and RubC1b.[25]

adenylation domain is located in the C-terminal region of the RubC1 and is called RubC1b, but the interesting part of the gene is called RubC1a and contains an amide synthetase, which catalysis the linkage between the aminocoumarin and the acyl moiety, followed by an unknown region of 180 amino acids. The adenylation region of the protein shows high similarity to the NRPSs CloH, SimH and Pcza361.18, which were discussed earlier [25]. Even though, these NRPSs are all dependent on a MbtH-like protein, which isn't present in the Rubradirin gene cluster, Tyrosine-adenylation activity of Rub C1 is observed [24]. Since overexpression was performed in *E. coli*, the presence of the natural MbtH-like protein YbdZ may explain the observed activity.
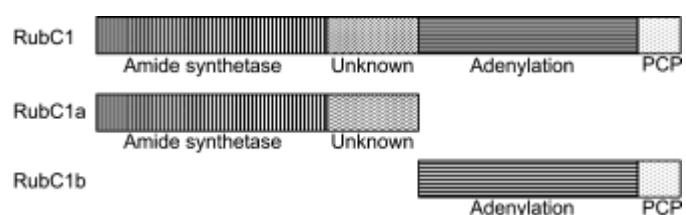
RubC1b activity is decreased without the attachment of RubC1a, but still some activity was measured due to the YbdZ gene. If *ybdZ* is deleted from *E. coli,* the RubC1b is completely inactive. So activity for RubC1b is dependent on MbtH-like proteins, just as for CloH, SimH and Pcza361.18 . Nevertheless, the RubC1 enzyme is able to catalyse and activate the adenylation of Tyrosine without the help of MbtH-like proteins and the catalytic efficiency with RubC1a is 6 times higher than with RubC1b+CloY. This shows that the activation of the adenylation domain can be achieved with something else than a MbtH-like protein. Fusion of the RubC1a to another protein domain can also be a possibility to achieve activation of the adenylation domain [25].

BIOLOGICAL APPLICATIONS

The functions of MbtH-like proteins are still unknown. But once they are known it will have a huge impact on several fields of Microbiology. Nowadays, most bacteria have already resistance towards antibiotics and there are only few resistance free strains. For that matter, it is important to find more compounds with antibiotic properties. MbtH-like proteins could help with combinatorial biosynthesis in which *mbtH*-like genes can facilitate expression of hybrid pathways. This is only possible if more information of MbtH-like proteins become available. During this process also improved production yields can be obtained, so multiple things can be accomplished. Earlier studies showed that Daptomycin[26] and A54145 [27] were established by genetic engineering. In both studies the MbtH-like genes, *dptG* and *lptG* were included in the design strategies, for the production of hybrid molecules, which contains the MbtH-like proteins and several NRPSs. But possibly the yield could be enhanced if the NRPS were co-expressed with a MbtH-like protein. This means that secondary metabolite production could be improved by strain improvement with MbtH-like proteins [1] and that difficulties producing secondary metabolites could perhaps be overcome through the addition of a MbtH-homolog. Another option is to use *mbtH* homologs to screen for a better production host to produce secondary metabolites. Most bacteria that contain NRPSs also contain MbtH-like proteins, which means that the *mbtH* homologs can serve as a beacon to identify potentially gifted strains, this leads to strains which contain new *mbtH* homologs that can be used for new production strains. Also the *mbtH* homolog can be used to identify new secondary metabolites. So in the end, four important biological applications can be improved if MbtH-like proteins are better understood: strain improvement, combinatorial biosynthesis, heterologous expression and genome mining.

CONCLUSION AND DISCUSSION

In this thesis, the important role of MbtH-like proteins for solubilisation, overexpression and activation of NRPS is discussed. Structural analysis of MbtH-like domains revealed the importance of three Tryptophan residues and the overall structure of the protein. The MbtH-like protein consists of a β-sheet bundle with 3 sheets, followed by an α-helix. It also revealed that none of the amino acids of MbtH-like proteins is involved in binding to the active site of the adenylation domain of NRPSs. This proves that there is still catalytic activity, but the MbtH doesn't contain an active site, so MbtH-like proteins are probably chaperons, stabilisers or facilitators.

The Tryptophan residues are important during the binding of the MbtH-like protein to the NRPS. Some data shows that especially the first two Tryptophan residues are important due to the formation of a small pocket, which is needed to form the interface between the MbtH-like protein and the NRPS. But the importance of the third Tryptophan residue is doubtful. Some data indicates that the third Tryptophan is only present in MbtH-like proteins, which are

attached through a small peptide chain to the NRPSs and that independent proteins need the third Tryptophan residue for the attachment to the NRPS, but the third residue contributes to the tertiary structures within the MbtH homologs[4]. This could be true since the crystal structure of SlgN1ΔAsubc did lack the third residue. Though SlgN1ΔAsubc is also fused to the A-core domain, but the third residue is still located within that protein. The structure revealed that even this Tryptophan is important during the formation of the interface, where it binds strongly to the NRPS. This means that all Tryptophan residues are conserved, most likely because of their important role during the attachment to NRPSs.

Another important thing to keep in mind is the function of MbtH-like proteins. All data available lacks a clear perspective on the biological function. The biological function could be predicted using structural information of similar proteins. The crystal structures of MbtH-like proteins is used for comparison using DaliLite to screen for possible structural similarities, during this search known structural information of other proteins or protein regions were compared to the MbtH-like protein structure, so structural alignments between two protein structures is performed The SlgN1ΔAsubc structure reveals the highest similarity to PA2412, which only differs from each other because of deviations in the loops at the N-terminal and C-terminal ends. The highest similarity with the A-core domain is found with PheA of *B. brevis*, which is considered to be an MbtH-independent NRPS because the genome lacks a *mbtH*-homolog. The difference between SlgN1ΔAsubc and PheA are mainly found at the interface, which isn't that surprising because of the lack of a MbtH-like protein in the genome sequence of *B. brevis* [9]. Also PA2412 is used for comparison, which results in two regions with 47 and 39 residues aligning the structure. Disappointingly they both have larger domains than the MbtH-like proteins and they lack the conserved residues, which are necessary for MbtH-like proteins. This means that structural information cannot provide information about the functions [6].

Luckily many depletion and reconstitution assays were performed and all revealed that the MbtH-like domains are necessary to partially or fully activate the adenylation domain of an NRPS. Even MbtH-like proteins from different species can complement each other, which indicates that they fulfil the same function. The only known independent NRPS in prokaryotes is Orf1[23], which doesn't need any MbtH-like protein to produce balhimycin. Many NRPS proteins in eukaryotic organisms are MbtH–like protein independent. Unfortunately the genome sequence of *Amycolatopsis balhimycina* was not screened for other MbtH-like proteins because of an incomplete sequence, which means that the result is inconclusive. Since other MbtH-like proteins may take over this function.

The existence of a protein-domain, which can actually replace MbtH-like proteins when attached to the NRPS, indicates that the attached proteins just fulfils the function of independent MbtH homologs and that the structure of the proteins is highly conserved. This means that there is a class of proteins that need MbtH-like proteins, a class which are MbtH-independent proteins and a class which contains a RubC1 like domain. Because of RubC1 and other data, it is known that MbtH-proteins have a specific function as a chaperone or facilitator of the adenylation domain of NRPSs [25].

REFERENCES

[1] 'Natural Products: The Secondary Metabolites', J. R. Hans, E. W. Abel, 2003, ISBN: 978-0-85404-490-0

[2] 'The chemical biology of modular biosynthetic enzymes', Jordan L. Meiera and   Michael D. Burkart, Chem. Soc. Rev., 2009,38, 2012-2045

[3] 'Nonribosomal peptide synthetases involved in the production of medically relevant natural products.' Felnagle EA, Jackson EE, Chan YA, Podevels AM, Berti AD, McMahon MD, Thomas MG., Mol Pharm. 2008 Mar-Apr;5(2):191-211

[4] 'Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*.' McMahon MD, Rush JS, Thomas MG, J Bacteriol. 2012 Jun; 194(11):2809-18

[4] 'Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery.' Baltz RH, J Ind Microbiol Biotechnol. 2011 Nov;38(11):1747-60

[5] 'Structural characterization of a 140 degrees domain movement in the two-step reaction catalysed by 4-chlorobenzoate:CoA ligase.' Reger AS, Wu R, Dunaway-Mariano D, Gulick AM., Biochemistry. 2008 Aug 5;47(31):8016-25

[6] 'The 1.8 A crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of *Pseudomonas aeruginosa*.', Drake EJ, Cao J, Qu J, Shah MB, Straubinger RM, Gulick AM., J Biol Chem. 2007 Jul 13;282(28):20425-34

[7] 'Solution structure of Rv2377c-founding member of the MbtH-like protein family.',Buchko GW, Kim CY, Terwilliger TC, Myler PJ., Tuberculosis (Edinb). 2010 Jul;90(4):245-51

[8] 'Conformation of twisted beta-pleated sheets in proteins.', Chothia C., J Mol Biol. 1973 Apr 5;75(2):295-302.

[9] 'Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes.', Herbst DA, Boll B, Zocher G, Stehle T, Heide L., J Biol Chem. 2013 Jan 18;288(3):1991-2003

[10] 'Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase.', Gulick AM., ACS Chem Biol. 2009 Oct 16;4(10):811-27

[11] 'Analyses of MbtB, MbtE, and MbtF Suggest Revisions to the Mycobactin Biosynthesis Pathway in *Mycobacterium tuberculosis*.', Matthew D. McMahon, Jason S. Rush, and Michael G. Thomas, J Bacteriol. 2012 June; 194(11): 2809–2818.

[12] 'Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin.', Quadri LE, Sello J, Keating TA, Weinreb PH, Walsh CT., Chem Biol. 1998 Nov;5(11):631-45.

[13] 'MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases.', Felnagle EA, Barkei JJ, Park H, Podevels AM, McMahon MD, Drott DW, Thomas MG., Biochemistry. 2010 Oct 19;49(41):8815-7

[14] 'N-Acylation during Glidobactin Biosynthesis by the Tridomain Nonribosomal Peptide Synthetase Module GlbF.', Heidi J. Imker, Daniel Krahn, Jérôme Clerc,Markus Kaiser, Christopher T. Walsh., Chemistry & Biology, Volume 17, Issue 10, 29 October 2010: 1077–1083

[15] 'Enzymatic Tailoring of Ornithine in the Biosynthesis of the Rhizobium Cyclic Trihydroxamate Siderophore Vicibactin.', John R. Heemstra, Jr., Christopher T. Walsh, and Elizabeth S. Sattely., J Am Chem Soc. 2009 October 28; 131(42): 15317–15329.

[16] 'Identification of the biosynthetic gene cluster for the pacidamycin group of peptidyl nucleoside antibiotics.', Wenjun Zhanga, Bohdan Ostash and Christopher T. Walsha, Proceedings of the National Academy of Sciences of the United States of America, vol. 107 no. 39:16828–16833

[17] 'MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor M145*.', Lautru S, Oves-Costales D, Pernodet JL, Challis GL., Microbiology. 2007 May;153(Pt 5):1405-12.

[18] 'Cloning and characterization of the biosynthetic gene cluster for kutznerides.', Fujimori DG, Hrvatin S, Neumann CS, Strieker M, Marahiel MA, Walsh CT., Proc Natl Acad Sci U S A. 2007 Oct 16;104(42):16498-503.

[19] 'Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins.', Zhang W, Heemstra JR Jr, Walsh CT, Imker HJ., Biochemistry. 2010 Nov 23;49(46):9946-7.

[20] 'NovG, a DNA-binding protein acting as a positive regulator of novobiocin biosynthesis.', Eustáquio AS, Li SM, Heide L., Microbiology. 2005 Jun;151(Pt 6):1949-61.

[21] 'Effects of deletions of mbtH-like genes on clorobiocin biosynthesis in *Streptomyces coelicolor*.', Wolpert M, Gust B, Kammerer B, Heide L., Microbiology. 2007 May;153(Pt 5):1413-23.

[22] 'Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis.', Boll B, Taubitz T, Heide L., J Biol Chem. 2011 Oct 21;286(42):36281-90

[23] 'The small MbtH-like protein encoded by an internal gene of the balhimycin biosynthetic gene cluster is not required for glycopeptide production.', Stegmann E, Rausch C, Stockert S, Burkert D, Wohlleben W., FEMS Microbiol Lett. 2006 Sep;262(1):85-92.

[24] 'Adenylate-forming enzymes of rubradirin biosynthesis: RubC1 is a bifunctional enzyme with aminocoumarin acyl ligase and tyrosine-activating domains.', Boll B, Hennig S, Xie C, Sohng JK, Heide L., Chembiochem. 2011 May 2;12(7):1105-14

[25] 'A domain of RubC1 of rubradirin biosynthesis can functionally replace MbtH-like proteins in tyrosine adenylation.', Boll B, Heide L., Chembiochem. 2013 Jan 2;14(1):43-4.
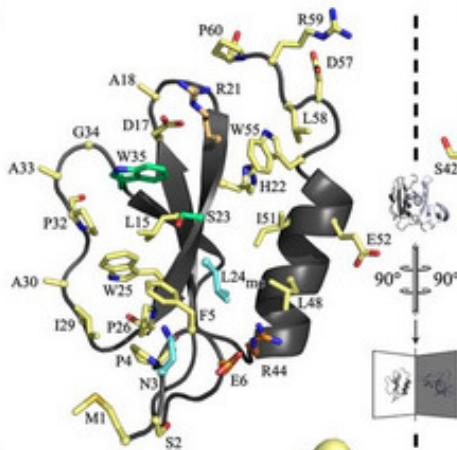
[26] 'Complementation of daptomycin dptA and dptD deletion mutations in trans and production of hybrid lipopeptide antibiotics.', Coëffet-Le Gal MF, Thurston L, Rich P, Miao V, Baltz RH., Microbiology. 2006 Oct;152(Pt 10):2993-3001.

[27] 'Development of a genetic system for combinatorial biosynthesis of lipopeptides in Streptomyces fradiae and heterologous expression of the A54145 biosynthesis gene cluster.', Alexander DC, Rock J, He X, Brian P, Miao V, Baltz RH., Appl Environ Microbiol. 2010 Oct;76(20):6877-87

[28] 'Motifs in the C-terminal region of the Penicillium chrysogenum ACV synthetase are essential for valine epimerization and processivity of tripeptide formation.' Wu X, García-Estrada C, Vaca I, Martín JF., Biochimie. 2012 Feb;94(2):354-64
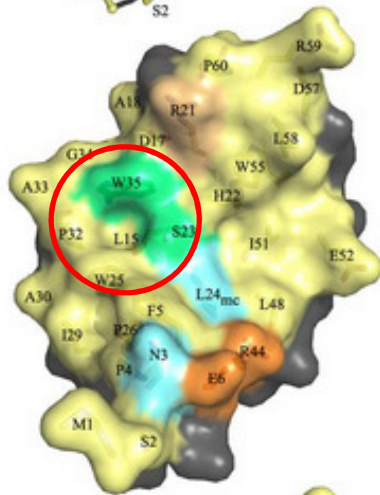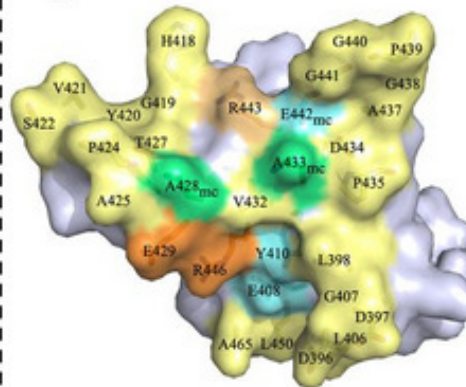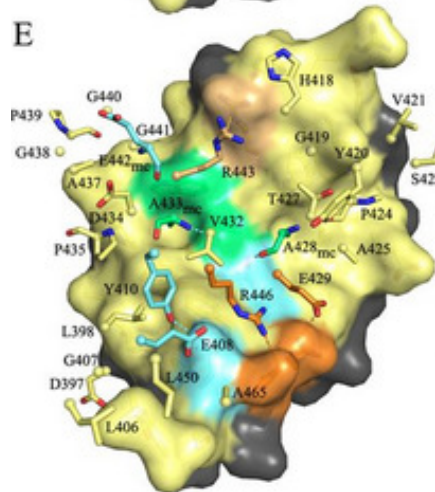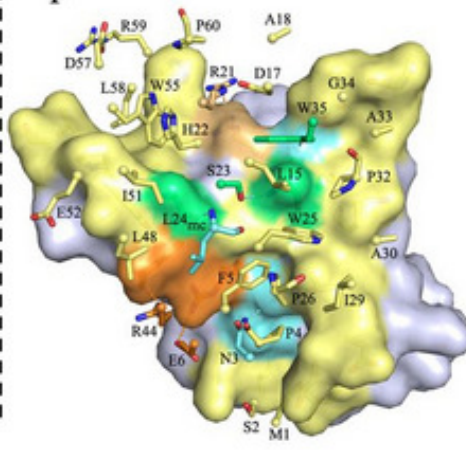
Attachment 1:

Attachment 1. The interface between the MbtH-like protein and the A-core domain of SlgN1. On the left side is shown the MbtH-like and at the right side is shown the A-core domain of SlgN1. The interface is flapped open along the inner axis. The red ring shows the mentioned flat ring, which is the negative imprint of Ala-433