

# Structural studies on DltA provide insights into the reaction mechanism of NRPS adenylation domains

## **Dissertation**

Zur Erlangung des akademischen Grades  
Doctor rerum naturalium (Dr.rer.nat.)

## **Vorgelegt der**

Naturwissenschaftlichen Fakultät I – Biowissenschaften  
der Martin-Luther-Universität Halle-Wittenberg  
Institut für Biochemie und Biotechnologie

Von

**Huma Yonus**

Geb. am 21. Mai 1979 in Aligarh, India

Gutachter:

1. Prof. Dr. Milton T. Stubbs, MLU Halle-Wittenberg
2. Prof. Dr. Henning Mootz, WWU Münster
3. PD Dr. Hauke Lilie, MLU Halle-Wittenberg

Verteidigt an 24.11.2015

Halle (Saale)

Dedication

*For my beloved  
parents, husband &  
kids.*

## **Acknowledgement**

This thesis dissertation marks the end of a very long and eventful journey for which there are many people that i would like to acknowledge.

First and foremost i would like to thank my PhD advisor, Prof. Dr. Milton Stubbs. He not only suggested an interesting topic, but was a source of good scientific knowledge and insightful discussions as well. I am very grateful to him for always encouraging me whenever i went through difficult times and most importantly for keeping his extreme patience with me.

I would also like to thank Prof. Dr. Mohamed A. Marahiel for collaborating with us on this project as well as allowing me to work in his laboratory in Marburg. Im also very thankful to Dr. Juergen J. May for providing me the DltA and DltC constructs as well as the DltA inhibitor. I also thank all lab members of the Dr. Marahiel group at Marburg.

I am highly indebted to Dr. Piotr Neumann for helping me in many aspects of structure solution. He introduced me to many crystallographic programs and was a source of immense knowledge in this field.

I sincerely acknowledge the co-operation of all lab members and technicians of the X-ray group. Im very thankful to all faculty members of this institute. My special thanks to late Prof. Dr. Rainer Rudolph, Dr. Elisabeth Schwarz as well Prof Dr. Ulbrich Hofmann for their moral support as well as valuable suggestions.

Im also highly grateful to Deutsche Forschungsgemeinschaft (DFG) for funding this project.

My special thanks to my husband, Ajmal for his remarkable patience, love and support during my research and also for the countless days of babysitting our beautiful daughter, Sarah. Im a proud mom of two daughters now (Sarah and Hiba) and i thank them both for bearing with me in many ways.

Above all i would like to acknowledge the tremendous sacrifices that my parents made to ensure that i had an excellent education. For this and much more, i am forever in their debt. Im also high thankful to my sisters and brothers for always being there for me.

## Zusammenfassung

DltA und DltC, die im Fokus dieser Arbeit stehen, werden durch das *dlt-Operon* kodiert, das in die D-Alanylierung der Teichonsäuren von Gram-positiven Bakterien involviert ist. Diese Modifikation hat einen wichtigen Einfluss in der Modulation der Oberflächenladung, der elektrochemischen Eigenschaften der Membran sowie in der Liganden-Bindung. DltA katalysiert die ATP-abhängige Aktivierung von D-Ala. Ausgehend von dem gebildeten energiereichen D-Alanyl-AMP-Intermediat wird D-Ala unter Ausbildung eines Thioesters auf den DltC-Phosphopantethein-Cofaktor übertragen. Sowohl die ATP-abhängige Aktivierung des Aminosäure-Substrates durch DltA, als auch die Involvierung von DltC als *carrier*-Protein ähnelt den nicht-ribosomalen Peptid-Synthasen (NRPSs). Diese modularen Enzyme sind an der Synthese von kleinen Peptiden mit vielseitiger pharmakologischer Relevanz beteiligt. Aufgrund der katalysierten Reaktion werden die Adenylierungs-(A)-Domänen der NRPSs der Familie der Adenylat-bildenden Enzyme zugeordnet. In den letzten Jahren gab es ein starkes Interesse an der Aufklärung der dreidimensionalen Struktur von Proteinen dieser Familie. DltA ist dabei ein besonderer Vertreter, da es, im Gegensatz zu NRPSs, die vorrangig L-Aminosäuren sowie Carboxysäuren aktivieren, die Aktivierung der D-Aminosäure D-Alanin katalysiert. Das aktivierte D-Ala wird anschließend auf den 4'-Phosphopantethein-Cofaktor des DltC übertragen, welches eine starke Ähnlichkeit zu den PCP-Domänen der NRPSs aufweist.

Beide Proteine wurden rekombinant hergestellt und durch den C-terminalen Hexahistidin-*tag* mittels Ni-NTA-Affinitäts-Chromatographie, gefolgt von einer Größenausschlusschromatographie gereinigt. Die Proteine wurden in hoher Reinheit erhalten und für umfangreiche Kristallisationsstudien eingesetzt. Die erhaltenen Kristalle wiesen gute Streueigenschaften auf. Jedoch war die Phasierung für DltC weder mittels *molecular replacement* unter Verwendung der vorhandenen NMR- und Röntgenkristallstrukturen, noch mittels Schwermetall-*soaking* erfolgreich. Selenomethionin-markiertes DltA wurde ebenfalls rekombinant hergestellt und die Kristallstruktur von DltA (Raumgruppe I222) konnte mittels SAD-Phasierung gelöst werden. Co-Kristallisations-Experimente von DltA und DltC resultierten in einer zusätzlichen Kristallform von DltA (Raumgruppe P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>), in denen weder Elektronendichte noch Platz für DltC vorhanden ist. Diese DltA-Struktur konnte mittels *molecular replacement* unter Verwendung der I222-Struktur als Model gelöst werden. Eine Interaktion zwischen DltA und DltC konnte weder mittels ITC, noch anhand nativer Gele beobachtet werden. Vermutlich ist die Interaktion zwischen den beiden Proteinen durch

eine niedrige Affinität charakterisiert, die *in vivo* durch zusätzliche Faktoren stabilisiert werden könnte.

DltA kann in eine große N-terminale und kleine C-terminale Domäne unterteilt werden und ähnelt den bereits beschriebenen Strukturen der Mitglieder der Familie der Adenylat-bildenden Enzyme. Das aktive Zentrum von DltA wird dabei von beiden Domänen ausgebildet. Obwohl D-Ala, MgCl<sub>2</sub> und ATP zur Kristallisation eingesetzt wurden, konnte nur Elektronendichte für ein AMP beobachtet werden. Durch Modellierung der D-Ala-Bindung in der Substratbindetasche konnte jedoch die Aminosäure C268 als potentieller Kandidat für die Determinierung der D-Ala-Spezifität identifiziert werden. Mutation von C268 zu Alanin resultierte in einer verringerten Substratspezifität innerhalb des ATP-PPi Austausch-Assays. Obwohl die Variante weiterhin D-Ala bevorzugt, konnte, im Vergleich zum Wildtyp-Protein, eine 10-fach höhere Aktivität gegenüber L-Ala beobachtet werden. Sequenzvergleiche ergaben, dass D-Ala aktivierende Enzyme von anderen A-Domänen durch das Vorkommen größerer Aminosäure-Seitenketten an Position 268 unterschieden werden können, da diese vorzugsweise Gly oder Ala an dieser Position aufweisen.

Für Adenylat-bildende Enzyme wurden unterschiedliche Orientierungen der beiden Domänen in Abhängigkeit des Stadiums der katalysierten Reaktion beobachtet. Die „offene Konformation“ repräsentiert die Liganden-freie Form, während die „Adenylierungs-Konformation“ mit der Adenylierung innerhalb der ersten Halbreaktion assoziiert ist. Die „Thiolierungs-Konformation“ ist dabei für die Thiolierung in der zweiten Halbreaktion verantwortlich. Für DltA konnte auch in Abwesenheit eines Akzeptors wie DltC oder CoA die Thiolierungs-Konformation beobachtet werden, sodass komplexere Mechanismen für die Konformationsänderungen angenommen werden können. Ein Vergleich der verfügbaren Strukturen sowie DltA-Modellierungs-Studien legen die Vermutung nahe, dass diese Enzyme in verschiedenen metastabilen Konformationen vorliegen können, die nur durch geringe energetische Unterschiede charakterisiert sind. Die Bindung der Substrate sowie die Produktbildungsreaktion führen dabei zu deutlichen Änderungen im aktiven Zentrum, die zusammen mit der gerichteten Bewegung der *hinge*-Region und des *P-loops* einen wichtigen Beitrag für die konformationelle Re-Orientierung liefern.

# TABLE OF CONTENTS

Chapters	Page no.
<b>1 Abstract</b> .....	<b>9</b>
<b>2 Introduction</b> .....	<b>11</b>
2.1 Cell wall of gram-positive bacteria.....	11
2.2 D-alanylation.....	12
2.3 Non-ribosomal peptide synthetases (NRPSs) .....	13
2.3.1 Producers of non-ribosomal peptides.....	13
2.3.2 Structural diversity of non-ribosomal peptides.....	13
2.3.3 The non-ribosomal machinery.....	14
2.3.4 Ribosomal verses non-ribosomal peptide synthesis.....	16
2.4 Other modular enzymes: FAS and PKS.....	17
2.5 The dlt operon.....	19
2.5.1 Pharmacological relevance of the dlt operon.....	22
2.6 The adenylation (A)-domains of NRPSs.....	25
2.6.1 Structural studies on adenylation domains reveal insights into the mechanism of substrate recognition and activation.....	27
2.6.2 Domain rearrangements among the adenylate forming enzymes.....	29
2.7 The peptidyl carrier protein-(PCP)-domain .....	31
2.8 The condensation-(C)-domain.....	32
<b>3 Objectives of the work</b> .....	<b>34</b>
<b>4 Materials and Methods</b> .....	<b>35</b>
4.1 Instruments, Materials and Chemicals.....	35
4.2 Softwares used.....	38
4.3 Synchrotron radiation facility.....	39
4.4 Expression of recombinant DltA.....	39
4.4.1 Cell disruption .....	39
4.4.2 Protein purification.....	40
4.4.3 Determination of protein concentration.....	40
4.4.4 Expression and purification of recombinant selenomethionine labelled DltA protein. ....	40

4.5	Expression and purification of recombinant DltC.....	41
4.5.1	Reversed phase HPLC analysis of DltC.....	41
4.6	ATP-PPi exchange assay.....	42
4.7	Isothermal titration calorimetry for studying interaction between DltA with DltC.....	43
4.8	Native gel shift assay.....	44
4.9	Protein crystallization of DltA and DltC.....	44
4.10	Measurements of protein crystals and data collection.....	45
4.11	Structure solution.....	46
4.12	Site directed mutagenesis of DltA.....	48
4.13	Sequence alignments.....	49
4.14	Modelling of functional states .....	49
<b>5</b>	<b>Results.....</b>	<b>51</b>
5.1	Recombinant protein expression, purification and characterization of DltA and DltC.....	51
5.1.1	Expression and purification of DltA.....	51
5.1.2	Expression and purification of DltC.....	52
5.1.3	Estimation of the degree of 4'-phosphopantetheine-containing DltC (holo-DltC) following <i>in vivo</i> post-translational modification.....	52
5.2	Substrate specificity of DltA.....	53
5.2.1	Activity assay for DltA.....	53
5.3	Studies on the formation of a DltA-DltC complex.....	54
5.3.1	Isothermal titration calorimetry to study complex formation of DltA with DltC.....	54
5.3.2	Native gel shift assay to test the presence of a complex between DltA and DltC.....	56
5.4	Crystallization of DltA and DltC.....	57
5.4.1	Crystallization of DltA.....	57
5.4.2	Crystallization of DltC.....	61
5.5	Structural studies on DltA and its comparison with members of AMP family.....	63
5.5.1	General features of the DltA structure.....	63
5.5.2	Overall topology of the structure.....	64
5.5.3	ATP binding pocket.....	66
5.5.4	D-alanine binding pocket.....	68

5.5.5	Domain reorganisation in DltA.....	71
5.5.6	The hinge region and P-loop.....	74
5.6	Conformation of DltA is not solely stabilized by crystal packing.....	75
<b>6</b>	<b>Discussion.....</b>	<b>78</b>
6.1	Stereo specificity of DltA is influenced by residue at position 268.....	78
6.2	Phylogenetic relationship of DltA to members of the superfamily.....	78
6.3	The DltA structure allows modelling of the different functional states of NRPS adenylation domains.....	84
6.3.1	Open conformation.....	84
6.3.2	Adenylation conformation.....	85
6.3.3	Thiolation conformation.....	87
6.4	Domain reorganisation as a dynamic feature of the superfamily.....	88
6.4.1	Role of Hinge region and P-loop.....	90
6.5	Role of domain alternation in catalysis.....	91
<b>7</b>	<b>Afterword.....</b>	<b>95</b>
<b>8</b>	<b>References.....</b>	<b>98</b>
<b>9</b>	<b>Appendix.....</b>	<b>108</b>
9.1	Structure of DltC.....	108
9.2	List of figures.....	110
9.3	List of tables.....	112
9.4	Abbreviations.....	113



# 1 Abstract

In this study, the proteins DltA and DltC belonging to the *dlt* operon were studied. The *dlt* operon is involved in the D-alanylation of teichoic acids of the gram-positive bacterial cell wall, a modification that plays significant roles in modulating the surface charge, electrochemical properties as well as the ligand binding abilities of these polymers. DltA catalyzes the activation of DAla at the expense of ATP to form a high energy D-alanyl-AMP intermediate, followed by a transfer of DAla as a thiol ester to the phosphopantetheinyl prosthetic group of the DltC. The ATP-dependent activation of the amino acid substrate by DltA, together with utilization of a carrier domain DltC for transport, closely resembles features of the non-ribosomal peptide synthetases (NRPSs), modular enzymes involved in the synthesis of small peptides of diverse pharmacological relevance. The reaction catalyzed by the adenylation (A) domains of NRPSs broadly classifies them into the superfamily of adenylate forming enzymes, and there has been much interest in 3-D structural analysis for this family in recent years. DltA is a unique example of an A-domain as it activates a D-amino acid i.e. D-alanine while most NRPS A-domains activate L-amino acids or carboxy acids. DltC subsequently picks up the activated D-alanine with its enzyme bound cofactor i.e. 4'-phosphopantetheine group and behaves much like the PCP domains of NRPSs.

Both proteins were expressed recombinantly with a C-terminal hexahistidine-tag and purified by Ni-NTA affinity chromatography followed by size exclusion chromatography. Diffraction quality crystals for both proteins could be obtained, however no phases for DltC could be obtained despite using molecular replacement with the existing NMR and crystallographic models. Heavy metals soaks were also not successful. DltA was overexpressed as selenomethionine labelled protein and the structure solved using the SAD phasing method (I222 form). A complex between DltA and DltC was also set up for co-crystallization, however the crystals obtained turned out to be a second crystal form of DltA ( $P2_12_12_1$  form) with neither density nor space for DltC. The structure of this crystal ( $P2_12_12_1$ ) could be solved by molecular replacement using the I222 structure as a search model. ITC and native gel shift assays failed to demonstrate complex formation, suggesting that a weak complex is formed between the two proteins which may be stabilized *in vivo* by other factors.

The overall structure of DltA was similar to the previously determined structures of the adenylate forming family, consisting of a large N-terminal domain and a small C-terminal domain with the active site in between. Despite being crystallized in the presence of its

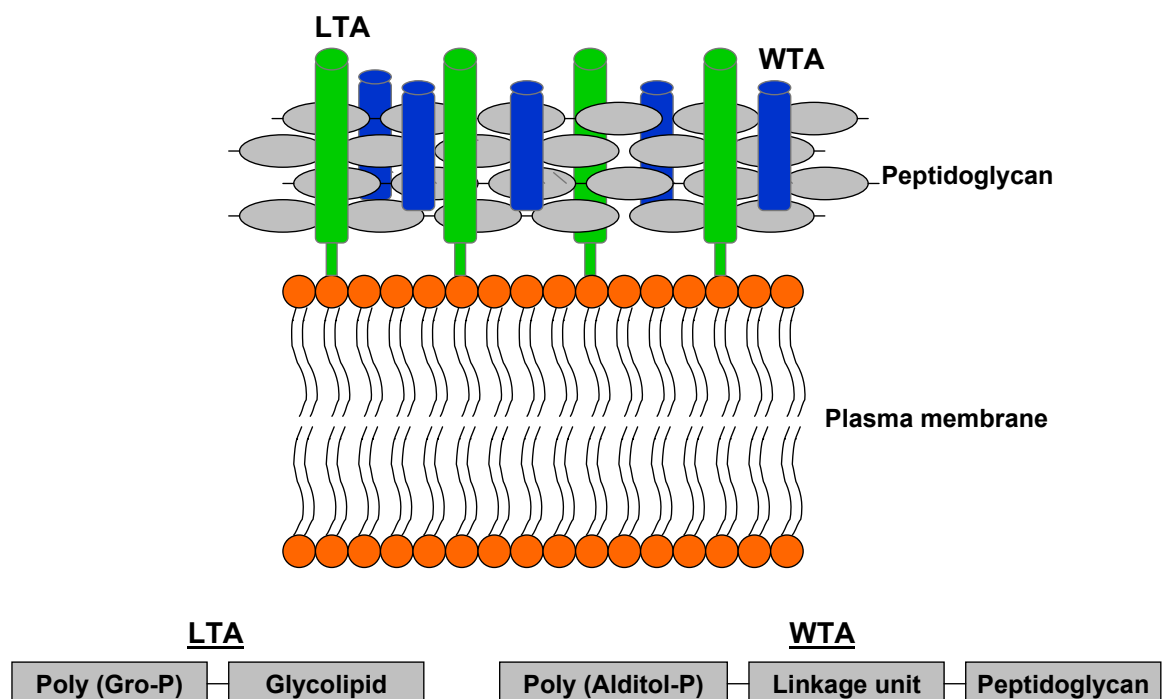
substrates, DAla, MgCl<sub>2</sub> and ATP, electron density was only apparent for the AMP moiety. Modelling of a DAla substrate into the binding pocket revealed determinants for the DAla specificity. C268 was selected as a potential candidate for DAla selectivity and mutated to an alanine. An ATP-PPi exchange assay showed that the mutation indeed relaxed the specificity: Although the mutant still exhibited high preference for DAla, it also showed a 10 fold higher activity for LAla than wt-DltA. Sequence alignments showed that the DAla activating enzymes can be distinguished from their counterparts by the presence of a bulky residue at position 268 while conventional A-domains have small Gly/Ala-268.

Different relative orientations of the two domains thought to be associated with the various catalytic stages of the adenylate forming enzymes have been observed: An 'open' conformation representing the unliganded state, an 'adenylation' conformation associated with the adenylation first half reaction and a 'thiolation' conformation held responsible for the thiolation second half reaction. In both crystal forms observed here, DltA displayed the thiolation conformation, even in the absence of thiol acceptors i.e. DltC or CoA, suggesting that the switch between the 'adenylation' and 'thiolation' conformations must be more complex than previously thought. Comparison of all the available structures as well as DltA modeling studies suggest that these enzymes exist in various metastable states separated by minor energy differences. Moderate changes around the active site brought about by substrate binding and product formation along with concerted movements of the hinge region and P-loop allow reorganization of the two domains, thereby driving the enzyme through the reaction cycle.

## 2 Introduction

### 2.1 Cell wall of gram-positive bacteria

The cell walls of gram-positive bacteria are built from macromolecular assemblies of cross-linked peptidoglycan, polyanionic teichoic acids (TAs), as well as surface proteins (Neuhaus and Baddiley, 2003). Teichoic acids constitute 30-60% of the cell wall and are basically polymers of 1,3-glycerol-phosphate (-Gro-P-) or 1,5 D-ribitol-phosphate(-Rbo-P-) linked via phosphodiester bonds. The teichoic acids are of two types: the wall teichoic acids (WTA) and the lipoteichoic acids (LTA) (Baddiley, 1962). WTA is mostly composed of either 1,3-glycerol-phosphate (-Gro-P-) or 1,5 D-ribitol-phosphate (-Rbo-P-) as repeating units and is phosphodiester linked to MurNAc residues of peptidoglycan via the linkage unit, i.e. (Gro-P)<sub>2</sub> or <sub>3</sub>ManNAc (β1-4) GlcNAc-P. LTA on the other hand is usually made up of poly(Gro-P) units and is hydrophobically anchored into the outer layer of the cytoplasmic membrane via its glycolipid moiety (Araki and Ito, 1989). The glycolipid is usually Glc(β1-6)Glc(β1-3)-(gentiobiosyl)diacyl-Gro in most genera of gram-positive bacteria (Fig. 1).

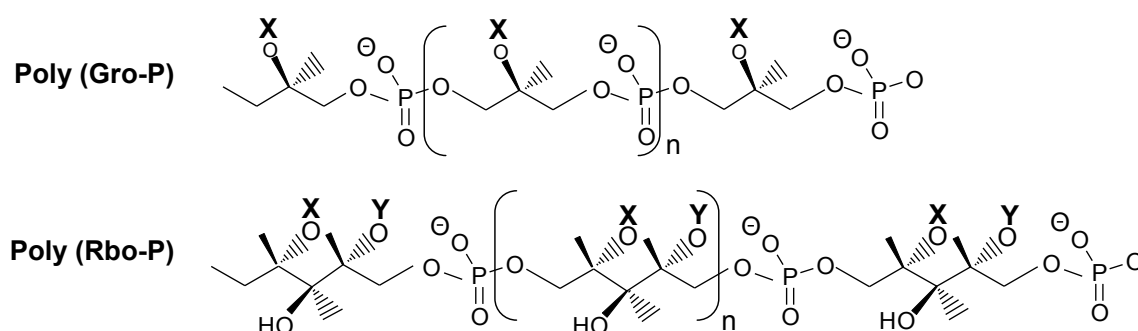


**Fig. 1. Cell wall of gram-positive bacteria showing the locations of lipoteichoic acid (LTA) in green and wall teichoic acid (WTA) in blue.**

LTA is hydrophobically anchored into the outer layer of the plasma membrane via a glycolipid moiety, while WTA is attached via a linkage unit to peptidoglycan.

## 2.2 D-alanylation

The teichoic acids are covalently modified with D-alanyl esters or glycosyl residues. The modification with D-alanyl esters always occurs at position 2 of either poly(Gro-P) or poly(Rbo-P) (Fig. 2) (Mirelman et al., 1970). When the 2'-OH of glycerol is substituted by a glycosyl unit as in group D streptococci, the D-alanyl esters are substituents on the sugar (Wicken and Baddiley, 1963). A wide structural diversity of teichoic acids is observed in the different genera of gram-positive bacteria that can be attributed to the presence and nature of D-alanyl esters, glycosyl substituents as well as the number of repeating units.



**Fig. 2. Monomers of teichoic acids showing the modifications with substituents.**

Position X may be occupied by either a hydrogen atom or a D-alanyl residue or an  $\alpha$ -glucosyl unit while position Y may be occupied by either  $\alpha$ -GlcNHAc or  $\beta$ -GlcNHAc (adapted from Neuhaus and Baddiley, 2003).

This modification of teichoic acids by D-alanine residues plays significant roles in modulating the biological activities of these polymers. Three roles of the D-alanylated LTA have been proposed: (a) modulation of the activity of autolysins; (b) maintenance of cation homeostasis and assistance in the assimilation of metal cations for cellular functions; and (c) definition of the electrochemical properties of the cell wall (Neuhaus and Baddiley, 2003).

Although LTA and WTA synthesis involves different enzymatic pathways, the D-alanyl esters of WTA are derived from those of D-alanyl LTA (Neuhaus and Baddiley, 2003; Perego et al., 1995). The synthesis of D-alanyl LTA in turn requires the *dlt* operon (DltA-D) (Abachin et al., 2002; Boyd et al., 2000; Clemans et al., 1999; Debabov et al., 2000; Perego et al., 1995; Peschel et al., 1999; Poyart et al., 2001). One interesting feature of the *dlt* operon is the two-step reaction catalyzed by DltA and DltC. DltA (D-alanine-D-alanyl carrier protein ligase) catalyzes the activation of DAla substrate at the expense of ATP to form a high energy D-alanyl-AMP intermediate, which is followed by transfer of DAla as a thiol ester to a phosphopantheynyl prosthetic group of the D-alanyl carrier protein, DltC. This ATP

dependent activation of the amino acid substrate by DltA, together with utilization of a carrier domain DltC for transport, closely resembles features of the non-ribosomal peptide synthetases (NRPSs).

NRPSs are modular enzymes involved in the synthesis of small peptides of diverse pharmacological relevance, including antimicrobial, antiviral, immunosuppressant and cytostatic agents (Finking and Marahiel, 2004; Schwarzer and Marahiel, 2001; Sieber and Marahiel, 2005). These multienzyme complexes are organized into modules that are responsible for the specific incorporation of monomeric components into the nascent peptide. Each module is further subdivided into a series of domains that catalyze individual steps of non-ribosomal peptide synthesis.

## **2.3 Non-ribosomal peptide synthetases (NRPSs)**

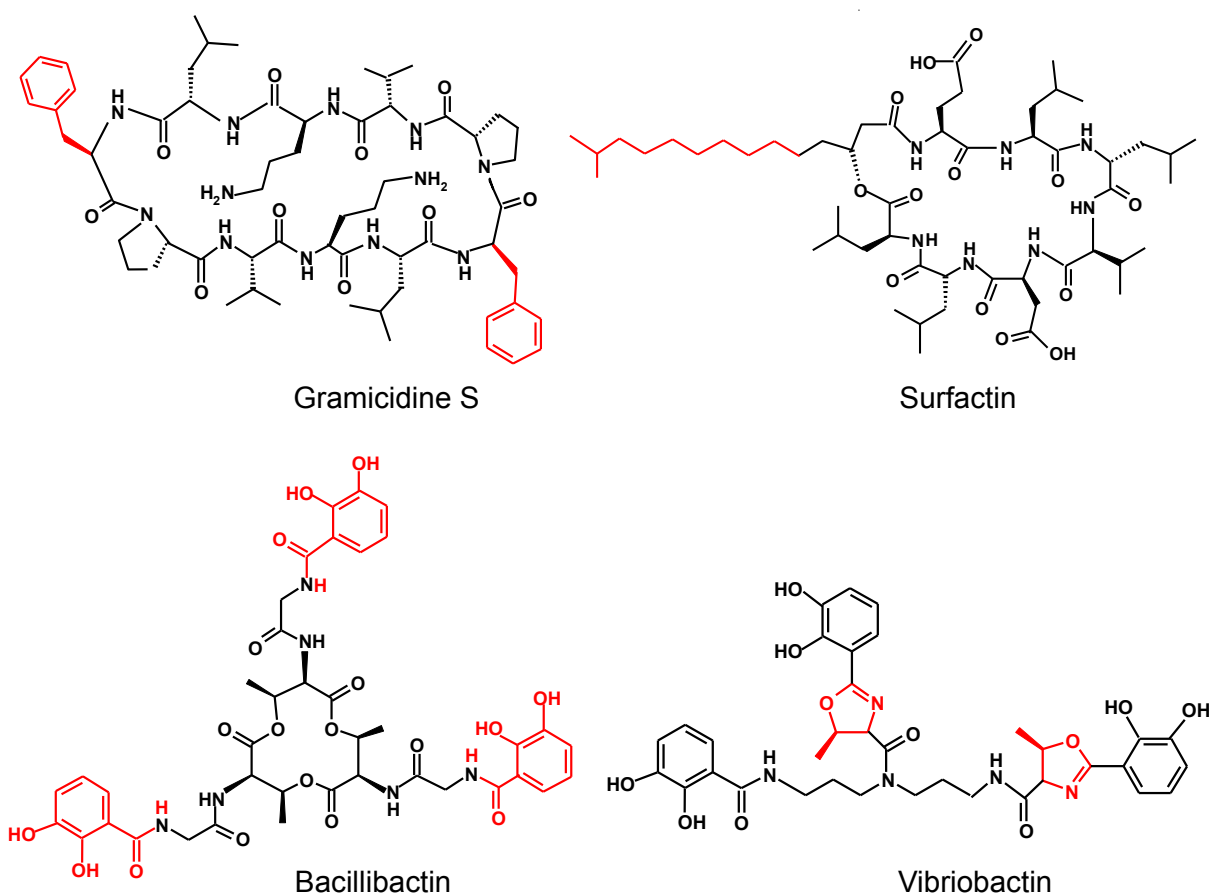
### **2.3.1 Producers of non-ribosomal peptides**

Non-ribosomal synthesized peptides are mostly produced by soil-inhabiting microorganisms, such as members of the gram-positive actinomycetes and bacilli, but also eukaryotic filamentous fungi. Many marine microorganisms also produce such secondary metabolites (Faulkner, 1998). Non-ribosomal synthesis of peptides is not limited to only simple organisms, in fact even higher eukaryotes harbour systems that resemble NRPSs. The mouse amino adipic acid semialdehyde dehydrogenase, U26, contains one set of adenylation/thiolation domains and is involved in the degradation of lysine to 2-amino adipic acid (Kasahara and Kato, 2003). Ebony from *Drosophila melanogaster* activates  $\beta$ -alanine for conjugation with biogenic amines to an aminoacyl-adenylate by an adenylation domain and covalently attaches it as a thioester to a thiolation domain in a non-ribosomal peptide synthetase (NRPS) related mechanism (Richardt et al., 2003).

### **2.3.2 Structural diversity of non-ribosomal peptides**

Non-ribosomal peptides are often cyclic molecules, built up of about 3 -15 amino acids, have a high density of non-proteinogenic amino acids and often contain amino acids connected by bonds other than peptide or disulfide bonds (Challis and Naismith, 2004). In addition, further modification such as N-, C- and O-methylation, acylation, glycosylation, heterocyclic ring formation, and conversion of the building blocks into their stereoisomer can be conducted by the enzymes. The released peptides can be linear, cyclic or branched-cyclic leading to macrocyclic lactams or lactones (Mootz et al., 2002). Non-ribosomally produced peptides

(Fig. 3) may contain: unnatural amino acids e.g. ornithine and D-phenylalanine in Gramicidin S (Hori et al., 1989),  $\delta$ -(L- $\alpha$ -amino adipic acid) in the ACV-tripeptide (Byford et al., 1997); fatty acids e.g. surfactin (Peypoux et al., 1999), mycosubtilin (Duitman et al., 1999), syringomycin (Hutchison and Gross, 1997); carboxy acids e.g. siderophores like bacillibactin (May et al., 2001), enterobactin (Gehring et al., 1998); heterocyclic rings (e.g. siderophores like vibriobactin (Marshall et al., 2001), bacitracin (Konz et al., 1997) as well as N-methylated residues e.g. cyclosporin (Thern et al., 2002) and pristanamycin.



**Fig. 3. Structural diversity of non-ribosomally produced peptides.**

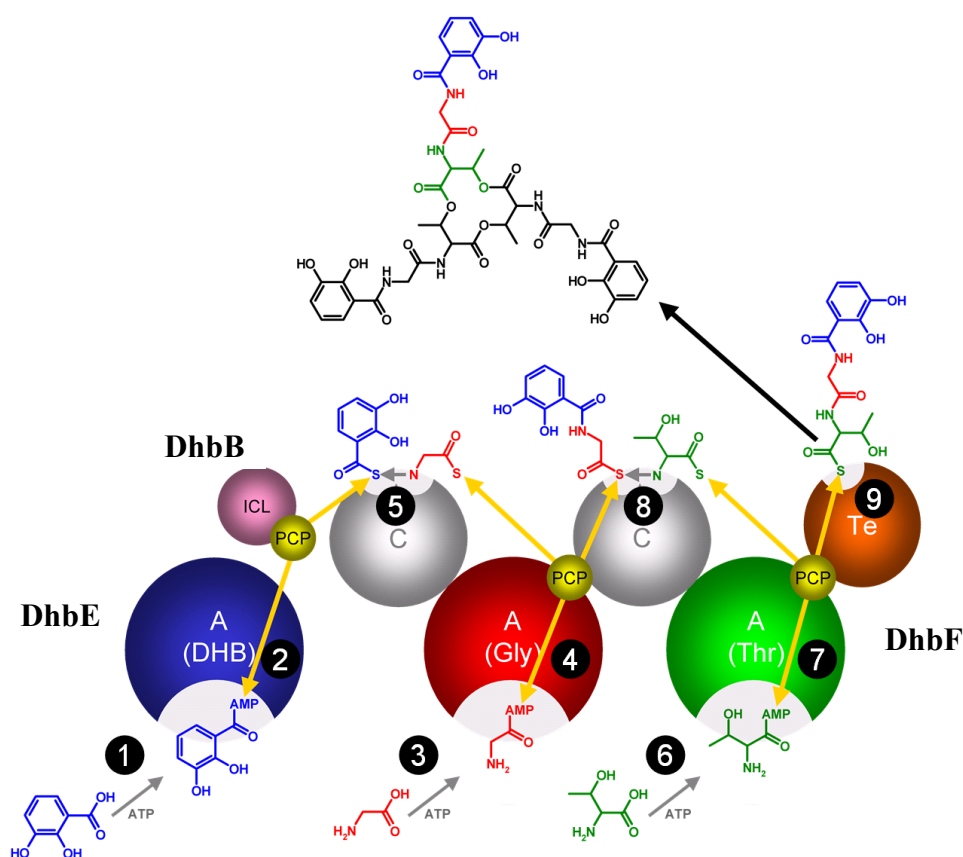
The characteristic structural features are highlighted in red.

### 2.3.3 The non-ribosomal machinery

NRPSs are built of repetitive catalytic units called modules. A module is a section of the NRPS polypeptide chain that is responsible for the incorporation of one amino acid into the final product (Marahiel et al., 1997). In all fungal NRPSs so far reported, these modules are part of one large polypeptide chain, whereas in bacterial systems these modules are often distributed on several enzymes (Schwarzer and Marahiel, 2001). The modules can be further

subdivided into domains, which represent the enzymatic units that catalyze the individual steps of non-ribosomal peptide synthesis (Stachelhaus and Marahiel, 1995). Three domains are necessary as the basic equipment of an NRPS elongation module: an adenylation (A)-domain for substrate recognition and activation as an acyl adenylate; a peptidyl carrier protein (PCP) domain for covalent binding of activated substrates as thioesters; and a condensation (C)-domain that catalyzes peptide bond formation. The last module of an NRPS usually contains an extra thioesterase (TE) domain to cleave off the fully assembled peptide from the PCP domain of the final module (Challis and Naismith, 2004). Some adenylation (A)-domains exist as separate enzymes i.e. not integrated within a module. Fig. 4 illustrates an example of such an NRPS with a stand-alone A-domain, DhbE, which like the stand-alone DltA enzyme, also activates a non-proteinogenic substrate. This NRPS is involved in the synthesis of a siderophore, bacillibactin (May et al., 2001).

In addition to these essential domains, extra enzyme activities can be found within some modules i.e. the epimerisation (E) domain, which is usually found after the PCP domains at the C-terminus of the module and catalyzes racemisation the PCP bound amino acid or of the C-terminal amino acid of the polypeptide chain (Stachelhaus and Walsh, 2000). The Methyltransferase (MT) domain may be inserted into the A-domain where it catalyzes the methylation of the nitrogen atoms of the peptide backbone (Turgay et al., 1992). The Cy-domain sometimes replaces the C-domain and catalyzes both heterocyclization of the side chains of serine, cysteine or threonine residues within the peptide backbone as well as peptide bond formation (Schwarzer et al., 2003).



**Fig. 4. Reaction cycle of an NRPS showing synthesis of the siderophore, bacillibactin from *Bacillus subtilis*.**

The stand-alone A-domain (DhbE) labelled as A (DHB) activates a 2,3 dihydroxybenzoate, which is subsequently transferred to the phosphopantetheine cofactor of the bifunctional isochorismate lyase/aryl carrier protein (PCP), DhbB. The dimodular DhbF adenylates threonine as well as glycine and covalently loads both amino acids onto their corresponding PCP domains. The condensation domains (C) of DhbF catalyze peptide bond formation to form the tripeptide, DHB-Gly-Thr. The thioesterase domain (Te) catalyzes the condensation of three DHB-Gly-Thr units and finally releases bacillibactin by intramolecular cyclization (adapted and modified from (Pfennig and Stubbs, 2012)).

### 2.3.4. Ribosomal versus non-ribosomal peptide synthesis

Ribosomal protein biosynthesis is a complex process requiring ribosomes, mRNA, tRNA, and several protein factors. Ribosomal peptide synthesis starts with activation of amino acids and the formation of the aminoacyl-tRNA, a reaction that is catalyzed by aminoacyl-tRNA synthetases. This is a two-step reaction; first the amino acid reacts with ATP forming an aminoacyl-AMP which is then transferred to tRNA leading to the formation of aminoacyl-tRNA. Each synthetase has a high degree of specificity for the correct tRNA and the correct amino acids. This specificity of the synthetase contributes to the first proof reading step and hence the accuracy of the translation process. The aminoacyl-tRNAs are then transported to the A site of the ribosomes so that they can form a complex with the mRNA. This process requires the elongation factor EF-Tu, a member of the G-protein family that binds the aminoacyl-tRNA only in the GTP form. GTP is hydrolysed to GDP only when an appropriate



complex between the EF-Tu-aminoacyl-tRNA complex and the ribosome is formed i.e. complementary base pairing between codon and anticodon has taken place, thus the free energy of GTP hydrolysis contributes to the second proof reading step of protein synthesis. Peptide bond formation then occurs when the correct aa-tRNA has been accommodated to the A-site, whereupon translocation can occur, regenerating the ribosome for subsequent rounds of synthesis (Finking and Marahiel, 2004).

As the name suggests, non-ribosomal peptide synthesis does not require ribosomes; instead, synthesis of short peptides occurs on large multi-enzyme complexes. The adenylation (A)-domains of NRPSs serve a similar function to the aminoacyl-tRNA synthetases, i.e. ATP dependent activation of substrates, although the two classes of enzymes share neither sequence nor structural homology (Eriani et al., 1990). NRPSs lack proof reading activity of aminoacyl-tRNA synthetases and often seem to have a relaxed substrate specificity (Conti et al., 1997; May et al., 2002; Stachelhaus et al., 1999). The activated substrates are subsequently transferred to a peptidyl carrier protein (PCP) domain which covalently attach them onto the thiol group of their 4'-phosphopantetheine (4'PP) cofactor, thus PCP domains serve as a transport unit, assuming a similar function to the tRNAs (Weber and Marahiel, 2001). Peptide bond formation is then catalyzed by the C-domains (Keating et al., 2000; Keating et al., 2002). The C-domains are believed to have an acceptor site for the nucleophile and a donor site for the electrophile thus when compared in terms of function, they exhibit some analogy to the A- and P-site of the ribosome, respectively. The absence of precision in non-ribosomal peptide synthesis has its own advantages: many bacteria and fungi can use this tool box for the synthesis of a wide spectra of secondary metabolites (not possible via ribosomal protein synthesis, which is accompanied by several proof reading steps because of the need for precision in primary metabolism) (Finking and Marahiel, 2004).

#### **2.4 Other modular enzymes: FAS and PKS**

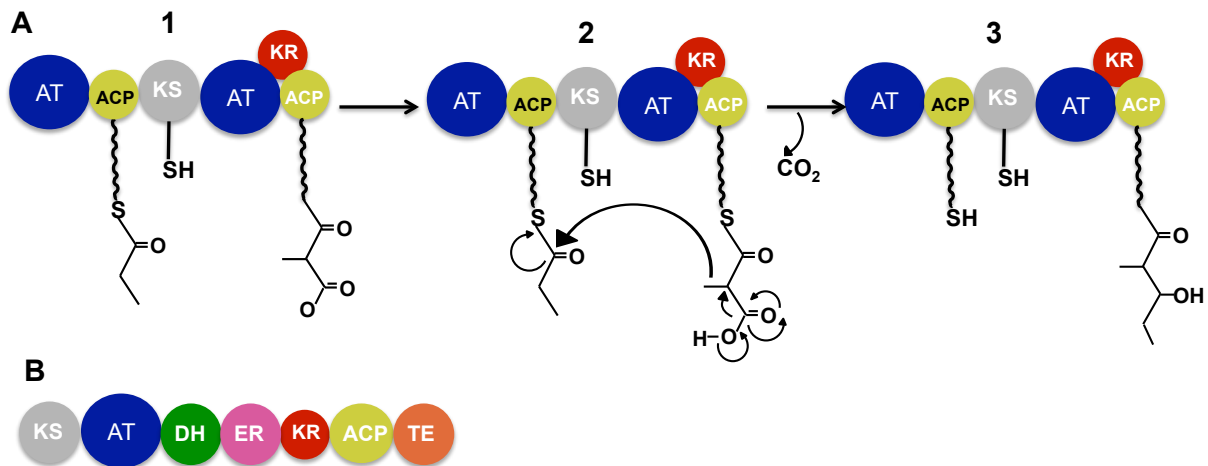
FAS and PKS are another group of modular enzymes that share a similar biosynthetic logic to the NRPSs. Fatty acids are a form of energy storage as well as forming building blocks for the cell and are thus important for primary metabolism. The PKS on the other hand generate a broad range of complex natural products i.e. polyketides, that includes plant flavonoids, fungal aflatoxins and a myriad of important pharmaceuticals that exhibit antibacterial, antifungal, immunosuppressive and antitumor properties (Finking and Marahiel, 2004; Smith and Tsai, 2007). There are two types of enzyme classes for both systems i.e. type I and type II

systems. In type I systems, the constituent catalytic components are covalently linked in multifunctional megasynthases like NRPSs (Joshi et al., 2003; Schweizer and Hofmann, 2004; Smith et al., 2003). Type I FASs are found in fungi and animals while type I PKS systems are found in bacteria, fungi, marine animals and plants. In contrast, the type II FAS systems (found in bacteria, chloroplasts and mitochondria) and the type II PKS systems (found in prokaryotes) are free standing monofunctional polypeptides, with each protein catalyzing a single step of multi-step reaction cycle (Black and DiRusso, 1994; White et al., 2005).

In both type I systems, the catalytic domains include a  $\beta$ -ketosynthase (KS), acyl transferase (AT), dehydratase (DH), enoylreductase (ER),  $\beta$ -ketoreductase (KR), acyl carrier protein (ACP) and thioesterase (TE). The animal type I FASs polypeptides contain a single copy of all seven functional domains, whereas the type I PKSs consist of multiple FAS-like ensembles, or 'modules', each containing a KS, AT and ACP with or without a full complement of the other catalytic domains (Fig. 5). In the type I FAS system, the same set of enzymes is used iteratively for all steps in the chain elongation process, whereas in the PKS system adjacent modules are used sequentially in an assembly line manner. The AT-domain catalyzes the transfer of activated substrate to the 4'PP of the corresponding holo-ACP-domain, thus the AT-domains can be compared to the A-domains while ACPs to the PCP domains of the NRPSs. In FAS, both the primer substrate (usually acetyl) and chain-extender moieties (usually malonyl) are loaded onto ACP phosphopantetheine by the same AT-domain, whereas loading of the primer (usually propionyl/acetyl) and chain-extender substrates (usually malonyl but also methyl-, ethyl-, or propylmalonyl) is catalyzed in PKSs by separate dedicated ATs located on different modules.

Another important difference is that the elongation reaction in FAS is performed by a single central ACP i.e. both the primer as well as extender moieties are loaded onto the same ACP, while PKSs have separate upstream and downstream ACPs. Thus ACP of the FAS complex plays a central role in shuttling substrate intermediates among the catalytic centres. During the elongation reaction, ACP passes the primer substrates over to the active site cysteine of the ketoacyl synthase (KS) domain. The subsequent extender moieties are collected either by the same (FAS) or another downstream ACP (PKS) and are then decarboxylated by the KS-domain to give the nucleophile that reacts with the thioester group of the ketide chain (primer substrate) attached to the KS-domain. The result is an ACP bound  $\beta$ -ketoacyl intermediate. The resulting  $\beta$ -ketoacyl product is either completely reduced (FAS) or is subjected to varying

degrees of  $\beta$ - carbon processing (PKS) by the action of KR, DH and ER domains. The Te-domain is responsible for the release of the final product. The multi-modular organisation of PKSs as well use of optional modification domains (KR, DH and ER) suggests close relationship between PKSs, FASs, and NRPSs with PKSs representing the mediator between primary (FAS) and secondary metabolism (NRPS). Currently high resolution structures are available for bacterial (White et al., 2005), fungal (Lomakin et al., 2007) as well as mammalian FAS (Maier et al., 2008) and for partial PKS modules (Tang et al., 2006)



**Fig. 5. (A) A dimodular PKS.**

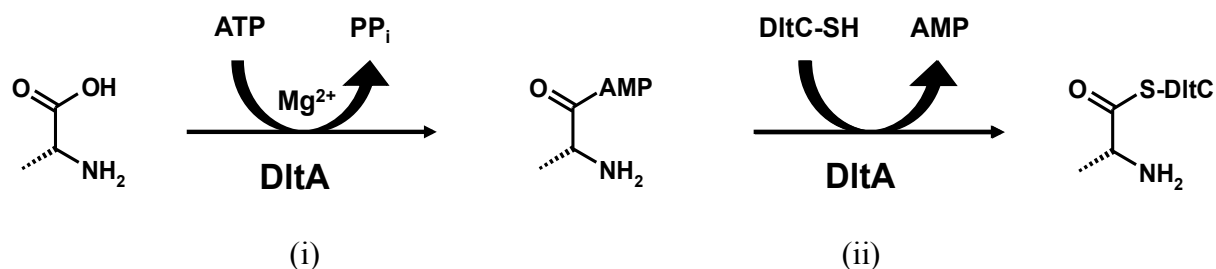
PKSs may also be multimodular like NRPSs while FASs are unimodular multidomain enzymes. The ACP of the first module is loaded with propionyl while the ACP of second module with methylmalonyl. The KR-domain has led to the reduction of the  $\beta$ -carbonyl to a hydroxyl group (adopted and modified from (Finking and Marahiel, 2004)) **(B) Domain organisation of an animal FAS** (adopted and modified from (Smith and Tsai, 2007))

## 2.5 The *dlt* operon

Genetic studies of LTA biosynthesis in various gram-positive bacteria have shown the incorporation of DAla residues into LTA requires the activity of four gene products (DltA to DltD), which are encoded by the *dlt* operon (Boyd et al., 2000; Clemans et al., 1999; Perego et al., 1995). Although the *dlt* operon comprises five genes, insertional inactivation of the fifth gene DltE had no effect on the incorporation of DAla ester content of both LTA and WTA, while the inactivation of any one of the genes, DltA-DltD, resulted in complete absence of DAla esters from both LTA and WTA (Perego et al., 1995).

DltA was first detected in the gram positive bacteria, *L. arabinosus*, *L. casei*, *B. subtilis* and *S. aureus* as an enzyme that activates a D-amino acids, D-alanine, using a pyrophosphoryl cleavage of ATP (Baddiley and Neuhaus, 1960). Subsequent cloning and expression in *E. coli*, clearly identified its role in the formation of the D-alanyl esters of membrane-bound

lipoteichoic acid. DltA, which is about 57kDa, closely resembles the adenylation domains of (A-domains) of non-ribosomal peptide synthetases (NRPSs) in catalyzing the ATP dependent activation of D-alanine forming the corresponding aminoacyl-adenylate [Fig. 6(i)]. The direct recognition of a non-proteinogenic amino acid by DltA makes it unique as most NRPS A-domains activate L-amino acids or carboxy acids, sometimes possessing epimerization domains that perform subsequent racemization to the D-form (Schwarzer et al., 2003).



**Fig. 6. The reaction catalyzed by DltA proceeds in two steps.**

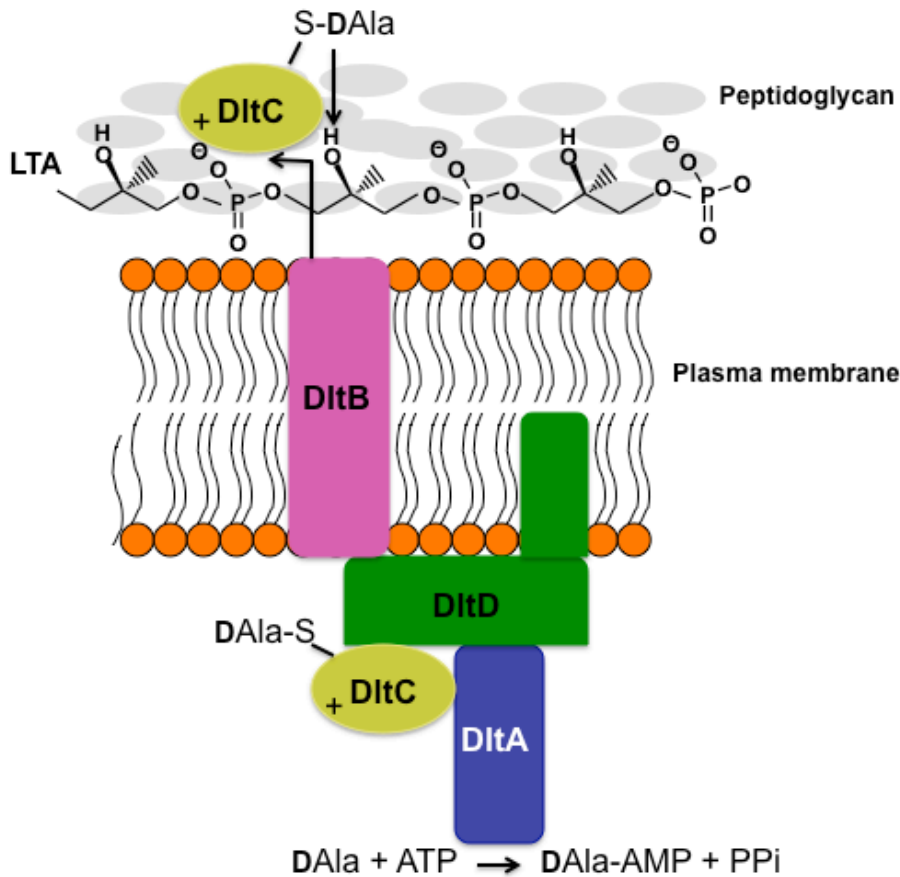
(i) the substrate DAla is adenylated with the release of pyrophosphate and then (ii) transferred to the phosphopantetheinyl prosthetic group of DltC *via* thioester formation.

DltC, which is about 9 kDa, encodes the corresponding D-alanyl carrier protein (DCP). DltC as well as ACPs and PCPs contain a highly conserved serine residue which is post-translationally modified with a 4'-phosphopantetheine prosthetic group in a reaction catalyzed by 4'-phosphopantetheine transferases (PPTases). DltC subsequently picks up the activated DAla with its enzyme bound cofactor 4'-phosphopantetheine and binds the amino acid covalently as a thioester [Fig. 6(ii)]. Despite low sequence identity (~ 20%), DltC was believed to be homologous to ACP of fatty acid synthesis, based on (i) the ability of DltC to be processed by holo-ACP synthase (AcpS) i.e. the PPTase for modification of ACP (ii) the ability of DltC and ACPs to be ligated with D-alanine in the reaction catalyzed by DltA, (iii) amino acid similarity around the phosphopantetheine attachment site, and (iv) the similarity of the predicted secondary structure (Debabov et al., 1996). The NMR structure of DltC from *L. rhamnosus* confirmed that the protein is indeed a true conformational homologue of the acyl carrier proteins (ACP-domains) involved in fatty acid and polyketide synthesis as well as to the PCP-domains of NRPSs (Volkman et al., 2001). All structures share a conserved three-helix bundle fold topology with few notable differences in lengths and orientation of helices. The presence of a specific binding site for LTA was proposed in DltC, based on earlier biochemical studies that showed that only D-alanyl-DltC was able to transfer D-alanine onto membrane associated LTA, while none of the acyl carrier proteins (ACPs) could replace the

function of DltC even though DltA ligates D-alanine to ACPs in the absence of DltD (Heaton and Neuhaus, 1994). Further it was also observed, that D-alanyl-DltC undergoes a time dependent hydrolysis when incubated with micellar LTA while D-alanyl-ACP was not hydrolysed (Kiriukhin and Neuhaus, 2001). An invariant Arg64 among the DltC family was proposed to participate in binding to the negatively charged poly(Gro-P) moiety of LTA (Neuhaus and Baddiley, 2003; Volkman et al., 2001).

DltD encodes a membrane protein which has no known homologue. It has a hydrophobic N-terminal sequence that is required to anchor this protein to the membrane (Fig. 7). In vitro assays showed that DltD bound DltC for ligation with D-alanine by DltA in the presence of ATP while the homologue of DltC, the *E. coli* acyl carrier protein (ACP), involved in fatty acid biosynthesis, was not bound to DltD (Debabov et al., 2000). Further the enzyme was seen to possess thioesterase activity and hydrolytically cleaved off D-alanyl-ACP if it becomes ligated with D-alanine which suggests that DltD might function in selection of the correct carrier protein, DltC, for ligation with D-alanine. These studies as well as others also suggest that DltD might also be involved in the catalysis of the final step of incorporation of DAla into LTA (Perego et al., 1995), however to support this hypothesis it would be necessary to first determine the topology of DltD (Debabov et al., 2000).

The exact role of DltB is unknown. The hydropathy profile of DltB from *B. subtilis* revealed the presence of 12 hydrophobic domains suggesting a transmembrane localization (Neuhaus et al., 1996). DltB showed sequence similarity to a variety of transport proteins especially proton antiporters that pump compounds from the cytosol at the expense of proton motive force. In light of these studies it has been proposed that DltB provides a channel for the secretion of D-alanyl-DltC to the site of D-alanylation (Neuhaus and Baddiley, 2003). DltB also showed similarity to undecaprenol phosphate transferases (Heaton and Neuhaus, 1994). Undecaprenol-P might be an intermediate membrane acceptor in the D-alanine incorporation system, and DltB might play a role in transferring D-alanine residues from DltC to undecaprenol phosphate.



**Fig. 7. Model for the incorporation of DAle into LTA by the enzymes of dlt operon.**

DltD might facilitate in the binding of DltA and DltC protein while transmembrane DltB might act as a channel for secretion of DltC to site of D-alanylation (adopted and modified from Neuhaus and Baddiley, 2003)

### 2.5.1 Pharmacological relevance of the dlt operon

Since the D-alanyl esters of LTA determine its net anionic charge and hence regulate the functions of this polymer, the role of these esters has been a point of recent investigation for several genera of gram-positive bacteria. The D-alanyl esters of LTA play important roles in the physiology of gram positive bacteria and several phenotypes have been observed, however there is no single phenotype that is common to all species examined i.e. aberrant cell formation (pleomorphs) resulting from inactivation of dlt operon was observed in *S. gordonii* (Clemans et al., 1999), *S. mutans* (Boyd et al., 2000), and *S. agalactiae* (Poyart et al., 2001) however no morphological changes could be observed in the dlt mutants of *S. aureus*, *S. xylosus* (Peschel et al., 1999), *B. subtilis* (Perego et al., 1995) and *L. monocytogenes* (Abachin et al., 2002).

A clear correlation between the D-alanyl ester content with the action of cationic antimicrobial peptides has been observed. *S. aureus* and *S. xylosus*, which tolerate high concentrations of several antimicrobial peptides, have shown increased sensitivity with the inactivation of the

dlt-operon. Sensitivity has been observed towards human defensin HNP1–3, animal-derived protegrins, tachyplesins, and magainin II, and to the bacteria derived peptides gallidermin and nisin (Peschel et al., 1999). The enhanced sensitivity to these cationic compounds is thought to result from the higher net polyanionic charge in the dlt mutants. It has also been reported that *S. aureus* strains resistant to glycopeptide antibiotics like vancomycin and teicoplanin, which represent the last options for the treatment of multidrug-resistant staphylococci, exhibited an at least threefold-increased sensitivity to these antibiotics in cells lacking DAla esters of teichoic acids (Peschel et al., 2000). The bactericidal action of human group IIA phospholipase A2 (PLA2) against the staphylococcal activity in human tears is enhanced by 30 to 100-fold with the inactivation of DltA, which has been attributed to increased  $\text{Ca}^{2+}$ -dependent activity of the bound PLA2 (Koprivnjak et al., 2002).

The D-alanine esters of LTA also play a role in virulence as established in *L. monocytogenes*, *S. agalactiae*, and *S. aureus*. The virulence of the DltA<sup>-</sup> mutant in *L. monocytogenes* was severely impaired in a mouse infection model and, *in vitro*, the adherence of the mutant to various cell lines (murine bone marrow-derived macrophages and hepatocytes and a human epithelial cell line) was strongly restricted (Abachin et al., 2002). It was suggested that the decreased cell adherence of the DltA<sup>-</sup> mutant might be as a result of the increased electronegativity of its charge surface and/or the presence at the bacterial surface of adhesins possessing altered binding activities. Inactivation of DltC in *S. mutans*, a pathogen that causes the formation of dental plaque also resulted in increased acid sensitivity and a defective ATR in the mutant strain (Boyd et al., 2000). It was speculated that the deficiency of D-alanyl esters is linked to increased proton permeability in this strain.

Reduction in the D-alanine content of the cell wall directly influences the autolytic mechanism. In *B. subtilis*, insertional inactivation of the genes of the dlt operon resulted in an increased rate of autolysis although the strain showed no aberrant morphology, cell growth or basic metabolism (Wecke et al., 1997). *S. faecalis* var. *zymogenes*, which produces a lytic agent (hemolysin/bacteriocin) that lyses walls and erythrocytes of gram-positive bacteria, conferred resistance against its own lytic agent (Davie and Brock, 1966). This strain contained a teichoic acid modified with D-alanine esters while some other streptococcal strains that were found sensitive to lysis produced an inactive ribitol teichoic acid which lacked the ester-linked D-alanine, therefore suggesting that the esters would appear to determine the number of binding sites on LTA for autolysins and removal of D-alanine from teichoic acid should enhance autolysis.

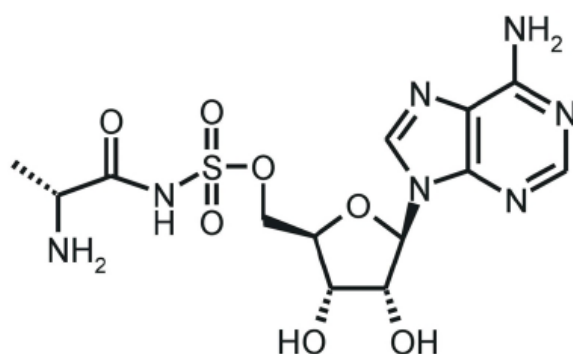
The D-alanine esters also play a vital role in the display of cell surface adhesins, bacterial coaggregation and formation of biofilms. Cell-surface adhesins on one cell type recognize and bind to complementary receptors on partner cell types. These interactions may lead to intergeneric as well as intrageneric coaggregation. *S. gordonii* DL1 (Challis) requires a 100 kDa surface adhesin for participating in intrageneric coaggregation (Clemans et al., 1999). Insertional inactivation of the *DltA* gene in this organism resulted in the loss of this adhesion, suggesting that this protein binds to D-alanyl LTA. Thus it seems that the DAla esters play a role in providing binding sites for this adhesin. Coaggregation also requires divalent metal cations, especially  $\text{Ca}^{2+}$ . It has been suggested that the D-alanyl esters of TA are required for the proper display of metal cations. It has been observed that when the ester content increases, the probability of the cation to bind to a single phosphodiester linkage (monodentate binding) as well as a mobile counterion increases (Neuhaus and Baddiley, 2003). This counterion could be replaced with an adhesin thus defining the role for the D-alanyl ester together with that of the cation. LTA with low D-alanyl ester content binds one cation for every two phosphodiester linkages (bidentate binding) so that the metal cation might no longer bind to the adhesin.

Production of bacterial biofilms is a major problem in hospitals as cells in biofilms are no longer susceptible to most antibiotics or to the immune response of the host. Biofilm formation involves two steps in which attachment of bacteria to a substrate surface is followed by the formation of multiple cell layers surrounded by a slimy matrix. Mutants of *S. aureus* lacking D-alanine residues in their LTA lose their ability to produce biofilms, while complementation of this mutant with plasmid bearing a copy of *dlt* operon not only restored the capacity to form biofilms but also led to production of stronger biofilm than the wild type (Gross et al., 2001). This increase corresponds with the higher amount of D-alanine esters for this complemented mutant due to presence of a high copy number plasmid. It is suggested that only the initial step of biofilm formation is affected because as the D-alanine content decreases there will be increase in net negative charge leading to electrostatic repulsion and thus preventing bacterial adherence.

Because of the pharmacological relevance of the *dlt* operon, a mechanism based inhibitor that specifically restrains the D-alanylation of the LTA and WTA in Gram-positive bacteria (May et al., 2005) was designed based on 5'-O-[N-(aminoacyl)-sulfamoyl] adenosine, nonhydrolysable analogues of amino acyl adenylates. Aminoacyl tRNA synthetases have



previously been inhibited by similar aminoacyl analogues (e.g. the alanyl-tRNA-synthetase from *E. coli* (Ueda et al., 1991)) and their use extended to the NRPS-system to inhibit the A-domains PheA and LeuA (Finking et al., 2003). DltA was observed to be inhibited by 5'-*O*-[*N*-(D-Alanyl)-sulfamoyl]-adenosine (Fig. 8) with a  $K_i$  value 60-fold lower than that of DltA with its cognate substrate D-alanine which suggested its suitability as an inhibitor (May et al., 2005). These studies also showed that the inhibitor was able to pass through the cell wall, reaching its target and improved effectiveness of antibiotics like vancomycin when used in combination. Thus such an inhibitor might be used either alone or in combination with antibiotics for treating pathogenic bacteria.

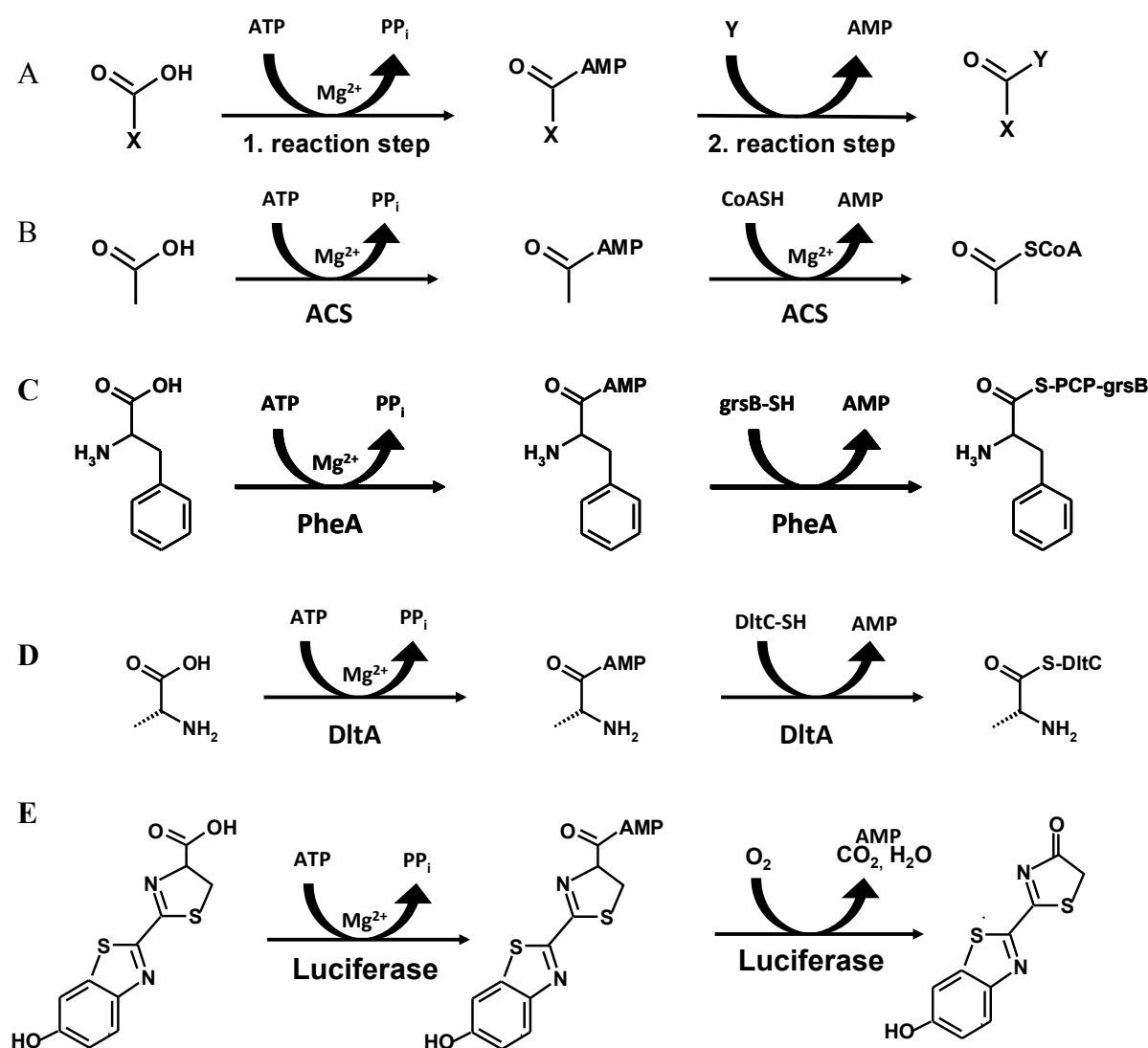


**Fig. 8.** The DltA inhibitor 5'-*O*-[*N*-(D-Alanyl)-sulfamoyl]-adenosine.

## 2.6 The adenylation (A)-domains of NRPSs

The adenylation (A)-domains of NRPSs belong to the superfamily of adenylate forming enzymes. This superfamily comprises three subfamilies: the acyl- and aryl-CoA synthetases, the adenylation (A)-domains of NRPSs and the firefly luciferase. All members of this superfamily catalyze a two-step reaction in which the first step is common to all, i.e. the ATP dependent activation of substrate to form an aminoacyl adenylate intermediate (Dieckmann et al., 1995), which is then used in a diverse set of second partial reactions (Fig. 9A). The enzymes of this superfamily share 20% sequence identity, are structurally homologous and catalyze very different overall reactions according to the second mechanistic step. This superfamily is widely spread in nature and comprises enzymes involved in both primary and secondary metabolism. The acyl-/ aryl-CoA synthetases activate a variety of different substrates, such as acetic acid, aromatic acids and long-chain fatty acids, to the corresponding acyl- adenylates, which are then transferred to the thiol group of CoA (Fig. 9B). The catalysis of acyl transfer to CoA is used by bacteria, plants and mammals for various metabolic functions (Gulick et al., 2004; Gulick et al., 2003; Jogl and Tong, 2004). As discussed in previous sections, the A-domains of NRPSs serve as entry points of substrates i.e. amino acids/carboxy acids that are activated as aminoacyl-adenylate intermediates and then

transferred onto the CoA-derived pantetheine cofactor of adjacent PCP domains (Fig. 9C) (Chang et al., 1997). The A-domains are about 550 amino acids in length and may exist either as an integrated part of an NRPS module or as single stand-alone enzymes. The DltA enzyme catalyzes a similar reaction to the NRPS A-domains, transferring the activated amino acid to the pantetheine cofactor of DltC (like a PCP domain), although neither protein is involved in peptide synthesis (Fig. 9D). Finally, firefly catalyzes the production of light in a similar two-step process; the second step involves an oxidative decarboxylation of the luciferyl adenylate intermediate however (Fig. 9E). The biological function of luminescence varies among species and ranges from distracting predators to attracting prey or mating partners.



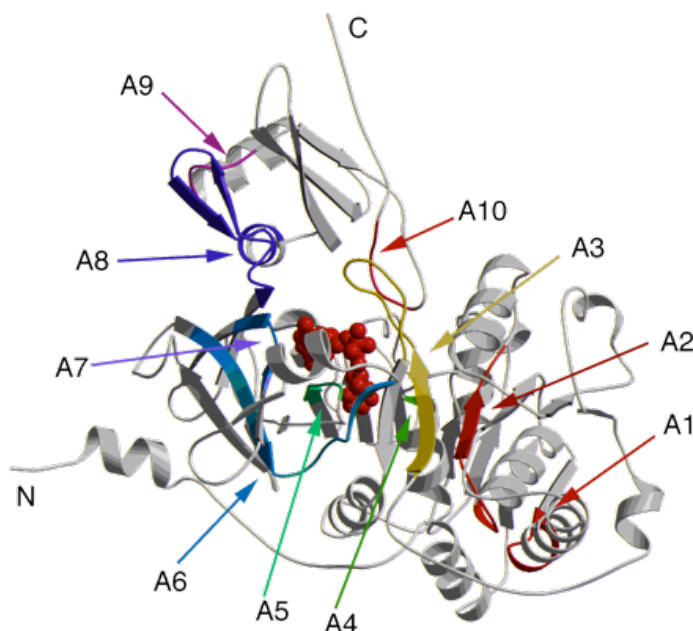
**Fig. 9. (A) Two-step reaction catalyzed by adenylate forming enzymes.**

Reaction step 1 involves formation of the activated acyl-AMP under hydrolysis of ATP (adenylation) while reaction step 2 involves the transfer of acyl-groups onto acceptor molecules by a nucleophilic attack on the carbonyl group of the acyl-AMP phosphoester. (B) Reaction catalyzed by Acetyl CoA synthetase (ACS) (C) Reaction catalyzed by the NRPS, PheA. (D) Reaction catalyzed by DltA. (E) Reaction catalyzed by the firefly luciferase.

### 2.6.1 Structural studies on adenylation domains reveal insights into the mechanism of substrate recognition and activation

The crystal structure of firefly luciferase from *Photinus pyralis* was the first of this superfamily to be reported (Conti et al., 1996). The structure confirmed the existence of a third scaffold for the ATP dependent formation of aminoacyl adenylates, in addition to those found in class I and class II aminoacyl tRNA synthetases. The enzymes fold into two distinct domains, a large N-terminal domain and a small C-terminal domain. [N.B the term “domain” has been used in the text to describe functional units within a module however the A-domains are further divided into two major structural domains, the N- and C-terminal (or large and small respectively) domains further discussed in this thesis].

The structure of the luciferase lacked any ligands and displayed an open active site; however the presence of distinct patches of conserved residues on the surfaces of the two domains facing each other suggested the approximate location of the active site. It was thus proposed that during the course of the reaction the two domains must come together to sandwich the substrates. This was later confirmed by the structure of the A-domain of gramicidin S-synthetase A, PheA from *B. brevis*, which clearly identified the active site of this enzyme with bound substrates phenylalanine as well as ATP and revealed the structural basis for substrate recognition and activation (Fig. 10) (Conti et al., 1997).



**Fig. 10. Crystal structure of PheA from *Bacillus brevis*.**

The substrates AMP and phenylalanine (red) are shown as spheres. The highly conserved core motifs (A1-A10) are marked in different colours (Conti et al., 1997)

The overall topology of PheA was found highly similar to the structure of firefly luciferase, although the two proteins only shared 16% identity in their primary sequence, suggesting that the other members of the adenylate-forming enzymes which show 30-60% similarity with PheA should have a very similar three-dimensional structure. The polypeptide chain of PheA folds into two compact domains, a large N-terminal and a smaller C-terminal portion, that are connected with a short hinge. The smaller C-terminal domain is rotated 94° with respect to the N-terminal domain when compared to the structure of luciferase. This rotation is interpreted as an active conformational change during the course of substrate recognition and activation since luciferase was crystallized in the absence of substrates.

The active site lies at the interface of the two domains. Ten highly conserved core motifs are found throughout the superfamily of adenylate-forming enzymes (table 1). As seen in the figure 10, these motifs are positioned around the substrate binding pocket and are thought to be involved in ATP binding, adenylation and hydrolysis reactions (Marahiel et al., 1997).

**Table 1: Core motifs of the A-domain**

<b>Core-motif</b>	<b>Sequence</b>	<b>Core-motif</b>	<b>Sequence</b>
A1	L(TS)YxEL	A6	GELxIxGxG(VL)ARGYL
A2	LKAGxAYL(VL)P(LI)D	A7	Y(RK)TGDL
A3	LAYxxYTSG(ST)TGxPKG	A8	GRxDxQVKIRGxRIELGEIE
A4	FDxS	A9	LPxYM(IV)P
A5	NxYGPTE	A10	NGK(VL)DR

Most residues involved in substrate binding are provided by the large N-terminal domain however, an invariant lysine residue (Lys517, core A10) located within the C-terminal domain provides a key interaction in the co-ordination of the  $\alpha$ -carboxylate group of the substrate amino acid. The core motif A3, called the P-loop, forms a signature sequence for this enzyme family, a motif also found in other ATP-binding proteins (Saraste et al., 1990). In the crystal structure of PheA, this loop is disordered suggesting high flexibility, however the subsequent structure of 2,3-Dihydroxybenzoic acid activating domain (DhbE) from *Bacillus subtilis* showed movement of this loop and it had been proposed that this loop plays an important role in coordinating the substrate ATP to initiate the adenylation reaction (May et al., 2002).

Based on the highly conserved three-dimensional structure of the A-domains, it was speculated that all other homologous A-domains of NRPSs, should have very similar main chain conformations and therefore substrate specificity should be mediated by the nature of residues lining the binding pocket of PheA. By performing sequence alignments of A-domains with PheA and by comparison of the residues that line the binding pocket of PheA with the corresponding moieties in other A-domains, it was possible to identify ten residues that are the major determinants of the substrate specificity of A-domains (Stachelhaus et al., 1999). Nine out of these ten residues lie within a 100-amino acids stretch between cores A4 and A5 of the A-domain, while only one residue, position 517 (core motif A10), lies outside this region. These residues are considered to be the “*codons*” of non-ribosomal peptide synthesis. Using the non-ribosomal code, it is now possible to predict the specificity of biochemically uncharacterized NRPSs simply by sequence analysis (Stachelhaus et al., 1999) and to alter substrate specificity of A-domains for the synthesis of novel peptides with modified biological and pharmacological properties. The subsequent availability of the crystal structure of DhbE (May et al., 2002) allowed the NRPS code to be extended to domains that activate carboxy acids.

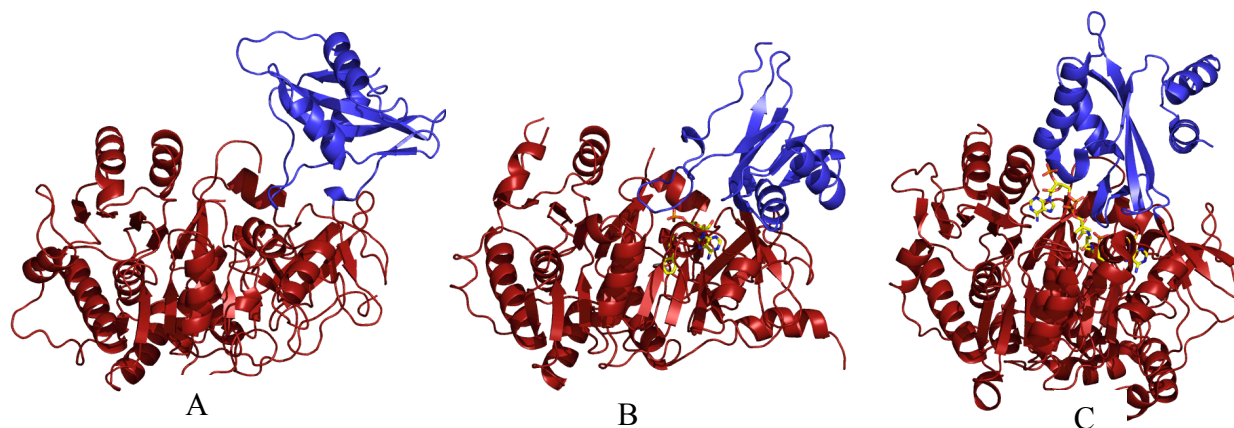
### **2.6.2 Domain rearrangements among the adenylate forming enzymes.**

Structural data are currently available for the following A-domains: firefly luciferase (Conti et al., 1996; Franks et al., 1998; Nakatsu et al., 2006) the phenylalanine activating A domain of gramicidin synthetase (PheA) from *Bacillus brevis* (Conti et al., 1997), the 2,3-Dihydroxybenzoic acid activating domain (DhbE) from *Bacillus subtilis* (May et al., 2002), acetyl-CoA synthetase (ACS) from *Salmonella enterica* (Gulick et al., 2003) and yeast (Jogl and Tong, 2004), the 4-chlorobenzoate:CoA ligase (CBAL) from *Alcaligenes sp.* (Gulick et al., 2004; Reger et al., 2008), long chain fatty acyl-CoA synthetase (LC-FACS) from *Thermus thermophilus* (Hisanaga et al., 2004), human medium-chain ACS (ACSM2A) (Kochan et al., 2009), DltA from *B. cereus* (Du et al., 2008; Osman et al., 2009) and the termination module of SrfA-C (Tanovic et al., 2008). These enzymes exhibit very similar secondary structural elements, folding essentially into a large N-terminal domain and a small C-terminal domain with the active site sandwiched between them. However, the relative orientations of the N- and C-terminal domains have been shown to vary between the various structures. An ‘open’ form conformation was first seen in the structure of the unliganded firefly luciferase (Fig. 11A). This enzyme catalyzes the ATP dependent activation of its substrate, D-luciferin, followed by an oxidation step to emit yellow-green light. In the open form, the large and the

small domains show only few interactions with each other leading to an open active site accessible to the bulk of solvent. Various forms of the open conformation have been observed in LC-FACS as well as adenylation domain of SrfA-C.

A second arrangement, first observed for PheA (Fig. 11B) and subsequently in DhbE, yeast ACS, CBAL, firefly luciferase (from *Luciola cruciate*), ACSM2A and DltA from *B. cereus* revealed the substrate binding pocket almost completely isolated from bulk solvent ('closed' conformation) through a close association of the two domains around the active site. When compared with the luciferase structure, the C-terminal domain of PheA is rotated by 94° relative to the N-terminal domain and is 5 Å closer to it. The conservation of crucial residues in the neighbourhood of the substrate-adenylate suggested that this conformation is responsible for the first half reaction, namely the formation of the substrate adenylate.

The third arrangement of the two domains, first observed for ACS from *Salmonella enterica* (Gulick et al., 2003) (Fig. 11C), LC-FACS (Hisanaga et al., 2004) and subsequently CBAL, (Reger et al., 2008) and different crystal forms of ACSM2A (Kochan et al., 2009) also reflects a 'closed' conformation; however, the C-terminal domain is rotated by approximately 140° with respect to the 'adenylation conformation'. The presence of coenzyme A in the crystals of *S. enterica* ACS (Gulick et al., 2003) suggested that this arrangement is responsible for the substrate transfer or 'thiolation half reaction'. It was therefore proposed that domain alternation, allowing a complete reconfiguration of the active site, is an important feature of the reaction cycle of adenylate forming enzymes, supported by mutational data on luciferase (Ayabe et al., 2005) and ACS (Reger et al., 2007), although it has been shown *in crystallo* that LC-FACS is able to catalyze the adenylation reaction in the 'thiolation conformation' (Hisanaga et al., 2004).

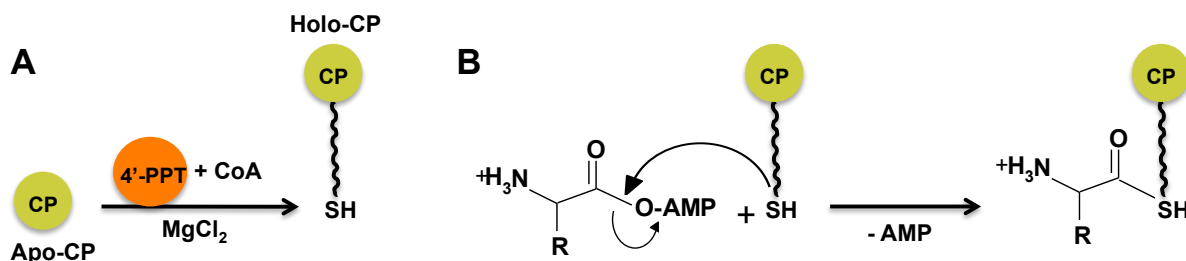


**Fig.11. Domain rearrangements observed in the adenylate forming enzymes.**

Three different orientations of the small C-terminal domains (blue) can be seen while the large N-terminal domains (maroon) are in equivalent orientation. (A) An ‘open’ conformation seen in the firefly luciferase crystallized in the absence of substrates. (B) The ‘adenylation’ conformation first seen in the PheA, reveals a closed active site with bound substrates phenylalanine and AMP (yellow). (C) The ‘thiolation’ conformation observed in the ACS crystallized in the presence of CoA and an inhibitor, adenosine-5'-propyl phosphate (yellow). The small domain in (C) is rotated by 140° with respect to that seen in (B) suggesting a reconfiguration of the active site.

## 2.7 The peptidyl carrier protein-(PCP)-domain

The PCP-domains of NRPSs, also known as carrier protein (CP) as well as thiolation (T) domains are about 80-100 amino acids in length and located downstream of the adenylation domain. PCPs are responsible for transportation of substrates and elongation intermediates between the catalytic centers. However the prerequisite for the functionality of the PCP is its modification with the 4'-phosphopantetheine (4'PP) cofactor which is post-translationally transferred from a coenzyme A to a highly conserved serine residue of the PCP by associated 4'-phosphopantetheinyl transferases (PPTases), thereby converting the inactive apo-PCP to its active holo-form (Fig. 12A) (Lambalot et al., 1996; Quadri et al., 1998). Activated amino acids get covalently attached to the thiol group of the PCP domain 4'PP cofactor as aminoacyl-S-PCP, which can now serve as a swinging arm to reach various catalytic centers (e.g. for condensation or modification domains) (Fig. 12B). This covalent tether, which is the hallmark of the so-called multiple carrier thio-template model of non-ribosomal peptide synthesis, is also found in polyketide as well as fatty acid synthesis and allows physical channelling through the multifunctional enzymes (Stein et al., 1996; Wu et al., 2001). These three classes form the CP superfamily.



**Fig. 12.** (A) Conversion of the carrier protein from the inactive apo-form to the active holo-form through transfer of the 4'PP cofactor of CoA by PPTases. (B) The free thiol moiety of the 4'PP cofactor serves to covalently bind the acyl reaction intermediates as thioesters.

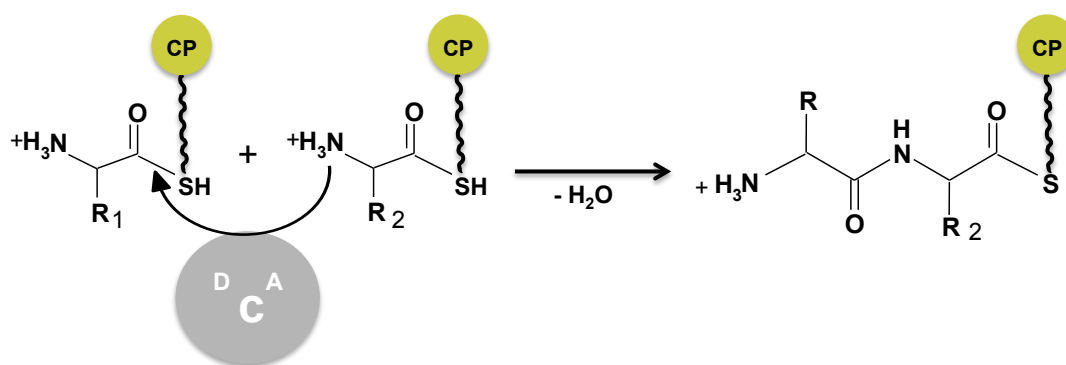
Although members of the CP superfamily assume different roles, their structures are quite similar, exhibiting a typical three-helix bundle fold topology (Crump et al., 1997; Holak et al., 1988; Parris et al., 2000; Volkman et al., 2001; Weber et al., 2000). The sequence similarity among the different members is also low apart from the sequence around the invariant serine residue (core T-motif), which serves as the attachment site for the 4'PP arm. The most obvious difference between ACPs and PCPs is the overall charge of the proteins. While ACPs have predominantly acidic side chains on their surface, the PCPs surface is much less polar, which corresponds to the charges of the corresponding phosphopantetheinyl-transferases AcpS and Sfp respectively. Previous NMR studies also revealed conformational diversity among PCP-domains (Koglin et al., 2006). Both apo-PCP and holo-PCP seemed to exist in two different stable conformations (A-state and A/H state for apo-PCP; H-state and A/H state for holo-PCP) of which one of these i.e. A/H state conformation is identical to both. A-state PCP was believed to be specifically recognized by Sfp (4'-PP transferase), while H-state PCP with SrfTEII (a thioesterase responsible for the regeneration of misprimed PCP domains). These studies also showed that the 4'-PP cofactor adopts two distinct positions in holo-TycC3-PCP, one being in proximity to the N-terminal end as seen in the A/H state and the other to the C-terminal end as seen in the H state, suggesting that structural rearrangements of the 4'-PP cofactor in the two holo-PCP forms modulates specific interactions with the A- and C-domains respectively. In contrast to these studies, the recent structure of an active site serine-to-alanine mutant of TycC3\_PCP in complex with Sfp and CoA revealed the PCP only in the A/H state conformation and no A-state conformation was observed (Tufar et al., 2014).

## 2.8 The condensation-(C)-domain.

Peptide bond formation in NRPSs is mediated by the C-domains which are about 450 amino acids in size. These enzymes catalyze peptide bond formation between the upstream



aminoacyl- or peptidyl-S-PCP moiety and the free amino group of the downstream aminoacyl-S-PCP, thus facilitating the translocation of the growing chain onto the next module (Fig. 13). According to some studies (Stein et al., 1996), C-domains should possess two distinct sites, one for the nucleophile (acceptor site) and another for the electrophile (donor site). This was also confirmed by the crystal structure of stand-alone VibH, which is an amide synthase resembling the NRPS C-domains and catalyzes peptide bond formation between DHB bound to its PCP, VibB, and the free diffusible substrate norspermidin (NS) (Keating et al., 2002). The structure of VibH shows that the protein consists of an N-terminal and a C-terminal face connected by a solvent channel. It has been proposed that the substrates, DHB-VibB and NS enter through the opposite sites of the enzyme i.e. C-terminal and N-terminal faces respectively, thus suggesting that the two faces would indeed correspond to the predicted donor and acceptor site of the C-domain. The crystal structure of SrfA-C, the termination module of the surfactin biosynthetic cluster revealed that the condensation domain closely associates with the adenylation domain to form a catalytic platform (Tanovic et al., 2008). The PCP domain, which is flexibly tethered to this platform, is believed to be stalled at the C domain's acceptor site in this structure. Based on the nature of interdomain linkers, it was speculated that both the C-domain and A-domain remain closely associated with each other to form a catalytic platform, while the PCP domain and the small C-terminal domain of the A-domain can rearrange to facilitate interaction of the PCP with both the adenylation and condensation domains.



**Fig. 13.** Peptide bond formation catalyzed by C-domains between 2 two adjacent PCPs loaded with their respective amino acids.

### **3 Objectives of the work**

The main objectives of this work were to gain insights into the catalytic reaction of DltA and its interactions with DltC. Although the stand-alone proteins DltA and DltC carry out different functions to NRPSs, their similarities to the A- and PCP-domains of these multi-enzyme complexes make them a good model system for understanding their more complex counterparts. DltA is particularly interesting as it is a stand-alone adenylation domain that specifically selects and activates a D-amino acid i.e. D-alanine (Perego et al., 1995) and so its structure could help provide a better understanding of the determinants for enantiomer selection. Both DltA and DltC could be expressed, purified and crystallized, and the structure of DltA solved by single wavelength anomalous dispersion (SAD). Based on the structure, biochemical analysis for the interaction of DltA with substrates DAla and DltC were also carried out. These studies provide insights into how substrates are channelled between different domains of an NRPS module.

## 4 Materials and Methods

### 4.1 Instruments, Materials and Chemicals

Table 2: Instruments and accessories used during the study.

Instruments and Accessories	Type and company
Automated DNA Sequencer	Li-Cor 4000 DNA-Sequencer, MWG-Biotech (Ebersberg)
Chromatography columns	HiTrap Chelating HP 5 ml, HiLoad 26/60 Superdex 75 prep grade, Amersham Biosciences (Freiburg) Vydac 218TP C4 reversed phase column
Centrifuges	Universal 32 R, Hettich, Biofuge pico, Heraeus, Avanti™ J-25 centrifuge, Avanti™ J-30 I centrifuge, Beckman (München)
FPLC-System	ÄKTA FPLC, Pump P-920, Monitor UPC-900, Valve INV-907, Mixer M-925 Fraction Collector Frac-900/901, Amersham Biosciences (Freiburg)
French Press	Gaulin –micron Lab 40 homogenisator, APV (Lübeck)
HPLC-System	GINA 50, Gynkotec (Idstein)
Photometer	Ultrospec 4000 UV/Visible Spectrophotometer Amersham Biosciences (Freiburg)
Scintillation counter	1900CA TRI-CARB Liquid Scintillation Analyzer Packard (Dreieich)
Thermocycler	Mastercycler gradient, Eppendorf (Hamburg)
Crystal Imager	OASIS L53 Protein crystal imaging system, Veeco (NY,USA)
Isothermal titration calorimeter	VP-ITC, MicroCal (Northampton, MA)

SDS PAGE assembly	Amersham biosciences (Freiburg)
Stereo microscope	Leica MZ 16
X-ray generator	R-AxisIV++, MSC Rigaku (USA)
Pipetting robot	Cartesian <sup>TM</sup> Dispensing system, Genomic solutions (USA)
Amicon Ultra-15, MWCO 5kDa, MWCO 30kDa	Millipore (Molsheim, France)

**Table 3: Kits**

Kit	Company
Crystallization kits House factorial screen Crystal screen 1, Crystal screen 2, Additive JBScreen Classic Kits 1-10	Prepared Inhouse (Jancarik and Kim, 1991) Hampton Research Jena Bioscience
QuickChange Site Directed Mutagenesis kit	Stratagene (La Jolla, USA)
QIAprep Miniprep	Qiagen

**Table 4: Enzymes and chemicals used for the study.**

Product	Company
Acetic acid	Roth
Acetonitrile	Roth
Acryl-/Bisacrylamide Rotipherese gel 30	Roth
Adenosin-5'-triphosphate (ATP)	Sigma
Adenosine-5'-[( $\alpha,\beta$ )-methyleno] triphosphate	Sigma
Agar Agar	Roth
Agarose	Roth
Ammonium chloride	Roth
Ammonium persulfate (APS)	Roth
Ampicillin	Sigma
Benzonase	Sigma
Bromophenol blue (BPB)	Roth
Coomassie Brilliant Blue	Merck

D-alanine	Sigma
5'-O-[N-(D-Alanyl)-sulfamoyl]-adenosine	Kindly provided by the group of Prof. Dr. Mohamed A. Marahiel, Marburg.
Disodium-Ethylenediaminetetraacetate-Dihydrate (EDTA)	Sigma
Disodium hydrogen orthophosphate (Na <sub>2</sub> HPO <sub>4</sub> )	Merck
Dithiothreitol (DTT)	Roth
Ethanol	Roth
Ethidium bromide	Riedel de Hæn
Glucose	Roth
4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic acid (HEPES)	Roth
Imidazole	Merck
Isopropanol	Merck
Kanamycin	Sigma
L(+)-Selenomethionine	Acros Organics
L-Isoleucine	Sigma
L-Leucine	Merck
L-Lysine	Sigma
L-Phenylalanine	Sigma
L-Threonine	Roth
L-Valine	Merck
Magnesium chloride	Merck
Magnesium sulphate (MgSO <sub>4</sub> .7H <sub>2</sub> O)	Roth
Monopotassium phosphate (KH <sub>2</sub> PO <sub>4</sub> )	Merck
N,N,N',N'-Tetramethyl-Ethylenediamin (TEMED)	Roth
Sodium-Dodecylsulfate (SDS)	Merck
PEG1000, PEG4000, PEG800	Fluka
[ <sup>32</sup> P]-Pyrophosphate ([ <sup>32</sup> P]-PPi)	Perkin Elmer
Rotiszint Eco Plus scintillation fluid	Roth

Sodium chloride	Riedel de Häen
Sodium hydroxide	Roth
β-mercaptoethanol	Sigma
Trifluoroacetic acid (TFA)	Roth
Trishydroxymethylaminomethane (Tris)	Merck
Trypton	Roth
Yeast extract	Roth

**Note:** The library of L-amino acids as well as D-amino acids used for the ATP-PPi exchange assays was provided in the form of already prepared stock solutions from the laboratory of Prof. Dr. Mohamed A. Marahiel, Philipps-University, Marburg.

**Table 5: Media**

LB Medium		M9- Medium (5X stock)	
Bactotrypton	10 g/L	Na <sub>2</sub> HPO <sub>4</sub>	30 g/L
Yeast extract	5 g/L	KH <sub>2</sub> PO <sub>4</sub>	15 g/L
NaCl	5 g/L	NH <sub>4</sub> Cl	5 g/L
		NaCl	2.5 g/L

LB-medium was used for the routine cultivation of *E. coli*, whereas M9-medium was used for the production of selenomethionine labelled DltA protein. All media were sterilized by autoclaving them for 30 minutes at 121°C and 1.5 bar pressure. The media were then cooled below 50°C and then supplemented with appropriate concentration of antibiotics.

For the M9-media the concentrated media was diluted to 1X in 1 litre of sterile water and to this was added 1ml 1M MgSO<sub>4</sub>.7H<sub>2</sub>O and 10 ml 20% glucose (carbon source)

## 4.2 Softwares used

Unicorn 4.11, Chromeleon, AxSys™ 81.4, Crystal Mine 1.025, Crystal Clear 1.3.5, Origin7, Pymol 0.99, O (Jones et al., 1991), CNS (Brunger et al., 1998), XDS, HKL, SHELX, LSQMAN (COLLABORATIVE COMPUTATIONAL PROJECT, 1994; Kleywegt, 1996), T-Coffee (Notredame et al., 2000)

### **4.3 Synchrotron radiation facility**

High resolution and SAD data sets were collected at the MPG/GBF wiggler beam line BW6/DORIS, DESY, Hamburg and PSF beam line, BESSY, Berlin.

### **4.4 Expression of recombinant DltA**

In this study the DltA gene was cloned and provided by Juergen J. May (May et al., 2005). This gene had been amplified from chromosomal DNA of *B. subtilis* MR168 and cloned into pQE60 vector (Qiagen, Hilden, Germany). The vector encoding the recombinant DltA with a C-terminal tag (RSHHHHHH) was then transformed into *E. coli* M15 [pREP4] cells (Qiagen, Hilden, Germany) using electroporation technique. *E. coli* M15 cells additionally contain the helper plasmid pREP4 that constitutively expresses the *lac* repressor protein that is inactivated by IPTG thereby inducing expression of recombinant protein. Glycerol stocks (15%) of cells grown to logarithmic phase were prepared and stored at -80°C. Using a flamed wire loop, an inoculum of bacteria was streaked across an LB agar plate containing the antibiotics ampicillin (100µg/ml) and kanamycin (25µg/ml). The plates were incubated at 37°C for 12 hrs and then stored at 4°C. A preculture was prepared by inoculating well isolated colonies from a freshly streaked plate into 50 ml of LB-medium, incubating it at 37°C for 12 hrs (overnight). For large-scale expression, 10 ml of overnight culture was then inoculated into 1000 ml of prewarmed LB-medium containing the appropriate concentration of antibiotics. The culture was then incubated at 37°C, shaking it at 250 rpm, to attain an OD600 of 0.6-0.7. At this stage the culture was induced with 1mM IPTG and was incubated further for 3-4 hrs at 37°C (250 rpm). The cells were then harvested by centrifugation at 6000 rpm for 15 minutes, weighed and stored at -20°C.

#### **4.4.1 Cell disruption**

The biomass from 1L expression was thawed and resuspended in 40ml of Hepes buffer A containing 10mM imidazole to reduce non-specific binding. The cells were then lysed by a french press, giving it 3-4 passages at a pressure of 900 bars. The cell lysate was further treated with benzonase for 30 minutes at room temperature to degrade nucleic acids and then centrifuged at 30,000 rpm for 45 minutes at 4°C, supernatant was collected filtering it through 0.2µM millipore filter and stored on ice.

#### **4.4.2 Protein purification**

The protein was purified in two steps ie Ni-NTA affinity chromatography and size exclusion chromatography. All purification steps were performed on a fast protein liquid chromatography (FPLC) system (Amersham Biosciences). For affinity purification, the supernatant after cell disruption and lysis (40ml) was applied on a 5ml prepacked Ni-NTA column (Amersham biosciences), that had been equilibrated with Hepes buffer A (50 mM Hepes, 300 mM NaCl, 10 mM Imidazole, pH 7.8), at a flow rate of 1ml/minute, collecting the flow through via a fraction collector. After binding, the column was washed with the same buffer to obtain a stable baseline. The protein was eluted with Hepes buffer B (50 mM Hepes, 300 mM NaCl, 250 mM Imidazole, pH 7.8), using a linear gradient of 150 minutes. The different fractions were then collected in 1.5 ml eppendorf tubes and kept on ice. The fractions containing the recombinant protein were then pooled together and concentrated to a volume of 2 ml using amicon ultra-15 centrifugal filter units having a membrane with a pore size of 30 KDa at a speed of 5000g. The concentrated protein was subsequently applied on a superdex G-75 gel filtration column (Amersham Biosciences) pre-equilibrated with Hepes buffer A (50 mM Hepes, 300 mM NaCl, pH 7.8). The different fractions were collected, further concentrated to 20 mg/ml and stored at -20°C.

#### **4.4.3 Determination of protein concentration.**

The concentration of the protein was determined based on ultraviolet (uv) absorption at a wavelength of 280 nm. For this 500 µl of diluted sample was placed in 1 cm quartz cuvette and absorption was measured at 280 nm . The theoretical molar extinction co-efficient i.e.  $49650 \text{ M}^{-1} \cdot \text{cm}^{-1}$  determined by the expasy proteomics server was used for calculations.

#### **4.4.4 Expression and purification of recombinant selenomethionine labelled DltA protein.**

Selenomethionine labelled DltA was prepared for the determination of phases by the method of single wavelength anomalous dispersion (SAD) (Hendrickson et al., 1990). For the incorporation of selenomethionine into proteins, either the cells producing the protein of interest should be auxotrophic for methionine (Cowie and Cohen, 1957) or alternatively methionine biosynthesis is suppressed (Doublet, 1997) by growing a nonauxotrophic prokaryotic strain in the presence of high concentrations of isoleucine, lysine and threonine, blocking *E. coli* methionine biosynthesis and forcing incorporation of selenomethionine from the medium into newly synthesized proteins.



For this, overnight cultures of *E. coli* M15 [pREP4] cells was prepared in LB-medium at 37°C. 1 ml of this culture was centrifuged and cell pellet was resuspended in 1ml M9 medium with a carbon source at 4g/litre and then this was used to inoculate 1 liter of prewarmed M9 medium containing the antibiotics ampicillin (100µg/ml) and kanamycin (25µg/ml). The culture was incubated at 37°C, and allowed to grow to mid-log phase (OD<sub>600</sub> 0.4), before addition of the amino acids: lysine, phenylalanine and threonine at 100mg/litre; isoleucine, leucine, and valine at 50 mg/litre and L-selenomethionine at 60 mg/litre. The culture was induced with 1mM IPTG after 15 minutes of addition of the amino acids and incubated overnight at 37°C shaking it at 250 rpm. The cells were harvested by centrifugation at 6000 rpm for 20 minutes and further stored at -20°C. Purification strategy was similar to that for the native protein except that all buffers used where degassed and also contained 2 mM β-mercaptoethanol.

#### **4.5 Expression and purification of recombinant DltC**

Cloning of the DltC gene was done by Juergen J. May (May et al., 2005). This gene was also amplified from chromosomal DNA of *B. subtilis* MR168 and cloned into pQE60 vector (Qiagen, Hilden, Germany). This protein also contained a C-terminal his-tag (RSHHHHHH) like DltA, however the expression strain used was *E. coli* BL-21 (Gold). The expression was done as described for DltA at 37°C in the presence of ampicillin (100µg/ml).

Purification also involved two steps ie Ni-NTA affinity chromatography and size exclusion chromatography. Purification scheme was similar to DltA except that for affinity chromatography, the binding buffers contained 5mM imidazole and elution was done with a linear gradient of 60 minutes. For Gel filtration, the buffer used was 50 mM Tris/HCl pH 7.0. The samples containing the purified protein were pooled together and concentrated using amicon ultra-15 centrifugal filter units having a membrane with a pore size of 5 KDa at a speed of 5000 g to 20mg/ml and stored at -20°C.

##### **4.5.1 Reversed phase HPLC analysis of DltC**

Reversed phase HPLC was performed in order to assess the amount of DltC expressed as holo protein bearing the phosphopantethiene cofactor (i.e. holo-DltC). In this technique, proteins with differing hydrophobicity are separated based on their reversible interaction with the hydrophobic surface of a chromatographic medium. Samples bind as they are loaded onto a

column (stationary phase) and then conditions are altered so that the bound substances are eluted differentially.

110  $\mu$ l of DltC at a concentration of 0.1mg/ml in 50mM Tris-HCl, PH 7.0 was applied on a C4 column (Vydac 218TP, 250mm, particle size 5  $\mu$ m). The column material consists of n-butyl groups bound to a 300 Å pore width silica gel. The column was pre-equilibrated with 0.05% (v/v) Trifluoroacetic acid (TFA) in ddH<sub>2</sub>O, and the protein was eluted using a linear gradient of 0-95% in 30 minutes with 0.05% (v/v) Trifluoroacetic acid (TFA) in acetonitrile. All measurements were performed at room temperature. The flow rate used was 0.7 ml/min and detection was performed at wavelengths of 205, 215, 260 and 280 nm.

#### **4.6 ATP-PPi exchange assay**

This radioactive assay is based on the adenylation reaction catalyzed by adenylate forming enzymes in which the substrate amino acids reacts with ATP to form an aminoacyl-adenylate intermediate with the cleavage of pyrophosphate (PPi). This reaction is reversible and so addition of tracer amounts of labelled [<sup>32</sup>P]-pyrophosphate to the reaction mixture will lead to its incorporation into ATP. This ATP will bind to activated charcoal which is then collected and the radioactivity is subsequently determined by liquid scintillation counting. The degree of adsorption of radioactive ATP to activated charcoal is a direct measure of enzymatic activity of the protein.

The ATP-PPi exchange assay was performed as described by Stachelhaus *et al* (Stachelhaus *et al.*, 1999) to examine the activity and specificity of both wt-DltA as well as for the DltA mutant C268A (see below). Specificity was checked for the cognate D-alanine as well as several L-amino acids and D-amino acids. Reaction mixtures (100 $\mu$ l final volume) contained 100 pmol of enzyme in assay buffer (50 mM Hepes, pH 7.8, 300 mM NaCl), 1 mM amino acid and 10 mM MgCl<sub>2</sub>. As a negative control the reaction mixtures contained water instead of amino acids substrate. The reaction was initiated by the addition of 2 mM ATP, 0.1 mM PPi and 0.15 $\mu$ Ci tetrasodium [<sup>32</sup>P] pyrophosphate (PerkinElmer) and incubated at 37°C for 10 minutes. Reactions were quenched by adding 0.5 ml of a stop mix containing 1.2 % (w/v) activated charcoal, 0.1 M tetrasodium pyrophosphate and 0.35 M perchloric acid. Subsequently, the charcoal was pelleted by centrifugation, washed 1-2 times with 0.5 ml water and resuspended in 0.5 ml water. After addition of 3.5 ml liquid scintillation fluid

(Rotiscint Eco Plus; Roth), the charcoal-bound radioactivity was determined by liquid scintillation counting (LSC) using a 1900CA Tri-Carb liquid scintillation analyzer (Packard).

#### **4.7 Isothermal titration calorimetry for studying interaction between DltA with DltC**

Isothermal titration calorimetry (ITC) was used to analyze the interaction between DltA and DltC in the presence of inhibitor. ITC can be used to study both protein-ligand and protein-protein interactions. It can directly determine the binding affinity ( $K_a$ ), enthalpy changes ( $\Delta H$ ), and binding stoichiometry ( $n$ ) of the interaction between two or more molecules in solution (Ladbury and Chowdhry, 1996). From these initial measurements, Gibbs energy changes ( $\Delta G$ ) and entropy changes ( $\Delta S$ ) can be determined using the relationship:

$$-RT \ln K_a = \Delta G^\circ = \Delta H^\circ - T\Delta S$$

where R is the gas constant and T is the absolute experimental temperature. From experiments involving titration of the reactants, one may also directly determine the stoichiometry of the interaction from the molar ratio at the equivalence point.

Both proteins DltA and DltC were dialysed against buffer containing 50 mM Hepes, 100 mM NaCl, pH 7.8. The inhibitor i.e. 5'-O-[N-(D-Alanyl)-sulfamoyl]-adenosine which was available as a salt was also dissolved in the same buffer so as to minimize any artefacts arising from mismatched buffer components. All dilutions were also performed with the same buffer. In the first set of experiments DltA (macromolecule) was titrated against the inhibitor (titrant). Both DltA and inhibitor solutions were degassed prior to loading into the sample cell and injection syringe, respectively. About 2 ml of DltA protein at a concentration of 1  $\mu$ M was used to fill the sample cell, taking care not to introduce any air bubbles that may interfere with the feedback circuit. The injection syringe was filled with 350  $\mu$ l of 10  $\mu$ M inhibitor solution and gently loaded into place. Titration was then performed at 25°C. The data were analyzed with the analysis software ORIGIN (Microcal, Northhampton, MA) provided with the VP-ITC. The second set of experiments was performed so as to study the interaction between DltA and its conjugate protein partner DltC in the presence of the inhibitor. First 10  $\mu$ M of DltA protein solution was mixed with 20  $\mu$ M of inhibitor solution and 2 ml of this mixture was used to fill the sample cell. Then DltC solution at a concentration of 100  $\mu$ M was filled into the injection syringe and titration was performed as described above.

#### **4.8 Native gel shift assay**

The interaction of DltA with DltC was also investigated by native polyacrylamide gel electrophoresis. DltA (0.268 nmols) and DltC were mixed in the following molar ratios; 1:1.2, 1:1.6, 1:2 respectively. One set of reactions was performed in absence of any substrates, making up the final volume of 20  $\mu$ l with Hepes buffer A, while another set of reactions was performed in the presence of substrates i.e. 2 mM D-alanine, 2 mM MgCl<sub>2</sub>, 5 mM ATP, incubated at 37°C for 10 minutes. The different samples were loaded on a 10 % acrylamide/bisacrylamide native gel and electrophoresis was performed. Gels were stained in Coomassie Brilliant Blue solution, destained and analyzed.

#### **4.9 Protein crystallization of DltA and DltC**

Protein crystallization occurs in supersaturated solutions where the concentration of protein exceeds its equilibrium solubility value (Weber, 1997). As protein solubility is influenced by a wide range of parameters (including protein concentration but also additives (alcohols), hydrophilic polymers, detergents, salt, pH and temperature), macromolecular crystallisation requires the empirical screening of a wide range of experimental conditions.

Native DltA protein at two concentrations of 10 and 20 mg/ml was incubated in the presence of its substrates 2mM D-alanine, 5mM ATP and 2mM MgCl<sub>2</sub> at 37 °C for 10 minutes to perform adenylation prior to crystallization. Both sitting drop as well as hanging drop vapour diffusion methods were used where equal volumes protein (2  $\mu$ l) and buffer solution (2  $\mu$ l) from the reservoir (500  $\mu$ l) were mixed in each drop. The protein was screened against a house factorial screen based on the matrix of Jancarik and Kim (Jancarik and Kim, 1991) consisting of 96 different conditions varying in salt composition and concentration, pH, precipitant concentration, and additives. The plates were sealed and further incubated at 25°C.

For the selenomethionine labelled DltA protein, further screening was done with identical conditions to native protein. As mostly small poorly diffracting crystals were obtained, both macro and microseeding were also performed. Hampton additive screens (Hampton Research), which are known to improve the quality of protein crystals, were also tested. With yet another batch of purified Se-met DltA protein, Hampton research crystal screens (HR2-110, HR2-112) were additionally tested.

Native DltA protein was also set up for crystallization in the presence of the nonhydrolysable ATP analogue adenosine-5'-[( $\alpha,\beta$ )-methylene] triphosphate, against the above mentioned screens. Setups were done both in the absence and presence of substrate i.e. D-alanine. To improve crystals, fine screening was done by varying protein concentration, pH, temperature and precipitant concentration.

DltC was also set up for crystallization against the house factorial screen, Hampton research crystal screen and various fine screens. Protein concentrations used were 10 mg/ml and 20 mg/ml. Both hanging drop and sitting drop vapour diffusion methods were used. Pipetting was performed manually as well as use of a pipetting robot (Zinsser Analytics, Genomic Solutions)

Attempts to co-crystallize DltA with DltC were also performed. For this DltA was incubated with DltC in a 1:1.2 ratio and crystallization was performed in the presence of substrates, DAla,  $Mg^{2+}$  and ATP. Later on, co-crystallization of DltA and DltC in the presence of the inhibitor, 5'-O-[N-(D-Alanyl)-sulfamoyl]-adenosine was also performed. Screening was done against house factorial, Hampton research and Jena bioscience screens as well as a self prepared screen based on the matrix developed by Radaev *et al* (Radaev and Sun, 2002) for crystallization of protein-protein complexes.

#### **4.10 Measurements of protein crystals and data collection.**

DltA crystals were mainly measured at the synchrotron, while DltC crystals were measured both at the synchrotron as well as on an in-house rotating anode x-ray generator (copper  $K\alpha$  wavelength 1.54178 Å). Native data sets for DltA crystals were collected at both the Deutches Elektronen-Synchrotron (DESY), Hamburg, beam line BW6 and at the PSF beam line at BESSY, Berlin. For measurements, the native DltA crystals were initially soaked in cryobuffer supplemented with 20 % glycerol as cryoprotectant. The crystals were mounted in loops and immediately flash cooled at 170 K in a nitrogen gas cryostream. After successfully collecting data for the native DltA crystal that belonged to the I222 space group, a SAD data set was collected for selenomethionine labelled DltA protein crystals at DESY, Hamburg, beam line BW6 ( $\lambda=0.97888\text{\AA}$ ). A second native data set (which later turned out as a crystal of  $P2_12_12_1$  DltA) was also collected for a crystal grown in the presence of DltA and DltC.

The DltC crystals were measured both in-house on a rotating anode (copper) x-ray generator (Rigaku) and at the synchrotron (BESSY). For one crystal form measured in house, 20% glycerol was sufficient as a cryoprotectant. The crystal was mounted in loops of appropriate size, immediately transferred to the cryostream (120 K) and centered on the goniometerhead. The generator was operated at a voltage of 20 kV and a tube current of 10 mA. The Cu-K $\alpha$  radiation ( $\lambda = 1.54178 \text{ \AA}$ ) could be selected with the help of an osmic confocal optic monochromator. The distance between the crystal and detector was kept to 100 mm in all measurements. The detector used was R-AXIS-IV<sup>++</sup> image plates. The software used for data collection and processing was CrystalClear (Rigaku). Another crystal form of DltC was also measured at PSF beam line at BESSY, Berlin. Heavy metal soaks were also performed with 0.1 M KI as well as Hg<sub>2</sub>Cl<sub>2</sub> for 10, 30 and 120 minutes and data collected in house.

#### 4.11 Structure solution

A typical X-ray diffraction experiment only measures the intensity of the reflections, which is proportional to the amplitudes of the diffracted waves, but the phase information is lost and cannot be measured directly. Once the phases have been determined, the electron density in a crystal can be calculated by the fundamental equation:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \cos 2\pi(hx - ky - lz - \alpha_{hkl})$$

whereby the electron density,  $\rho$ , at position  $(x, y, z)$  in a unit cell of volume  $V$ , is expressed as a Fourier transformation of the structure factors  $F(h k l)$ . The amplitudes of these structure factors can be obtained from the intensity of the diffracted beam by the relation  $I(h k l) = |F(h k l)|^2$ , but the phase angles  $\alpha(h k l)$  cannot be derived directly and a number of methods can be used to obtain the phases. Molecular replacement can be used if a good model (sharing at least 40%-50% sequence identity) is available for the unknown structure in the crystal. In this method, the orientation (rotation function) as well as the position (translation function) are determined for a molecule in the unit cell using a previously solved structure. On the other hand, Multiple Isomorphous Replacement (MIR) involves introducing heavy atoms into the crystal, usually by soaking experiments. The structure factors of heavy-atom derivative ( $\mathbf{Fp}$ ) and native protein ( $\mathbf{Fp}$ ) and heavy atom ( $\mathbf{Fh}$ ) have the following relationship;  $\mathbf{Fp} = \mathbf{Fp} - \mathbf{Fh}$ . It is possible to determine the position of the heavy atom in the unit cell using the Patterson difference map, thus giving both the amplitude as well as phase for the heavy atom ( $\mathbf{Fh}$ ). With the amplitudes of  $\mathbf{Fp}$  and  $\mathbf{Fp}$  measured experimentally, and  $\mathbf{Fh}$  determined, it is possible to calculate the phases for both  $\mathbf{Fp}$  and  $\mathbf{Fp}$  geometrically.

Multiple wavelength Anomalous Dispersion (MAD) requires introduction of an anomalous scatterer (often selenium, Se) into the protein. Anomalous scattering occurs as the wavelength of the incoming X-rays approaches that of the absorption edge. Selenium, with an absorption  $K$  edge at  $0.9800\text{\AA}$ , has become the most common choice for the anomalous scatterer, as it can be easily introduced into recombinantly produced proteins. The form factor  $f$  of the atom in question becomes modified by the relationship  $f_{anom} = f + f' + if''$ , where  $f'$  represents the real component of the absorption,  $i$  is the imaginary number  $\sqrt{-1}$ , and  $f''$  is the imaginary component of the form factor. Similar to MIR, the anomalous scatterers may be located using an anomalous Patterson map. Initial phasing then proceeds in a manner similar to that used for MIR. MAD experiments can be conveniently performed at synchrotrons where it is possible to tune the wavelengths. The position of the adsorption edge of the atom in question is determined by measuring the X-ray fluorescence of the crystal as a function of X-ray energy (Energy  $E$  can be converted into wavelength  $\lambda$  using the relationship  $E=h\nu=hc/\lambda$ ). The peak of the fluorescence scan corresponds to the maximum absorption ( $f'$  is maximum), whilst the inflection of the scan corresponds to the maximum dispersion ( $f''$  is maximum). Data sets are then collected from the crystal at wavelengths equivalent to the peak and the inflection, as well as a remote wavelength. Assuming that radiation damage is negligible, the 3 or 4 resulting data sets then differ only in the anomalous scattering contribution if collected from a single crystal.

Native data sets for both DltA crystal forms were evaluated using XDS (Kabsch, 1993). All attempts at molecular replacement for both crystal forms of DltA using various known A-domain structures were unsuccessful. Thus a single wavelength anomalous diffraction data was collected at the peak wavelength of  $\lambda=0.97888\text{\AA}$  to  $3\text{\AA}$  for the Se-methionine labelled protein at DESY. The data were evaluated using the HKL package (Otwinowski and Minor, 1997). All 12 substituted Se-atoms could be found using the SHELX package (Schneider and Sheldrick, 2002), providing an initial interpretable electron density. Model building was performed using the programme O (Jones et al., 1991), and the structures were refined using CNS (Brunger et al., 1998). The  $P2_12_12_1$  crystal form was solved by molecular replacement using the I222 structure as search model, and rebuilt and refined as described above. Several attempts at molecular replacement for DltC using all available models of ACPs, PCPs, as well as the apo-Dcp were performed but no solution could be obtained. Automated phasing were also performed using the Phaser software but with no solution.

#### 4.12 Site directed mutagenesis of DltA

Cys292 located in the active site was mutated to an alanine. For this, quick change site directed mutagenesis method was used (Stratagene). The sequences of the primers were as follows: DltA forward, cys-ala: 5'-CAT GCG GAC ACA TTT ATG TTT GCG GGA GAG GTT CTT CCG G-3' ; DltA reverse, cys-ala 5'-CCG GAA GAA CCT CTC CCG CAA ACA TAA ATG TGT CCG CAT G-3' (Metabion). For this, plasmid DNA i.e. pQE60 containing the DltA gene was isolated by performing a miniprep (Qiagen). The PCR reaction was performed in thin walled tubes recommended by Stratagene ensuring ideal contact with the temperature cyclers heat block. The reaction was set up with 1 µl of the template DNA, 5 µl of 10x reaction buffer, 1 µl each of both forward and reverse primers, 1 µl of dNTP mix. Distilled water was added to make up a final volume of 50 µl and finally 1 µl *PfuTurbo* DNA polymerase (2.5 U/µl) was added. The thermal cycling parameters were as follows;

**Table 6: PCR programme for site directed mutagenesis**

Cycles	Temp(°C)	Time(min)
1	95	2
18	95	0.75
	55	0.75
	68	10

Following temperature cycling, the reaction mixture was placed on ice for 2 minutes to cool it to  $\leq 37^{\circ}\text{C}$ . In order to digest parental DNA template, *Dpn* I digestion was performed. For this, 1 µl of the *Dpn* I restriction enzyme (10 U/µl) was added to the amplification reaction and the mixture was gently and thoroughly mixed. Subsequently the tubes were spun down in a microcentrifuge for 1 minute, and the reaction was immediately incubated at  $37^{\circ}\text{C}$  for 1 hour to digest the parental supercoiled dsDNA. For initial transformation XL1-Blue supercompetent cells were used. For this 5 µl of the *Dpn*-I treated PCR product was transferred to 50 µl of prechilled supercompetent cells, mixed thoroughly and incubated on ice for 30 minutes. The transformation reactants were given a heat pulse for 45 seconds at  $42^{\circ}\text{C}$  and further incubated for 2 minutes on ice. Then 0.5 µl of NZY+ broth preheated to  $42^{\circ}\text{C}$  was added and the reactions were further incubated for  $37^{\circ}\text{C}$  for 45 minutes, shaking it at 250 rpm. Then 250 µl of the transformation reaction was plated on LB-agar plates



containing ampicillin. The plates were further incubated at 37°C overnight. Few colonies were selected from the plates and their plasmid DNA was sent for sequencing (GATC) in order to verify whether or not they contain the desired mutation.

### 4.13 Sequence alignments

A structure based sequence alignment was performed for DltA with known structures of the AMP family i.e. PheA, DhbE, ACS, Luciferase, CBAL, and LC-FACS. Initially the entire N-terminal domain as well as C-terminal domains of known structures were aligned separately using the program LSQMAN. However aligning the entire N-terminal domains revealed high RMSD values in certain regions. This led to alignment of the individual subdomains that form the N-terminal domain ie subdomain A, B and C (Conti et al., 1996). Initially, each of these subdomains were aligned explicitly based on their  $\beta$ -sheets. Then an improve step was done considering the whole N-terminus, in this way one can get an alignment for the whole N-terminal domain based on a single subdomain.

Sequence based alignments were also done taking *B. subtilis* DltA sequence as query using PSI-BLAST at NCBI server. A maximum number of 20000 sequences were targeted. The DAla-specific *bona fide* NRPS starter domain LnmQ (Tang et al., 2007) was also included in the alignment. A further BLAST search was performed taking LnmQ as a query sequence and then aligned with the DltAs. A phylogenetic analysis was also performed with representative members of the adenylation domain family which showed significant sequence similarity in the blast search database. The selected sequences were aligned using clustalW, phylogeny was characterized using BioNJ, and a phylogenetic tree generated with Treedyn.

### 4.14 Modelling of functional states

Modelling was performed by Milton T. Stubbs on the DltA structure to represent the different functional states of NRPS adenylation domain structures. For this, the following pdb entries were used as templates to complete the picture of the catalytic reaction: *B. subtilis* PheA (1amu (Conti et al., 1997): modelling of DAla into the active site and orienting the small domain in the ‘adenylation’ conformation); *B. subtilis* DhbE (1md9 (May et al., 2002): ATP bound conformation); *h-ras* p21 (5p21 (Pai et al., 1990): ATP bound conformation); luciferase from *L. cruciata* (2d1s (Nakatsu et al., 2006): modelling of P-loop in ‘adenylation’ conformation) and from *P. pyralis* (1ba3 (Franks et al., 1998): modelling of hinge region and small domain in the ‘open’ conformation); ACS from *S. enterica* (1pg4 (Gulick et al., 2003):

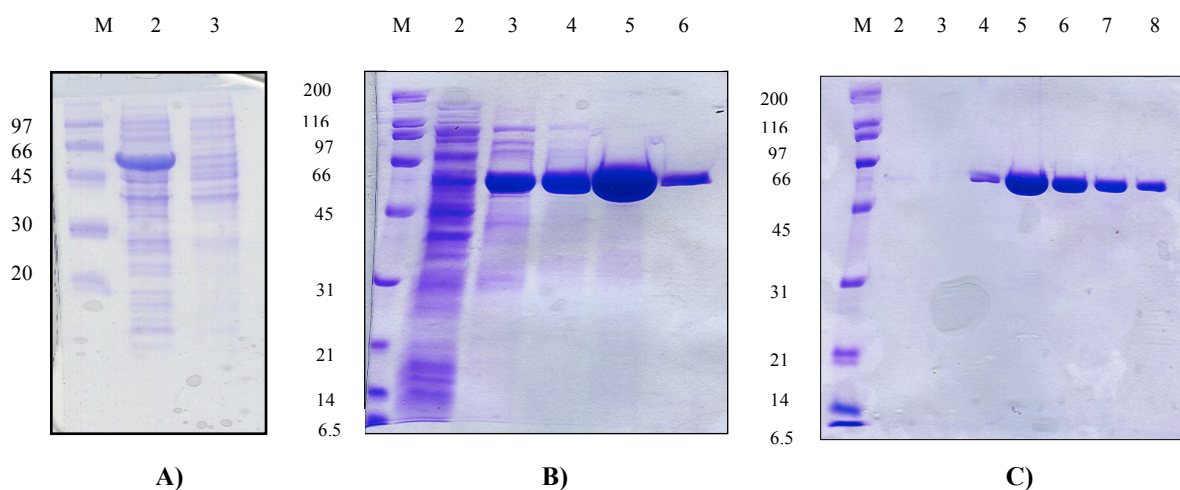
modelling of phosphopantetheinyl moiety in the 'thiolation' conformation). Initial superposition of the large domains to that of DltA in the I222 crystal form using LSQMAN (Kleywegt, 1996) was followed by transformation of the corresponding regions of DltA (small domain, hinge region, P-loop) to obtain the various functional states; where necessary, torsion angles of individual amino acids were adjusted to those of the target structure using O (Jones et al., 1991). No attempts at energy minimisation or molecular dynamics were made for any of the modelled structures.

## 5 Results

### 5.1 Recombinant protein expression, purification and characterization of DltA and DltC

#### 5.1.1 Expression and purification of DltA

Expression of DltA in 1 litre of LB-medium at 37 °C yielded 4 g of biomass, with DltA produced as a soluble protein (Fig. 14A). The protein was purified using an Ni-NTA affinity column (yielding protein of ~90% purity, Fig. 14B) and gel filtration, resulting in a single band on a Coomassie Brilliant blue-stained SDS-PAGE (> 99 % purity, Fig. 14C). The final yield of purified DltA determined at  $A_{280}$  using a molar extinction coefficient,  $49650 \text{ M}^{-1} \cdot \text{cm}^{-1}$ , was about 80 mg/L of cell culture. The C292A DltA mutant (see below) was expressed and purified as wild type DltA with similar results.



**Fig. 14. Coomassie Blue-stained 10% SDS-polyacrylamide gel.**

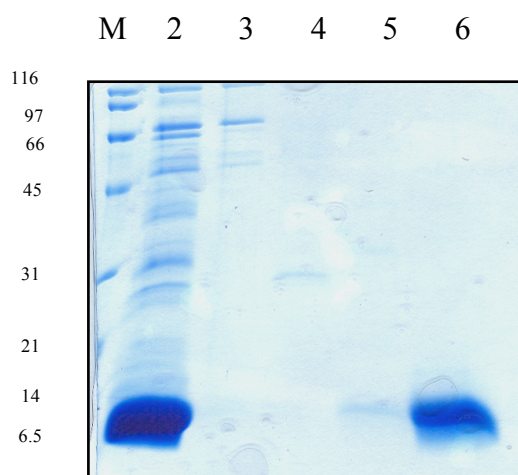
(A) expression of DltA. Lane 2 represents whole cells of induced culture, demonstrating a prominent band at ~57 kDa, while lane 3 shows whole cells of uninduced culture; lane M: molecular mass standard in kDa. B) DltA after Ni-NTA affinity chromatography. Some unbound protein can be seen in the flow through (lane 2) however most protein bound to the column via its C-terminal hexa-histidine tag and could be easily separated from the contaminating impurities with a linear gradient of up to 250 mM imidazole. Lane 3 of the gel shows some loosely bound protein that comes off shortly before the main peak of the protein, so that this fraction was discarded. Lanes 4-6 show fractions from the main peak eluting at imidazole concentrations from 50 – 187.5 mM; these fractions were pooled for further purification. C) DltA after gel filtration chromatography. Samples after Ni-NTA step were concentrated to a volume of 3 ml and applied onto a superdex G-75 gel filtration column (Amersham Biosciences). DltA eluted as a symmetrical peak at a column volume about 120 ml.

The selenomethionine labelled DltA protein expressed in M9 medium yielded about 2 g of biomass after overnight induction and had a lower yield than that of the native protein ie 20 mg per liter cell culture. This was expected as cells tend to grow much slower in minimal medium and reach stationary phase at a lower final cell density (Doubie, 1997).

Purification of selenomethionine DltA was identical to that of the native protein and a single band could also be seen on 10% SDS gels after gel filtration.

### 5.1.2 Expression and purification of DltC

DltC, also expressed at 37 °C, yielded approximately 4 g of wet biomass, and was produced solubly. Purification by Ni-NTA as well as gel filtration chromatography yielded protein of sufficient purity (> 99 % purity, Fig. 15). Using gel filtration chromatography, a single band of pure DltC could be obtained (lane 6). An earlier eluting fraction (lane 5) might represent a putative holo-DltC dimer as reported in earlier studies (Debabov et al., 1996). The final yield of the purified protein determined at  $A_{280}$  using molar extinction coefficient  $5810 \text{ M}^{-1} \cdot \text{cm}^{-1}$  was approximately 30 mg of DltC from a 1 litre cell culture.

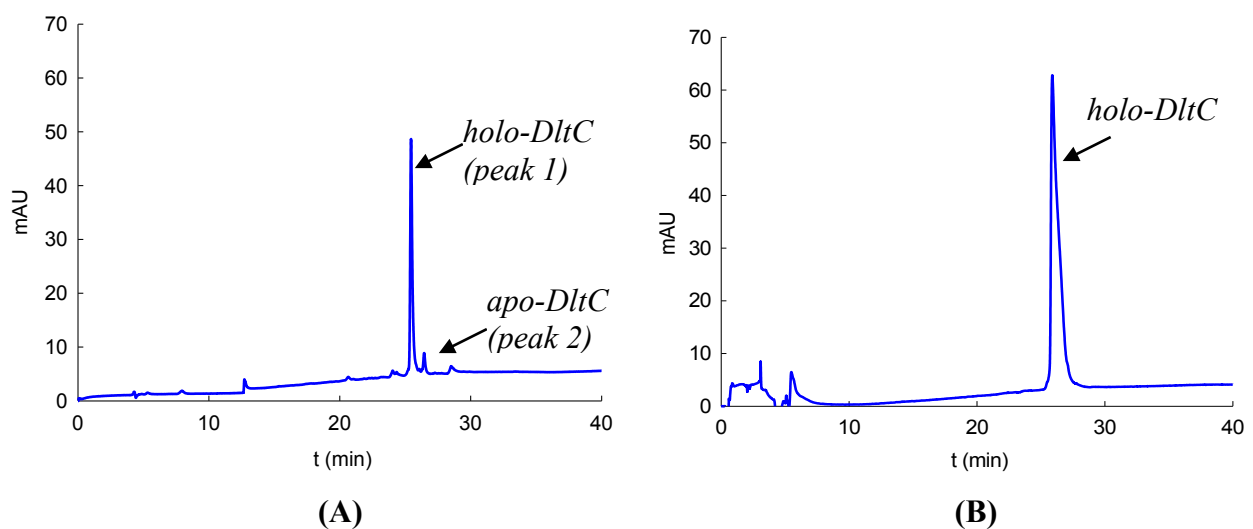


**Fig. 15.** Coomassie Blue-stained 15% SDS-PAGE gel after Ni-NTA affinity chromatography (lane 2) and after gel filtration (lanes 3-6). Lane M shows the molecular mass standard.

### 5.1.3 Estimation of the degree of 4'-phosphopantetheine-containing DltC (holo-DltC) following *in vivo* post-translational modification

Reversed phase HPLC was performed to assess the amount of carrier protein bound to the 4'-phosphopantetheine (4'PP) cofactor after expression and *in vivo* post-translation modification (by the ACP synthase of the host strain (Debabov et al., 1996)) (Fig. 16), followed by mass spectrometry. DltC elutes at a retention time of approximately 25 minutes. In one batch (Fig. 16A), two peaks are observed; DltC is predominantly in the holo-form (measured mass 10415.4 Da, theoretical value 10415.5 Da) with a small fraction of apo-DltC (measured mass 10075.9 Da, theoretical value 10075.2 Da). In a second batch (Fig. 16B), the entire protein

was found in the holo-form. This predominance of the holo-form was observed in most batches of expression; in one batch only, both the holo and apo-forms were present in almost equal amounts (data not shown). This was unexpected, as previous studies had revealed an overall low amount of holo-DltC: 2-3% after *in vivo* expression of DltC from *L. casei*, where the majority of the protein had not been processed by holo-ACP synthase of the expression strain (Debabov et al., 1996); and 51 % of holo-DltC from *B. subtilis* (May et al., 2005). This difference in the amount of holo-DltC from *B. subtilis* may be due to the different *E. coli* strains used for expression, although batch variability could also play a role.

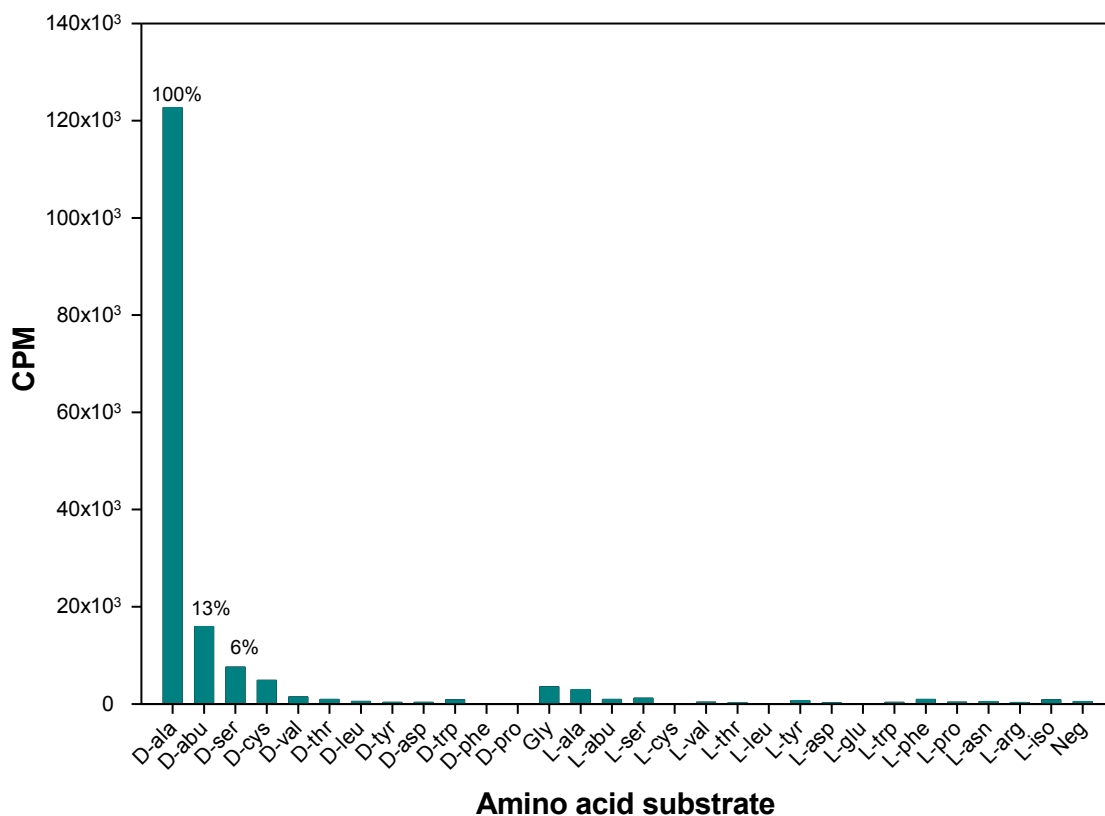


**Fig. 16.** HPLC analysis for estimating the amount of holo-DltC formed after *in vivo* post-translational modification from two separate expression batches. A. Chromatogram showing separation of holo-form (measured mass 10415.4 Da) from apo-form (measured mass 10075.9 Da) on a C4 HPLC column. B. Chromatogram showing only a single peak of DltC expressed only as holo-protein

## 5.2 Substrate specificity of DltA

### 5.2.1 Activity assay for DltA

Purified DltA was checked for activity as well as substrate selectivity with the help of ATP-PPi exchange assay (section 4.6). The specificity was tested in presence of the natural substrate, D-alanine as well as for almost all proteinogenic amino acids and several D-amino acids (Fig 17). The negative control (Neg) represents the activity detected in the absence of any amino acids. DltA showed highest activity for its cognate substrate, D-alanine, with side-specificity for the non-proteinogenic amino acids, D-amino butyric acid (DAbu) (13%). The enzyme also showed some minor activity for DSer (6%) with virtually no activity for LAla as well as the much smaller glycine. DltA also showed no activation of the L-enantiomers of LAbu and LSer, nor for amino acids with larger side chains, hydrophobic as well as charged amino acids.

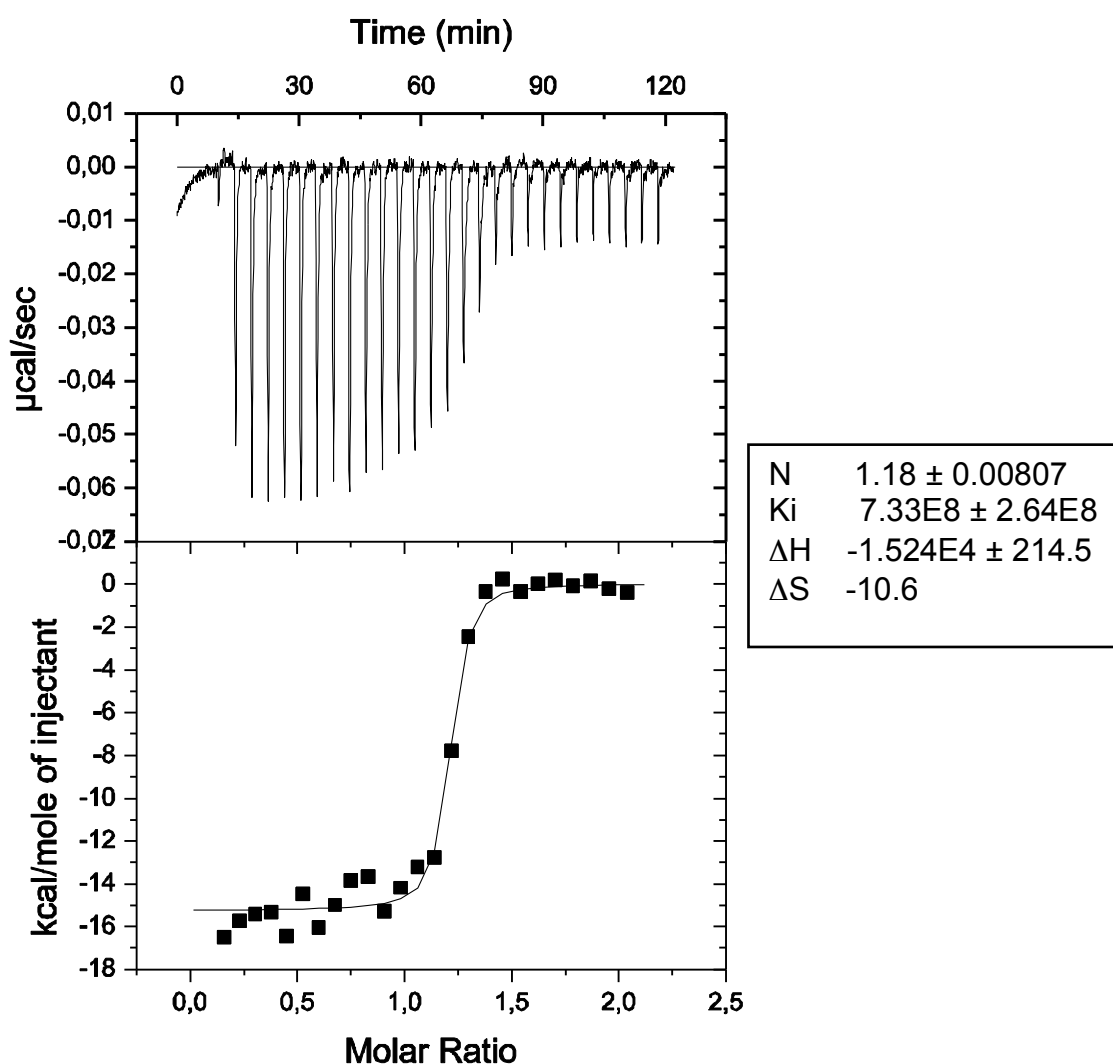


**Fig. 17.** Amino acid specificity of wild type DltA determined by the ATP-PPi exchange assay. The highest exchange activity for the cognate DAla substrate was defined as 100% and the values obtained for other amino acids were set accordingly.

### 5.3 Studies on the formation of a DltA-DltC complex

#### 5.3.1 Isothermal titration calorimetry to study complex formation of DltA with DltC

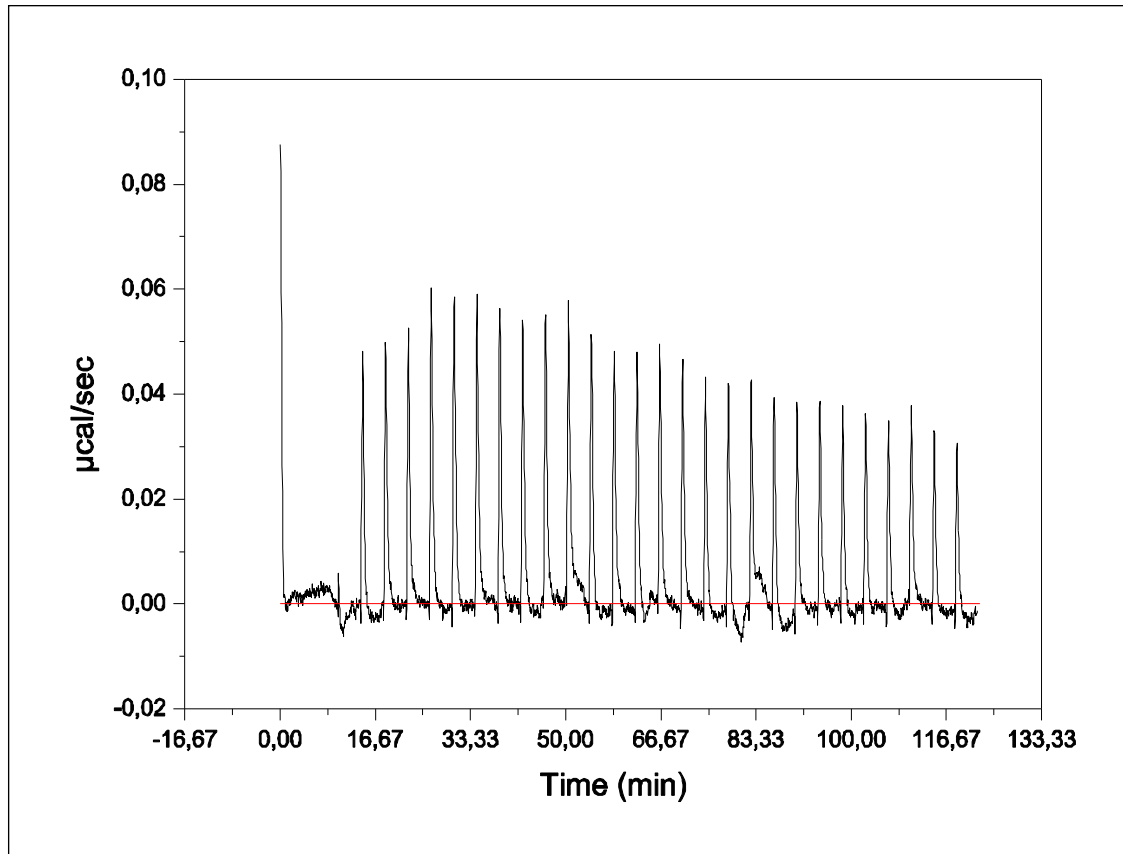
The association of DltA with its binding partner DltC was investigated by ITC. Initial experiments tested the binding of DltA with its known inhibitor, 5'-O-[N-(D-Alanyl)sulfamoyl]-adenosine (May et al., 2005). Since typical ITC experiments are performed with initial ligand concentrations which are 10-20 times greater than the macromolecular concentration, 10  $\mu\text{M}$  of inhibitor was titrated against 1  $\mu\text{M}$  DltA solution. As can be seen from the Fig. 18, the inhibitor binds with very high affinities to DltA resulting in a very steep curve. The  $n$ ,  $K_a$  and  $\Delta H$  parameters were determined as 1.2,  $0.73 \cdot 10^9 \text{ M}^{-1}$  and  $-15.2 \text{ kcal mol}^{-1}$  respectively using a 1:1 binding model.



**Fig. 18. Differential binding isotherm for the binding of 5'-O-[N-(D-Alanyl)sulfamoyl]-adenosine to the DltA.**

Top: Raw ITC data. Bottom: Baseline-adjusted binding isotherm, created by plotting the integrated heat flow per mole of added ligand against the molar ratio of ligand to macromolecule present in the cell

The  $K_d$  for the reaction was estimated at 1.4 nM, which is in the range observed in previous studies (May et al., 2005). The stoichiometry, determined from the molar ratio of the inhibitor to protein at equivalence point was 1.2. This error is probably due to an imprecise determination of protein concentrations. To study complex formation between DltA and holo-DltC, the inhibitor was used as a substrate instead of the natural substrates, D-alanine and ATP in order to prevent substrate turn over. DltA at 10  $\mu$ M was mixed with 20  $\mu$ M of inhibitor and titration was performed with 100  $\mu$ M of holo-DltC solution, however as seen from Fig. 19, no binding could be observed even at such high concentration of reactants.

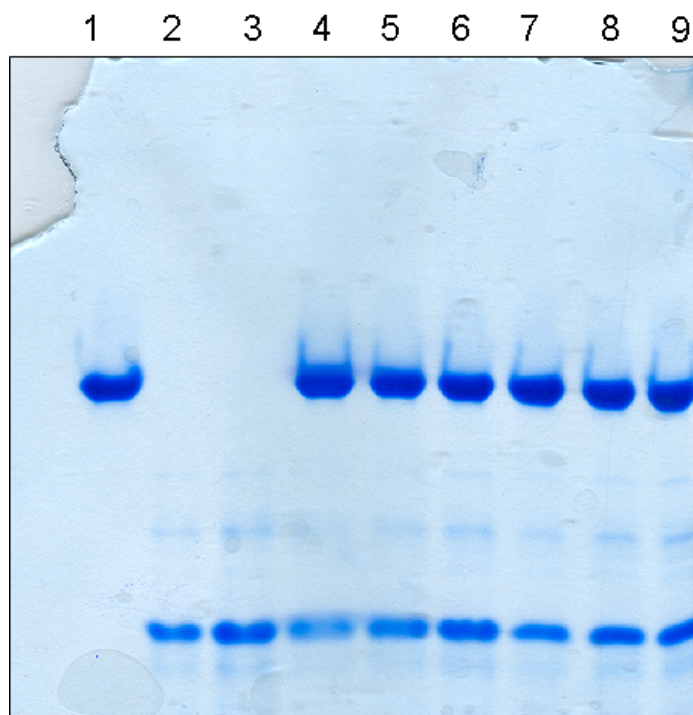


**Fig. 19.** Raw ITC data for the binding reaction of DltC to the DltA-inhibitor complex at 25 °C.

### **5.3.2 Native gel shift assay to test the presence of a complex between DltA and DltC**

The affinity of DltA to DltC was tested by a native gel shift assay which should preserve any non-covalent protein-protein interactions. As seen from Fig. 20, no other band other than those seen with uncomplexed forms of DltA (lane 1) and DltC (lanes 2 and 3) could be observed in lanes 4-6 which contains a mixture of DltA and holo-DltC in the absence of substrates as well as lanes 7-9 which additional contain the substrates. Also DltC shows some additional faint bands other than the main DltC band.





**Fig. 20. Native gel shift assay to analyze the DltA-DltC complex formation.**

Lane 1 contains only DltA, while lanes 2 and 3 contain only DltC at increasing concentrations. Lanes 4, 5 and 6 contain DltA and DltC mixed in the following ratios 1:1.2, 1:1.6 and 1:2 respectively while lanes 7, 8 and 9 contain DltA and DltC mixed in the same ratio as lanes 4, 5 and 6 but additionally contain the substrate i.e. 2 mM D-alanine, 2 mM MgCl<sub>2</sub>, 5 mM ATP.

## 5.4 Crystallization of DltA and DltC

### 5.4.1 Crystallization of DltA

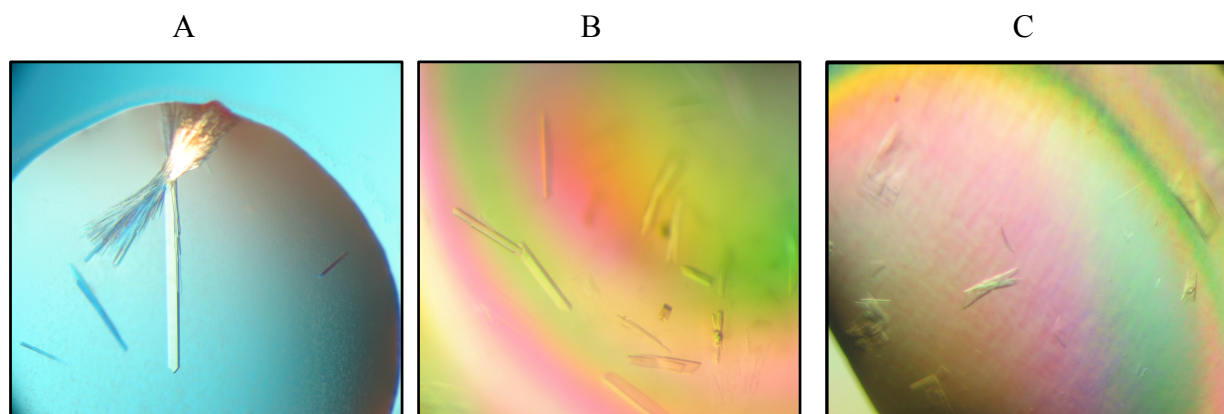
Native DltA protein was set up for crystallization as described (section 4.9). The protein crystallized in various conditions however most conditions yielded only very thin needles that diffracted very weakly. One crystal, obtained at a concentration of 20 mg/ml in the presence of substrates, D-alanine, MgCl<sub>2</sub> and ATP with a fine screen around one of the 96 conditions of the house factorial screen (House factorial No. 53) i.e. 0.1 M sodium citrate, 0.2 M ammonium acetate, 20 % PEG 4000 at pH 6.0, grew out of the needles over a period of 2 months using the sitting drop vapour diffusion method (Fig. 21A). The crystals, which were measured at MPG beamline BW6 at DESY, Hamburg, diffracted to about 2.28Å resolution. They belonged to the orthorhombic space group I222, with cell constants  $a = 90.56$ ,  $b = 94.99$ , and  $c = 132.55$ Å and contained 1 molecule per asymmetric unit. Details of crystallographic statistics are given in table 7.

DltA protein was also set up for crystallization in the presence of a nonhydrolysable ATP analogue, Adenosine-5'-[( $\alpha,\beta$ )-methylene] triphosphate, both in the absence and presence of

D-alanine substrate. This analogue was used to capture the 'preadenylation' state of the reaction i.e. when the substrates, ATP and DAla bind and before the adenylation reaction can proceed. Crystals under two conditions ie 0.1 M sodium citrate, 0.1 M ammonium sulphate, 30% PEG 8000, pH 6.1 and 0.2M ammonium sulphate, 0.1 M Na-cacodylate, 30%PEG8000, pH 6.5 could be obtained, however they were only thin needles that did not diffract. Attempts to improve these crystals by varying concentration of precipitants, pH as well as streak seeding failed to improve them.

As all attempts to solve the structure by molecular replacement failed (a phenomenon that has been observed for many adenylation domains), selenomethionine labelled DltA protein was crystallized. This material yielded only very thin needles when grown under conditions for the native protein, and both macro and microseeding resulted in thin plates that diffracted poorly. Similar results were obtained using the Hampton additive screen. Suitable crystals of selenomethionine labelled DltA in the presence of substrates could finally be obtained from a solution containing 0.1M sodium cacodylate, 0.2M ammonium sulphate and 30% PEG 8000 at pH 6.5 from the Hampton research crystal screen (Fig. 21B). An SAD dataset was collected at beamline BW6, DESY, Hamburg. The crystals diffracted to about 3.0 Å. and were isomorphous to the native DltA crystals.

Co-crystallization experiments of DltA with DltC were also carried out in the presence and absence of the natural substrates i.e. DAla, Mg<sup>2+</sup> and ATP, as well as the inhibitor 5'-O-[N-(D-Alanyl)-sulfamoyl]-adenosine. This non-cleavable substrate analogue can be used to mimic the intermediate amino-acid-adenylate, thereby stabilizing the conformation for the second half-reaction during which the holo-form of DltC should bind. One crystal could be obtained in the presence of DltA mixed with holo-DltC in a 1:1.2 ratio as well as substrates (Fig. 21C) from house factorial screen No. 86 (30 % PEG 1000 at pH 6.5). They belonged to the orthorhombic space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with cell constants a = 45.8, b = 91.4, c = 122.4 Å. They diffracted to 2.6 Å and again contained 1 molecule per asymmetric unit. These crystals however revealed neither density nor space for the carrier protein (DltC), corroborated by the specific volume of the crystal. This was thus taken as another crystal form of DltA i.e. P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> (table 7).



**Fig. 21. Crystals of DltA.**

(A) Native DltA crystals belonging to the orthorhombic space group,  $I222$ . The buffer conditions were 0.1 M sodium citrate, 0.2M ammonium acetate, 20 % PEG 4000, pH 6.0. (B) Selenomethionine labelled DltA crystals obtained under buffer conditions, 0.1 M sodium cacodylate, 0.2 M ammonium sulphate and 30% PEG 8000 at pH 6.5. These crystals were isomorphous to native DltA crystals. (C) Native DltA crystals belonging to the orthorhombic space group,  $P2_12_12_1$ . The buffer conditions were 30 % PEG 1000 at pH 6.5.

**Table 7: Crystallographic statistics of DltA**

<b>Data collection</b>	<i>SeMet peak, I222</i>	<i>I222</i>	<i>P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub></i>
Wavelength $\lambda$ (Å)	0.97888	1.05000	1.05000
a (Å)	89.58	90.56	45.77
b (Å)	94.42	94.99	91.39
c (Å)	133.71	132.55	121.47
Max. resolution (Å)	3.04	2.28	2.60
$R_{merge}$	8.5 (28.5) <sup>b</sup>	6.1 (27.2)	9.3 (23.4)
% complete	98.0 (92.0)	94.6 (84.7)	96.9 (97.3)
No. of reflections	65255	71086	44316
No. of unique reflections	11079	25284	15971
$\langle I/\sigma(I) \rangle$	8.75 (2.46)	14.16 (4.21)	10.13 (4.79)
<b>Refinement</b>			
Resolution range (Å)		20.00–2.28 (2.31–2.28)	20.00–2.60 (2.60–2.64)
Completeness (working +test) (%)		96.2 (58.6)	97.8 (98.3)
No. of reflections ( $F > 0$ )		25261 (36)	15932 (53)
Wilson B (Å <sup>2</sup> )		23.5	24.5
$R_{cryst}$ (%)		20.50 (30.60)	17.90 (25.50)
$R_{free}$ (%)		25.60 (32.20)	23.80 (32.80)
No. of non-hydrogen atoms			
Protein		4059	4037
Water		367	211
Ligand atoms		28	23
R.m.s.d. from ideality			
Bond lengths (Å)		0.006	0.006
Bond angles (°)		1.30	1.30
Dihedrals (°)		23.4	23.5
Improper (°)		0.92	0.84
<b>Average B-factor (Å<sup>2</sup>)</b>		41.1	39.4
Protein atoms		40.83	39.71
Main chain		40.28	38.07
Water		45.09	36.18
Ligand		27.31	16.26
PO <sub>4</sub>		56.52	-

$$R_{merge} = \frac{\sum |I - \langle I \rangle|}{\sum \langle I \rangle}$$

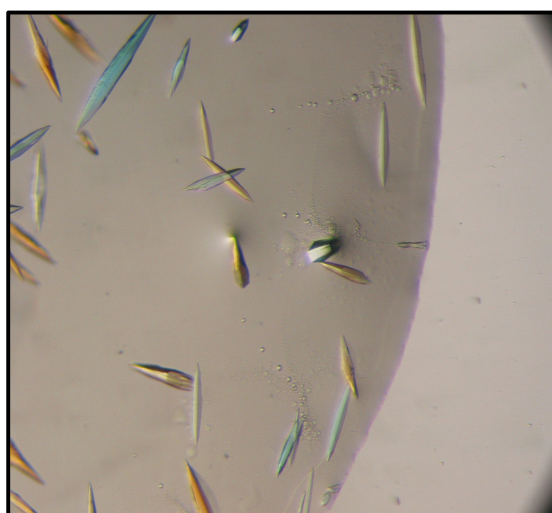
Values in parentheses correspond to the highest resolution shell.

$$R_{cryst} = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

$R_{free}$  is calculated as for  $R_{cryst}$  but for a test set comprising reflections not used in the refinement (5%).

### 5.4.2 Crystallization of DltC

Holo-DltC crystals could be grown from a solution containing 0.1M sodium citrate, 20% Isopropanol, 10% PEG 4000, PH 7.0 using the house factorial screen with some fine screening (Fig. 22). Although these crystals deteriorated rapidly upon exposure to air, a native data set could be collected with an in house x-ray generator (Rigaku) for one crystal that diffracted to 2.9 Å. They belonged to the hexagonal space group P6<sub>5</sub>22 with cell constants a = 72.956, b = 72.956, c = 110.356 and contained one molecule in the asymmetric unit. However attempts to solve the structure by molecular replacement with all available NMR (Crump et al., 1997; Holak et al., 1988; Koglin et al., 2006; Volkman et al., 2001; Weber et al., 2000) as well as crystallographic models (Roujeinikova et al., 2002) failed. More crystallization setups under identical conditions as well as some fine screening around these conditions were performed in order to carry out multiple isomorphous replacement (MIR) with heavy metals. However no crystals could be obtained under these conditions, so entirely new screening was done with another protein batch. This new screening yielded crystals that grew from 0.1 M sodium acetate, 0.2 M ammonium acetate, 30 % PEG 4000, pH 5.7. The best crystals diffracted to 2.2 Å resolution and data were collected at the PSF beam line at BESSY (table 8). These crystals were isomorphous to the ones measured in-house. Heavy metal soaks were performed with potassium iodide as well as mercury chloride at different time intervals, however heavy metal binding sites could not be detected.



**Fig. 22.** Crystals of DltC grown by hanging drop method. They crystallized under conditions 0.1M sodium citrate, 20% Isopropanol, 10% PEG 4000, PH 7.0.

**Table 8: Crystallographic statistics for holo-DltC**

<b>Data collection</b>	<b><i>Holo-DltC, P6<sub>1</sub>22</i></b>
Beamline	BESSY 14.1
Wavelength $\lambda$ (Å)	0.9184
Temperature (K)	100
Molecules/asu	2
Unit cell constants	
a, b, c (Å)	72.68 72.68 110.3
$\alpha, \beta, \gamma$ (Å)	90 90 120
Resolution range (Å)	62.94-2.2 (2.32-2.2)
No. of unique reflections	9293(1300)
Multiplicity	11.2(11.5)
Completeness(%)	99.9 (100)
I/ $\sigma$ I	16.6(3.8)
R <sub>meas</sub>	9.6(80.6)
<b>Refinement</b>	
R <sub>cryst</sub> (%)	20.9
R <sub>free</sub> (%)	23.6
R.m.s.d bond length (Å)	0.002
R.m.s.d bond angle (Å)	0.534
Mean B (Å <sup>2</sup> )	38.76
Ramachandran plot (%)	
Preferred	96.1
Allowed	3.9
Disallowed	0

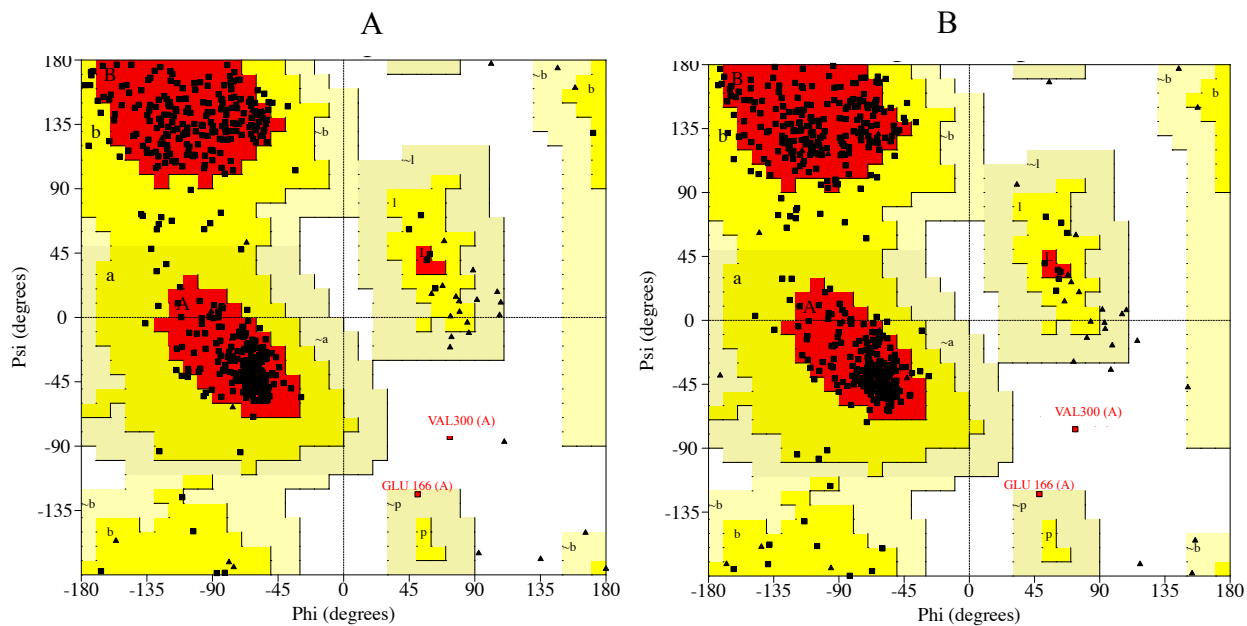
Values in parentheses correspond to the highest resolution shell.  $R_{meas} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$   
 $R_{cryst} = \sum ||F_{obs}| - |F_{calc}|| / \sum |F_{ob}$

## 5.5 Structural studies on DltA and its comparison with members of AMP family.

### 5.5.1 General features of the DltA structure.

DltA protein crystallized in two different orthorhombic space groups, each with one molecule per asymmetric unit. The final model contains 508 residues in both crystal forms; the crystallized protein also contains a C-terminal RSHHHHHH tag used for purification, of which the last three histidine residues are disordered in both crystal forms. The DltA molecules in both crystal forms exhibit very similar conformations, with an rmsd for main chain atoms of 0.62 Å. There are two disordered regions that exhibit high temperature factors in both crystal forms. One of these loops i.e. residues 108-115, is by far one of the most variable regions among all members of this family; another region is loop 447-454, which is also quite different in all members. Additionally, the P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> crystal form exhibits two more poorly ordered regions; loops 334-337 and 484-492. The P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> crystal form contains several disordered side chains which mainly lie on the surface of the molecule.

Most of the residues (99.6%) are located in the most favoured and additional allowed regions of the Ramachandran plot (Ramakrishnan and Ramachandran, 1965) (Fig. 23). Only one residue, V300, is located just outside the allowed regions of the Ramachandran plot. This residue has dihedral angles ( $\phi = 72^\circ$ ,  $\psi = -83^\circ$ ). Interestingly, the residue equivalent to V300 in the PheA (Conti et al., 1997) is I330 which also has unfavourable dihedral angles ( $\phi = 74^\circ$ ,  $\psi = -64^\circ$ ). Such energetically unfavourable conformations are commonly associated with a functional role (Herzberg and Moulton, 1991). In PheA, the main chain carbonyl oxygen of this residue forms a hydrogen bond with the  $\alpha$ -amino group of the phenylalanine substrate. This residue is also part of the specificity conferring code, and is conserved or has an equivalent in all amino acid activating domains, while in DhbE (May et al., 2002), the equivalent of this residue i.e. V337 is not at the position seen in PheA and DltA, due to an abrupt change in the main chain conformation around the region of core A5 (*328QVFGMAEGLVN338*). Structure based alignments also reveal that this residue is conserved only among members which have a substrate bearing an  $\alpha$ -amino group i.e. PheA and DltA. This V300 is also completely conserved among all D-alanine activating enzymes as seen from sequence alignments. It is likely that in the DltA structure, V300 would be involved in a hydrogen bond with the  $\alpha$ -amino group of the D-alanine substrate.



**Fig. 23. Ramachandran plot for DltA. A) I222 crystal form B) P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> crystal form**

**I222 crystal form**

Residues in most favoured regions [A,B,L]	391	88.7%		
Residues in additional allowed regions [a,b,l,p]			48	10.9%
Residues in generously allowed regions [~a, ~b, ~l, ~p]			1	0.2%
Residues in disallowed regions			1	0.2%
			----	-----
Number of non-glycine and non-proline residues			441	100.0%
Number of end-residues (excl. Gly and Pro)			381	
Number of glycine residues (shown as triangles)			37	
Number of proline residues			28	
			----	-----
Total number of residues			887	

**P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> crystal form**

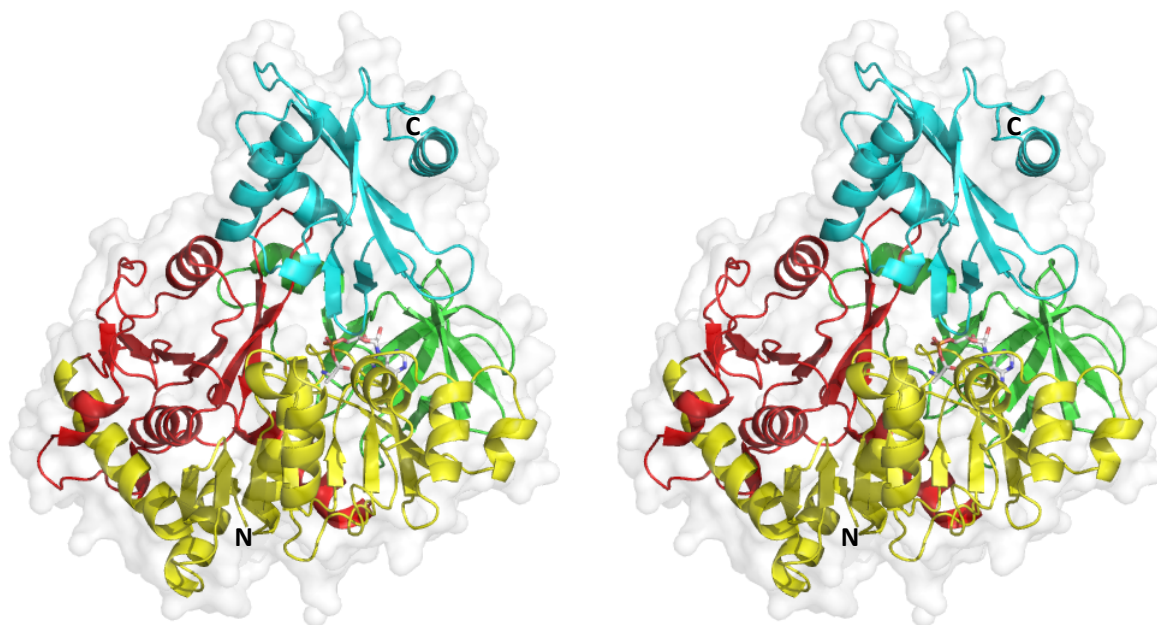
Residues in most favoured regions [A,B,L]	383	86.8%		
Residues in additional allowed regions [a,b,l,p]			56	12.7%
Residues in generously allowed regions [~a, ~b, ~l, ~p]			1	0.2%
Residues in disallowed regions			1	0.2%
			----	-----
Number of non-glycine and non-proline residues			441	100.0%
Number of end-residues (excl. Gly and Pro)			221	
Number of glycine residues (shown as triangles)			37	
Number of proline residues			28	
			----	-----
Total number of residues			727	

**5.5.2 Overall topology of the structure**

The overall structure of DltA is similar to the structures of previously determined members of the adenylate forming family, consisting of two distinct domains, a large N-terminal domain (residues 1-395) and a small C-terminal domain (res 400-508) with a four residue ‘hinge’ in between. Like other members, the N-terminal is further divided into three subdomains. Subdomain A and B share a similar topology consisting of two  $\beta$ -sheets flanked on either side by  $\alpha$ -helices, while subdomain C is a distorted  $\beta$ -barrel (Fig. 24).



Subdomain A is formed by a contiguous segment of the polypeptide chain (residues 47-178), and comprises of 6  $\beta$ -strands and 4  $\alpha$ -helices, while subdomain B is constructed from two non-contiguous segments, the first is a small segment (residues 1-46) containing 2 antiparallel  $\beta$ -strands and 2  $\alpha$ -helices starting at the N-terminus and ending shortly before it joins subdomain A, while the second segment is longer (residues 179-317) and comprises of 6 mostly parallel  $\beta$ -strands and 5  $\alpha$ -helices. The two  $\beta$ -sheet subdomains of the large N-terminal domain are arranged parallel to each other and form a five layered  $\alpha\beta\alpha\beta\alpha$ -double sandwich structure with  $\alpha$ -helices occupying both the inner as well as outer faces of the sheets.



**Fig. 24. Ribbon diagram of DltA bound to AMP (stereo representation).**

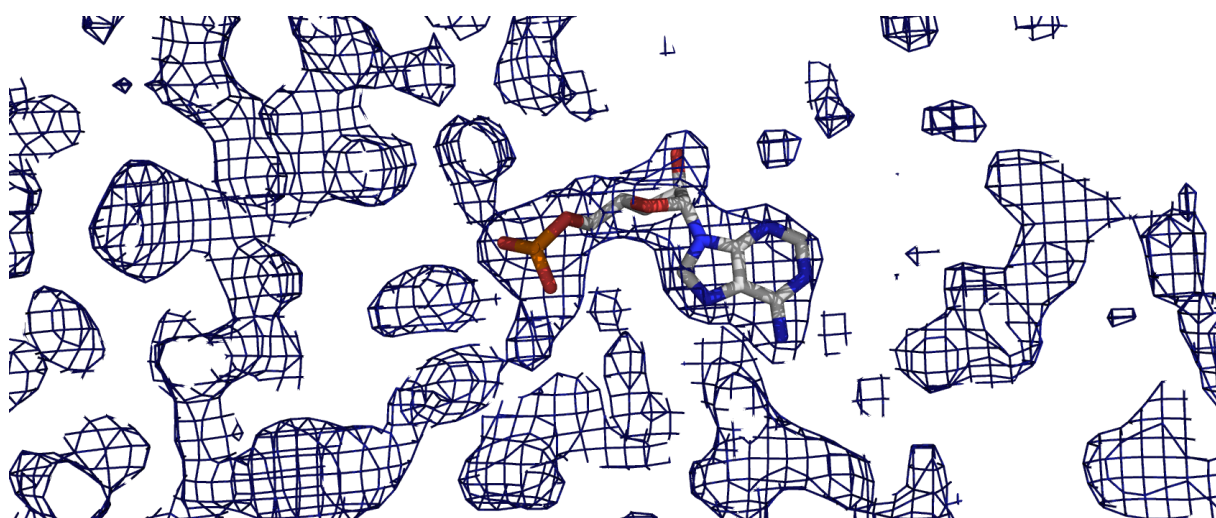
The large N-terminal domain is divided into 3 subdomains; subdomain A (Red), subdomain B (Yellow), subdomain C (Green). The small C-terminal domain is shown in Cyan. The position of the active site, indicated by the bound AMP (sticks), is found at the interface of the large and small domains

The  $\beta$ -barrel subdomain is also built from a single portion of the polypeptide chain (residues 318-395) and is made up of 7  $\beta$ -strands and 1  $\alpha$ -helix. The barrel is distorted and packs against the side of both the  $\beta$ -sheet subdomains forming two shallow depressions on the surface of the molecule. Superposition of the N-terminal domain with other known adenylation domain structures reveals the closest similarity to PheA, with an rms deviation of 1.70 for 357 residues. Further analysis indicated that relatively high rmsds observed for DltA to the large domains of the other adenylation domains are a result of significantly different arrangements of the three subdomains A, B and C. The C-terminal domain, also termed as the “lid”, is separated by a large cleft from the N-terminal domain and is constructed from 5  $\beta$ -strands and 3  $\alpha$ -helices. The small C-terminal domain has been seen to assume different orientations with respect to the large domain in the various known structures. In DltA, this

domain is seen in a very similar orientation to that observed in ACS (Gulick et al., 2003) which involved a large rearrangement of the small domain with respect to the PheA/DhbE structures.

### 5.5.3 ATP binding pocket

The structure determination of DltA, crystallized in the presence of its substrates D-alanine, ATP and MgCl<sub>2</sub>, revealed unambiguous electron density for AMP (Fig. 25) although no density was apparent for the D-alanine substrate even at the final stages of refinement.

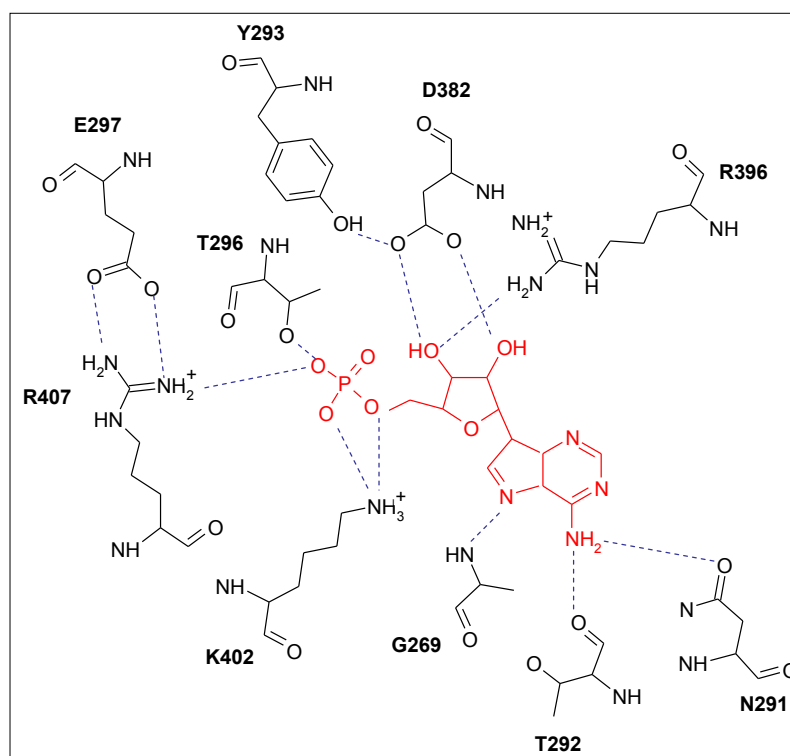


**Fig. 25. Electron density for the AMP ligand.**

The AMP moiety is positioned in a similar way as the PheA/DhbE structure and is bound in a cleft at the surface of the large N-terminal domain between residues of subdomain B and subdomain C. The crystal structure of DltA from *B. cereus* with a bound D-Ala-adenylate (Du et al., 2008) was also determined at about the same time as to our DltA structure from *B. subtilis*. The structure displayed the location of the small domain as seen in the ‘adenylation’ conformation. The AMP binding pocket of both the structures are essentially the same. The adenine ring lies in a slot sandwiched between the side chains of Y293, V320 on one side and the main chain atoms of residues <sup>269</sup>GEV271 on the other side (Fig. 26). In addition to hydrophobic as well as van der Waals interactions, the adenine ring is held in place by hydrogen bonds of the N6 amino group with N291 O<sub>δ1</sub> and T292 O and possibly N7 of AMP with G269 NH.

The 2'- and 3'- hydroxyl groups of the ribose ring are involved in hydrogen bonding interactions with D382, a completely conserved residue belonging to the core A7 motif. This interaction with D382 fixes the central portion of AMP. D382 also forms a hydrogen bond

with the OH of Y293, a residue conserved in many members of this family. The same is also observed in PheA (Conti et al., 1997) while in DhbE (May et al., 2002) the corresponding residue is F331 and instead a water molecule at the vertex of this residue forms a hydrogen bond with the highly conserved D413. The 3'-hydroxyl also interacts with R396, a strictly invariant residue among the adenylate forming family of enzymes; such an interaction of the 3'-hydroxyl was also observed in ACS (Gulick et al., 2003) but not in PheA. In DltA, the side chain of R396 is positioned by D398 just like in ACS while in PheA the homologous residue is positioned further. The ribose O-5' as well as the O2P of AMP hydrogen bonds with K402, this is also observed in LC-FACS (Hisanaga et al., 2004), while the equivalent lysine in PheA and luciferase lie much further from the active site due to the conformational change and are disordered and instead another lysine, K517 (PheA no.), a highly conserved residue belonging to the A10 motif interacts with both the ribose ring oxygens as well as the substrate carboxylate oxygen. The O3P of AMP interacts with the highly conserved T296 as well as with R407, this interaction with R407 is also observed in ACS but not in PheA as the equivalent arginine is much further away from the active site lying on the surface and is disordered. Interestingly, R407 is involved in a salt bridge to the invariant E297 just like in ACS, its guanidinium moiety occupying the position of the active site  $Mg^{2+}$  ion in previously determined structures of PheA and DhbE.



**Fig. 26. Schematic representation showing interactions between the protein (black) and the AMP ligand (red).**

Possible hydrogen bonds are shown as in dotted lines (blue)

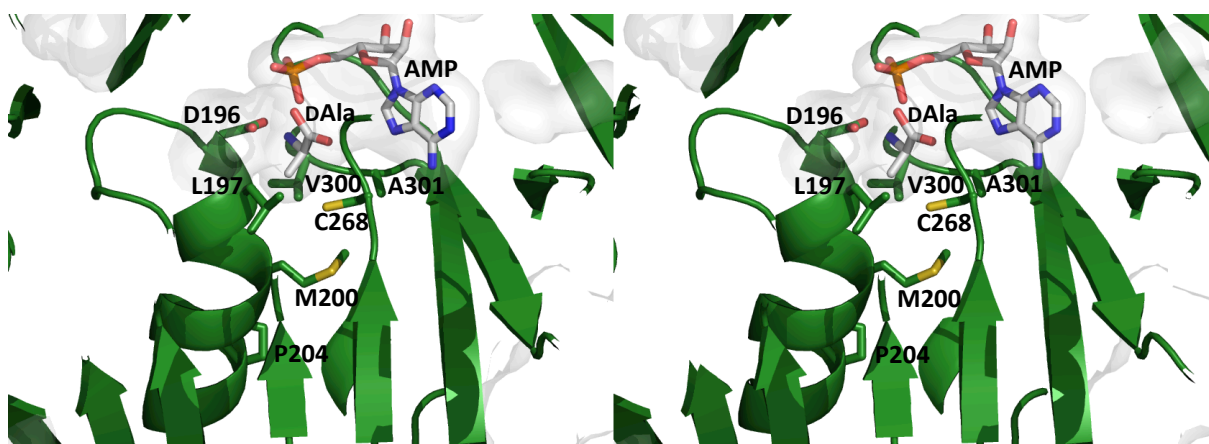
#### 5.5.4 D-alanine binding pocket

Although D-alanine was included in the crystallization conditions, no density for the substrate could be observed. Alignment of the active site with PheA and DhbE revealed that the binding pocket of D-alanine should lie in more or less similar position as PheA/DhbE near the intersection of the three subdomains that form the large N-terminal domain. This was later confirmed with the structure of DltA from *B. cereus* which contained a bound DAla substrate (Du et al., 2008). The specificity pocket of DltA is much shallower than the adenylation domains that accept the larger substrates, mainly forming interactions with the  $\beta$ -strands and an  $\alpha$ -helix of subdomain B of the large domain. The invariant D196 at the start of the “active site helix” conserved among all amino acids activating A-domains, is at the same position as in PheA and plays a role in stabilizing the  $\alpha$ -amino group of the substrate. In contrast the aryl acid activating domain DhbE, which lacks an  $\alpha$ -amino group, bears the neutral Asn at this position that hydrogen bonds to 2'-OH group of DHB. The D-alanine binding pocket is mainly surrounded by hydrophobic amino acids having large side chains that would mediate specificity for only smaller substrates. L197, lining the sides of the D-alanine pocket corresponds to the much smaller A236 in PheA and the side chain of this residue would block binding of substrates with larger side chains. In addition, the active site  $\alpha$ -helix is distorted due to the presence of P204, resulting in a further compression of the specificity pocket. Such a phenomenon was previously reported in DhbE which suggested that these enzymes utilize subtle changes in structure to configure the active site for selectivity towards cognate substrate. This distortion in the specificity helix is also seen in LC-FACS and ACS. The D-Ala pocket is lined at the bottom by M200, on one side by V300, A301 and by C268 on the opposite side.

Comparison of the binding pocket with that of *B. cereus* DltA containing the DAla substrate shows a very high similarity both in sequence as well as structural location of residues (Du et al., 2008). C268 is seen in one of the alternate conformations suggested by our DltA structure. All the specificity mediating residues are conserved. However due to difference in domain rearrangements between the two structures, catalytically important residue K491 (relevant to the ‘adenylation’ half reaction) is seen close to binding pocket positioned for substrate binding. Also the salt bridge between R407 and E297 is absent as a consequence of domain movement. Another important residue close to the binding pocket i.e. F195 (DltA numbering) is seen in a different side chain conformation to that of *B. cereus* DltA. The two different

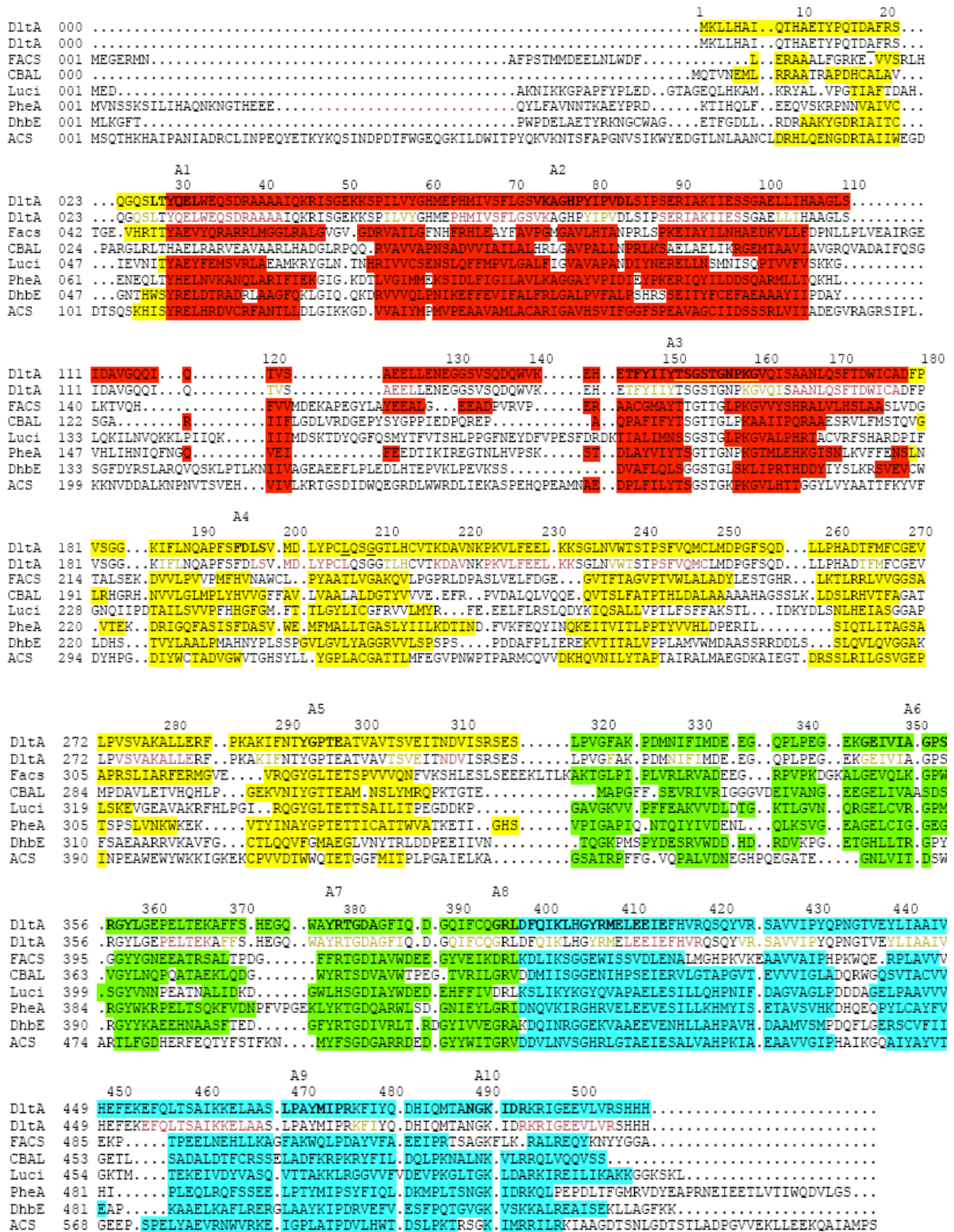
locations of the side chain of this residue may play significant role in the catalysis of the both half reactions of this enzyme family.

In order to explain the selectivity towards D-alanine, a DAla substrate was modelled into the structure following superposition of the phenylalanine-containing PheA structure on DltA. Fig. 27 shows how a DAla substrate would easily fit into the specificity pocket without any hinderance from the surrounding residues; however the side chain of C268 would restrict access of L-alanine to the pocket. With the exception of the ACS, this position (C268) is occupied by an alanine or glycine residue in other A-domains (Fig 28).



**Fig. 27. Specificity pocket of DltA.**

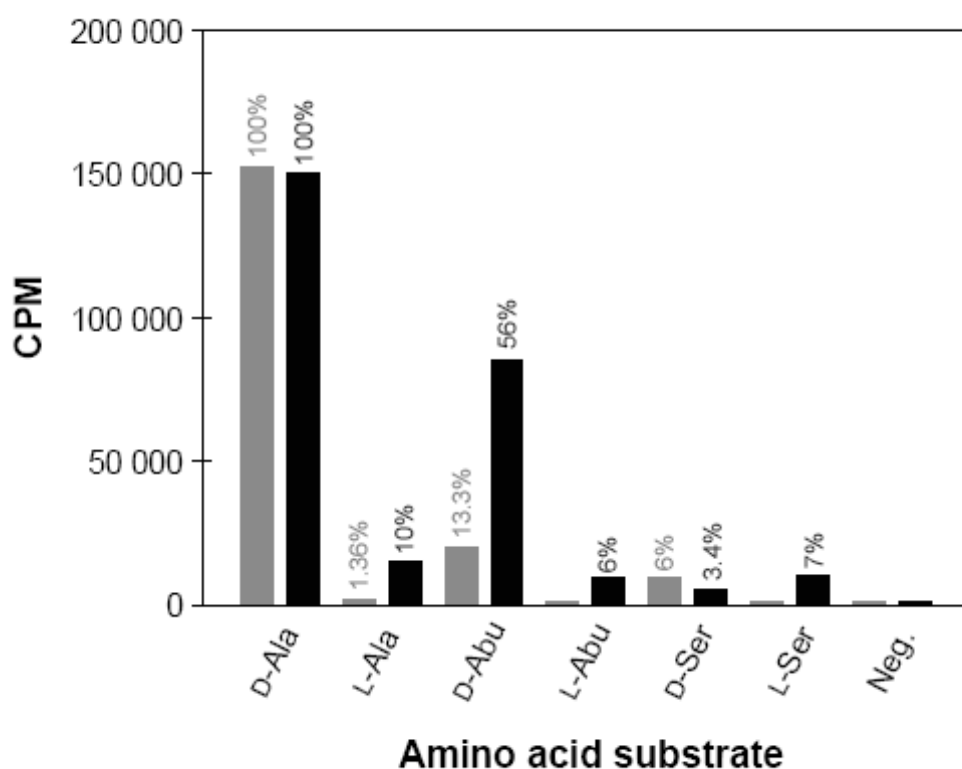
A DAla substrate has been modelled into the active site close to the AMP moiety. The specificity mediating residues are labelled. The active site helix is distorted due to P204.



**Fig. 28. Structure based sequence alignments of DltA with members of the adenylate forming family for which structures are available.**

Alignment is based on individual subdomains, colour coding is according to Fig. 24, where red denotes subdomain A, yellow denotes subdomain B, green denotes subdomain C, and cyan denotes the c-terminal domain. The core regions are shown in bold.

To test the postulated role of C268 as a selectivity determinant for D-alanine, this cysteine residue was mutated to an alanine by site directed mutagenesis and the ATP-PPi assay was performed for the C268A DltA mutant and compared to wild type (Fig. 29). The C268A mutant still retained its high activity for D-alanine, however it was also able to accept an L-alanine with an activity up to 9.7% of that for the cognate D-alanine and nearly 10 fold more than that shown by wt-DltA for L-alanine. The mutant showed more than 4-fold increase in activity against DAbu (56%). The mutant also showed several fold increase in activity for LABu as well as L-Ser, while the activity against D-Ser decreased.



**Fig. 29. ATP-PPi exchange assay for wt-DltA (grey) and C268A mutant (black).**

The highest exchange activity for the cognate D-alanine substrate was defined as 100% and the values obtained for other amino acids were set accordingly.

### 5.5.5 Domain reorganisation in DltA

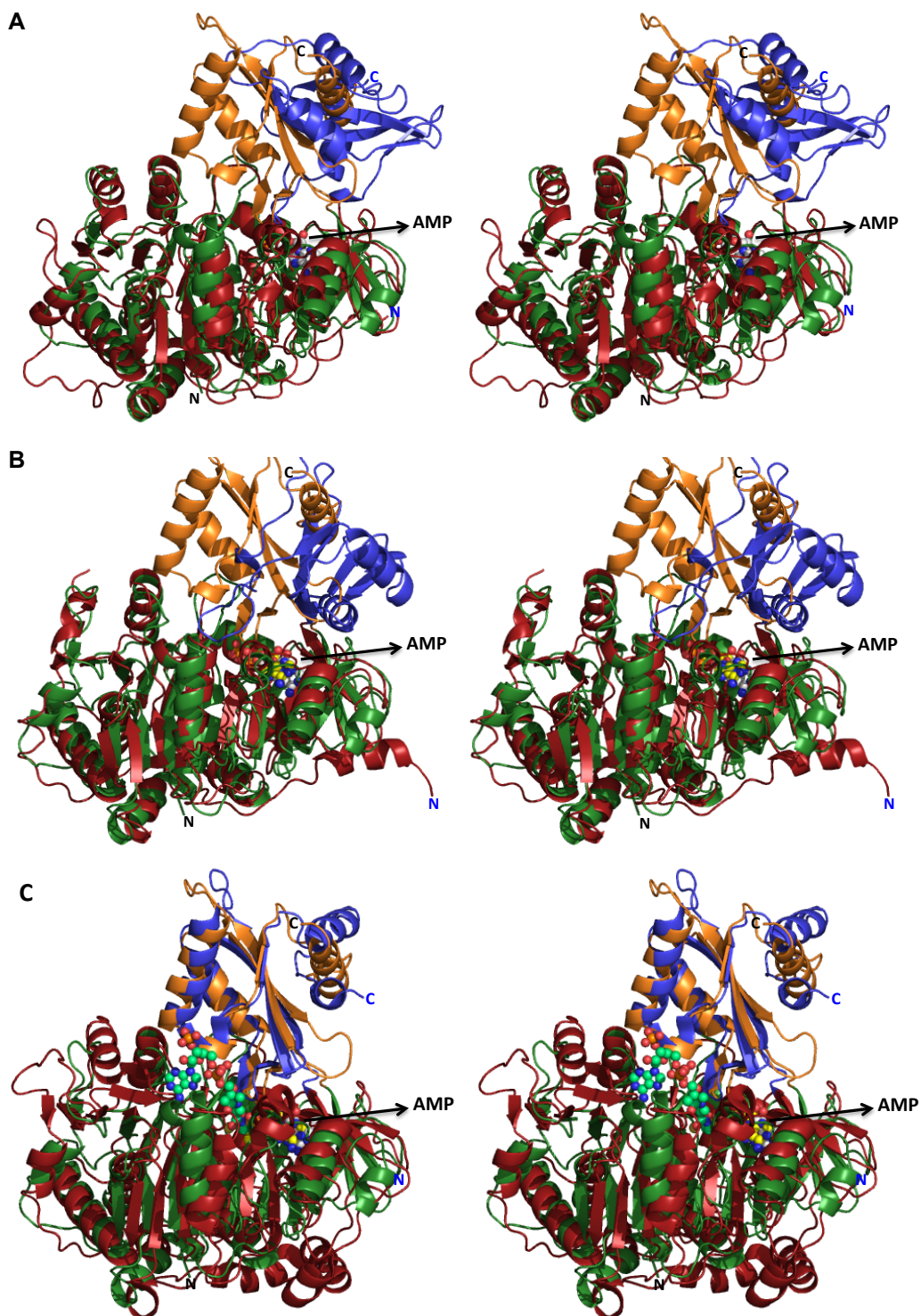
In contrast to the open conformation first seen for the firefly luciferase (Conti et al., 1996), the structure of DltA revealed a closed conformation of the two domains where the large N-terminal and small C-terminal domains come together to form the active site that is almost completely isolated from bulk solvent (Fig. 30A). This closed conformation is different from the closed conformation that has been seen in the previous crystal structures of PheA (Conti et al., 1997), DhbE (May et al., 2002), yeast ACS (Jogl and Tong, 2004), CBAL (Gulick et al.,

2004) and firefly luciferase (Nakatsu et al., 2006) as well *B. cereus* DltA (Du et al., 2008) in respect to the orientation of the small domain relative to the large domain and a rotation of approximately 140° is required to move the small domain of DltA onto that of PheA (Fig. 30B).

However the structure of DltA displayed an identical conformation to that observed in the ACS from *Salmonella enterica* (Gulick et al., 2003) (Fig. 30C). This conformation was also later seen in the crystal structures of LC-FACS (Hisanaga et al., 2004). In this conformation, the strictly invariant residue, K491 of DltA is situated away from the active site, in exactly the same location as in ACS. This residue is part of the active site as seen in the structure of PheA and DhbE and mutation as well as acetylation of this residue has been shown to affect the adenylation half reaction while having no effect on the thiolation reaction (Branchini et al., 2000; Starai et al., 2002). With the knowledge of all these structures, it has been proposed that members of the AMP family will adopt different conformations along their reaction cycle, with the conformation of PheA as a representative for the adenylation half reaction and that of ACS for thiolation half reaction.

DltA, which was crystallized in the presence of its substrates DAla, ATP, and MgCl<sub>2</sub> (required for the adenylation reaction), revealed the thiolation conformation even in the absence of DltC, its conjugate PCP partner, required for the thiolation half reaction. Thus the structure of DltA revealed for the first time that A-domains of NRPSs that have PCP partners will also adopt a similar conformation like the CoA binders of this family to catalyze the thiolation reaction but the switch in conformation must be more complex than that proposed previously (Gulick et al., 2003).



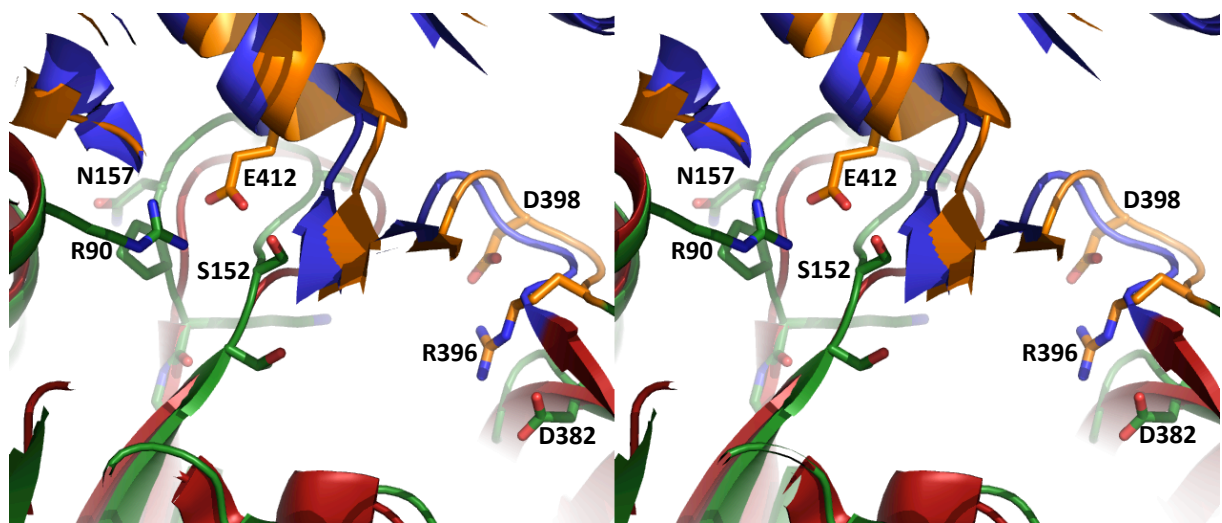


**Fig. 30. Comparative overlay of the different domain rearrangements among the adenylate forming family.**

A) Luciferase overlaid onto DltA B) PheA onto DltA C) ACS onto DltA. DltA is always shown in green (large domain) and orange (small domain). Luciferase, PheA and ACS are all in same colour ie maroon (large) and blue (small). The N and C-termini are labelled N and C respectively, AMP is shown as spheres.

### 5.5.6 The hinge region and P-loop

In addition to the more or less structurally conserved large and small domains, comparison of the different A-domains reveals two regions as structurally variable. Connecting the large and small domains, the hinge region in DltA (G395-F399) adopts an open turn conformation. Supporting this conformation is invariant residue R396; the guanidinium group is sandwiched between the carboxylate side chains of (absolutely conserved) D382 and highly conserved D398. The P-loop (phosphate binding loop or walker A motif, residues T151-K159) (Saraste et al., 1990) represents the most highly conserved region of A-domains. Despite evolutionary conservation, the loop is able to adopt multiple conformations, and is frequently disordered in crystal structures (Conti et al., 1996; Conti et al., 1997). In the DltA structures described here, the loop is well defined and adopts a conformation similar to that observed in *S. enterica* ACS (Gulick et al., 2003) (Fig. 31). The P-loop makes close contacts to residues in the small domain, and is stabilised in particular by charge-assisted hydrogen bonds (from S152 O<sup>-</sup> and N157 NH) to the carboxylate group of absolutely conserved E412. The latter residue also approaches R90, highly conserved among DAla-activating enzymes (Fig. 34).

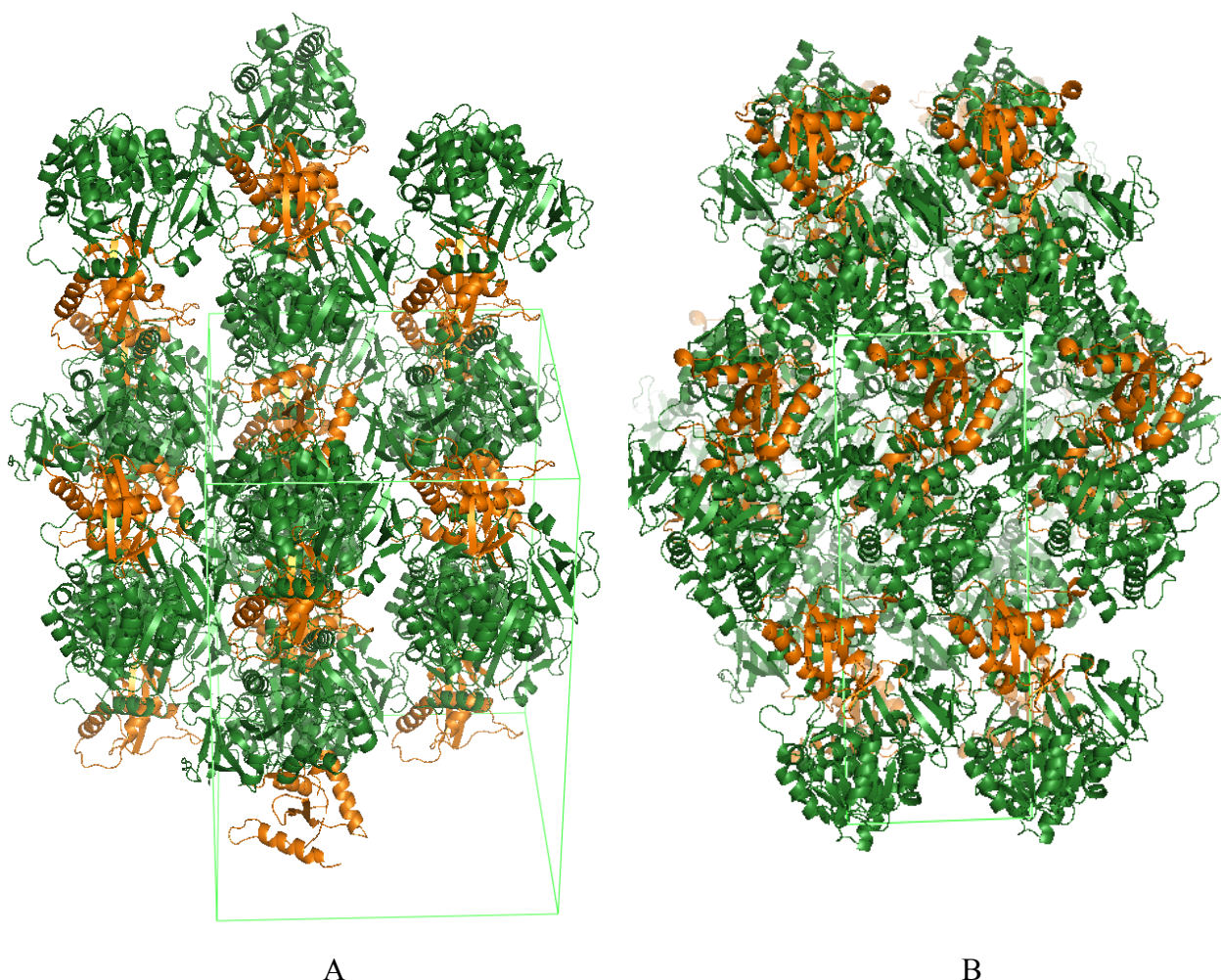


**Fig. 31. Stereo representation of the P-loop and hinge region.**

DltA was overlaid with the ACS. The large and small domains of DltA are in green and orange respectively while those of ACS in red and blue. The P-loop of both DltA and ACS adopt a similar conformation and interacts with strictly invariant E412. The hinge regions are also seen in a very similar conformation and side chain of R396 is seen sandwiched between the side chains of D382 and D398. Residues are shown for only DltA as sticks.

## 5.6 Conformation of DltA is not solely stabilized by crystal packing.

It is unlikely that the conformation of the small domain observed for DltA is solely stabilized by crystal packing. In fact the change in conformation of the small domain seems to have taken place first, as a part of the catalytic cycle of the enzyme, and then the crystal was packed. This is supported by the fact that both crystal forms of DltA, although crystallized in two different space groups, I222 and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, displayed an identical conformation of the small domain. Fig. 32 shows crystal packing in both crystal forms of DltA. It can be seen that the small domains are solely surrounded by the large domains of adjacent symmetry related molecules and there are no interactions with small domains from the neighbours.



**Fig. 32. Crystal packing in the I222 (A) and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> (B) crystal form of DltA.**

The small domains (orange) are seen to be surrounded by the large domains (green) from symmetry related molecules. The unit cell is shown in green.

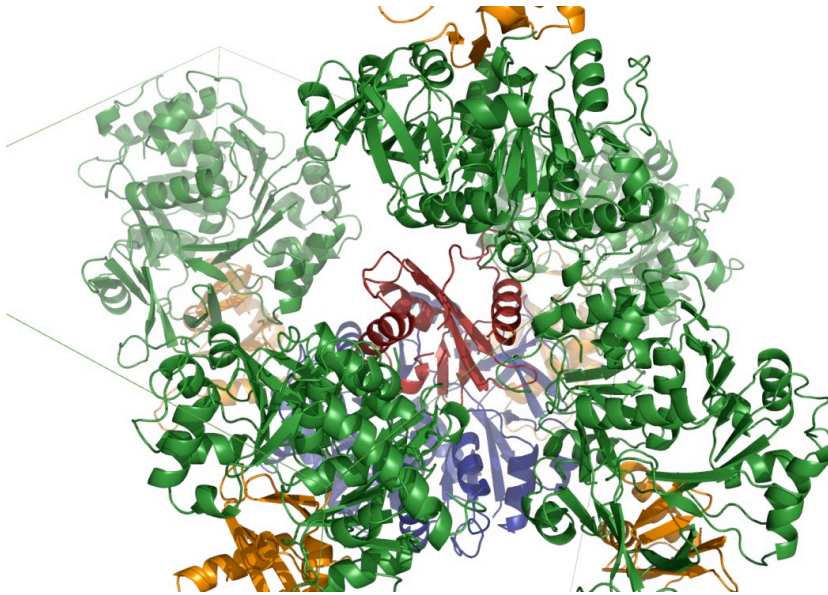
Table 9 shows the contacts formed by the small domain with symmetry related molecules in both crystal forms. All contacts are observed with the large domains, besides there are not so many interactions and most of them are also not so strong. There are a few salt bridges but they are also not formed with identical residues in both crystal forms i.e. K486 forms a salt bridge with two different glutamates, E397 in I222 crystal form and E359 in P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> crystal form. Interactions with only one symmetry related molecule i.e. (7) for I222 and (3) for P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> are nearly identical. All other interactions with neighbouring symmetry related molecules are different, suggesting that despite differences in interactions, the small domain adopts identical conformation in both crystal forms.

**Table 9. Interactions formed by the small domain with symmetry related molecules in the both the I222 and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> crystal forms of DltA (Distance cut off of 2.0-4.0 Å).**

The +/- signs indicate whether the interaction exists or not in the particular crystal form. The numbers in brackets indicates the interacting symmetry related molecule.

Residues involved in contacts from the small domain. (Reference molecule)	I222	Contacting residues from the symmetry related molecules	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	Contacting residues from the symmetry related molecules
Y455	+	E125, S72 (7)	+	E125, S72 (3)
N458	+	I142 (7)	+	Q141 (3)
E462	-		+	E125 (3)
E474	-		+	H285 (1)
S483	+	N352 (6)	+	E358, E359 (4)
K486	+	E397 (6)	+	E359 (4)
K487	+	E365 (6)	-	
R500	+	E74 (7)	+	E74 (3)
Q505	+	Y38 (5)	-	
H507	+	G207, G208 (5)	-	
N513	-		+	S206, K209 (1*)
L526	+	Q68 (7)	-	
	+	Q68 (7)	+	Q68 (3)
	+	S72, G73 (7)	+	G73 (3)
V527	+	I71, G73, Q162 (7)	+	I71 (3)
S529	-		+	Q40 (1)

Finally, it is also worth noting that the small domain is not so tightly packed by the surrounding symmetry related molecules and there is enough room for it to rotate at least to some degrees in either crystal form (Fig. 33).



**Fig. 33. DltA in the I222 space group surrounded by close symmetry related molecules.**  
The original molecule is shown in blue (large domain) and maroon (small domain) while the symmetry related molecules are shown in green (large domain) and orange (small domain).

## 6 Discussion

### 6.1 Stereo specificity of DltA is influenced by residue at position 268

DltA is a unique example of an adenylation domain that activates a non-proteinogenic amino acid, D-alanine. ATP-PPi exchange assays for wild type DltA revealed a high specificity for the cognate DAla substrate with virtually no activity for LAla. The lack in activation of LAla clearly demonstrated the high stereo-selectivity of DltA. Minor activation of the non-proteinogenic amino acids, D-amino butyric acid (DAbu) (13.3%) and DSer (6%), both of which possess small side chains, might be due to similarity in structure to cognate D-alanine substrate. The fact that DltA did not activate LAbu and LSer also further emphasizes selectivity of the enzyme for D-enantiomers.

The structure of DltA provides an explanation for the stereo selectivity of DltA. Modelling studies with the DltA structure indicated that the side chain of C268, a residue close to the DAla binding pocket, would restrict access of LAla as its  $\alpha$ -methyl group would lie too close to the side chain thiol of C268 resulting in steric clashes. In order to test this, a variant was constructed where C268 was mutated to a smaller alanine. ATP-PPi assay for C268A variant revealed that while the mutant enzyme still exhibited a strong preference for DAla, a high activity was also observed towards DAbu (56%). Most significantly, the variant now accepted the L-amino acids LAla (9.7%), LAbu (6%) and LSer (7%), thus demonstrating the introduction of a single mutation, C268A, was sufficient enough to yield a variant of relaxed specificity therefore C268 seems critical for determining enantiomeric selection by DltA. Such a mutation in combination with other site specific mutations may further relax the specificity of DltA.

### 6.2 Phylogenetic relationship of DltA to members of the superfamily

A search with the *B. subtilis* DltA sequence as query was performed using PSI-BLAST at NCBI server targeting a maximum number of 20000 sequences. The DltAs were aligned separately. The DAla-specific NRPS starter domain LnmQ (Tang et al., 2007) was also included in the alignment. The alignment reveals that the specificity-determining residues described for DltA (section 5.5.4) are highly conserved (Fig. 34). The streptococcal DltAs differ from those of other bacteria at residue 268 (*B. subtilis* numbering): while most DltA sequences exhibit a cysteine residue in this position, the streptococcal enzymes are demarcated by an aspartic acid residue. Since this residue is an important specificity

determinant, the inhibitor used against *B. subtilis* may also not be very suitable and thus the streptococcal class of enzymes may require design of different inhibitors for their effectiveness. *Clostridium difficile* also displayed a residue different to Cys or Asp for position 268 i.e. Ile. It is therefore apparent that DAla-activating enzymes can be distinguished by the presence of a bulky residue in position 268, while most other conventional A-domains of NRPSs that accept larger substrates like PheA, DhbE, exhibit a much smaller residue at this position i.e. Gly or Ala (Fig. 28)

The DAla-specific NRPS, LnmQ (Tang et al., 2007), also shows conservation for the equivalents of C268 as well as L197, although the ‘kink-forming’ residue P204 is replaced by a cysteine. A further BLAST search was performed taking LnmQ as a query sequence. Although several A-domains containing G/A for position 268 were observed, a series of hypothetical A-domains from diverse microorganisms for which position 268 is consistently occupied by a cysteine residue could also be identified. Like LnmQ these enzymes showed deviations from the DltA specific P204 and M200 residues. Though it seems that these uncharacterised proteins activate D-amino acids and their preferred substrates is possibly DAla, however the fine rearrangements near the active site described here for DltA (due to P204) and for DhbE (*cis*P241 (May et al., 2002) as well as the differences in organisation of individual subdomains, makes it uncertain to classify their substrate specificity.

A phylogenetic analysis was also performed with a few representative members of the adenylation domain family which showed significant sequence similarity in the blast search database. Many members for which structures are known could not be found in the search database (i.e. the maximum number sequences that could be targeted was 20000) nevertheless similar enzymes from the closest genera were selected for phylogenetic analysis. Fig. 35 shows that the earliest members seem to be the acyl and aryl-CoA ligases. Luciferases lie close to the acyl-CoA ligases. These are followed by single NRPS enzymes ie the stand-alone aryl-acid activating domains of NRPSs followed by the DltAs. The NRPS A-domains (amino acid activating domains) seem to have diverged the furthest. These results are also in agreement with earlier studies that suggested that NRPS A-domains may have evolved from a common ancestral form of adenylate forming enzymes that utilize CoA-SH as a small acceptor molecule for their synthesized reactive adenylates (Linne et al., 2007). In these studies a set of six structurally related enzymes were tested for their ability to synthesize acyl-

4CoAs in vitro. DltA showed highest activity for acylation of CoA-SH after ACS. This may be because DltA is a single enzyme involved in the D-alanylation of lipoteichoic acid of bacterial cell wall which is highly conserved among gram positive bacteria (related to primary metabolism) and thus may be closer to ACS also belonging to primary metabolism. The aryl acid activating domain ie DhbE showed reduced activity in comparison to DltA thus reflecting a divergence in function required for its role secondary metabolism. In contrast, the NRPS A-domains showed only residual activity, which may be because these are module integrated enzymes required for secondary metabolism and thus have diverged furthest from ACS. Recent studies by Koetsier *et al* (Koetsier et al., 2011) have shown that some CoA ligases have considerable activity for  $\alpha$ -amino acids resulting in the formation of  $\alpha$ -aminoacyl-CoAs however these thioesters are far less stable than the normal acyl-thioesters thus posing evolutionary demands. This may be the reason why  $\alpha$ -aminoacyl-CoAs are not used as intermediates in secondary product biosynthesis. The thioester-activated amino acids of NRPSs remain covalently attached to the enzyme, preventing their hydrolysis and thus would be much more stable than the  $\alpha$ -aminoacyl-CoAs.



D1tA 001  
P39581  
P0C397  
Q8Y8D4  
Q03AZ2  
Q183T6  
P0A398  
A3CR87  
Q53526  
Q99ZA6  
A4VU15  
Q8GGN5  
A3WUG7  
Q0LNQ7  
Q12QC2  
A3GT14

1 10 20  
M K L L H A I Q T H A E T Y P Q T D A F R S  
M K L L H A I Q T H A E T Y P Q T D A F R S  
M T D I I N K L Q A F A D A N P Q S I A V R H  
M T T S I I E R I D A W A E K T P D F P C Y E Y  
M I D N V I T A I D R V A A E H P T R V A Y D Y  
M K I I E G I K K Y S N T . D R T A L M C  
M S N K P I A D M I E T I E H F A Q T Q P S Y P V Y N V  
M S N K V I H D M I E A I E H F A Q V Q P D F P V Y D I  
M A N K K I K D M I A T I E N F A Q E Q A E F P V Y N I  
M I K D M I D S I E Q F A Q T Q A D F P V Y D C  
M D R E D L S I M S F K E E S I K M I S S M L E R V E H F A T E N P D Y P V Y H  
M S G A K L L H W F L D G L A R N P R G T A L R I  
M T P L A A S F V K S A T Q F R D K A A L W V  
M M I E F . N Q M L . . . E V V P P C Q S E P N Q A H A T S Q L H S G F L Q S V Q R Y P N N I A L T I  
M I Q L R D C S L T F A H R P A L W V  
M P Y N S A T D L L Q R E G Y G V V H N Q F I N N M A M Y C E T R H S I L A A I Q A N R H P A L W V

30 40 50 60 70 80 90 100 110  
D1tA 047  
P39581  
P0C397  
Q8Y8D4  
Q03AZ2  
Q183T6  
P0A398  
A3CR87  
Q53526  
Q99ZA6  
A4VU15  
Q8GGN5  
A3WUG7  
Q0LNQ7  
Q12QC2  
A3GT14

Q Q Q S L T Y Q E L W E Q S D R A A A A I Q K R I S G E K K S P I L V Y G H M E P H M I V S F L G S V K A G H P Y I P V D L S I P S E R I A K I I E S S G A E L L T H A A G L  
T D E L T Y Q L M D E S S K L A H R L Q G . . . S K K P M L F G H M S P Y M I V G M I G A I K A G C G Y V P V D T S I P E D R I K M I N K V Q E F V E N T D E  
A G T R L S Y K E L K R Q S D A F G S F L K N L I T D K E K P I I V Y G H M S P L M L V A F L G S I K S G R A V V P V D V S M P V E R I E Q I K K A A D P S M F I C T E E L  
E G T O Y T Y A Q L K E G S D R L A G F F A E . T . L P E H E P I I V Y G G Q T F D M V E V F L G L S K S G H A Y I P I D T H S P N E R I T Q V Q V A H T P A I I E V A P L  
N G D K L S Y K D L N E Y S D A I S V F L K D V . Y K E E D T P I V I G N K E N M M A C M I G A L K S G R A V V P L D I S F P I D R V F E V T K E I K P K V L F N F S D E  
L G Q E H T Y G D L K A D S D S L A A V I D Q L . G L P E K S P V V V F G G Q E Y E M L A T F V A L T K S G H A Y I P I D S H S A L E R V S A I L E V A E P S L I I A I S A F  
L G Q V H T Y G D L K D S D S L A A Q I D R L . G L P D K S P V V V F G G Q E Y E M L A T F V A L T K S G H A Y I P I D S H S A L E R V A A I E V A E P S L I I A I N D F  
L G E I H T Y G E L K A D S D S L A A H L D Q L . D L T A K S P V V V F G G Q E Y A M L A S F V A L T K S G H A Y I P I D H S A L E R I E A I L E V A E P S L V I A V D F  
L G E R R T Y Q L K R D S D S I A A F I D S L . A L L A K S P V L V F G A Q Y T D M L A T F V A L T K S G H A Y I P V D V H S A P E R I L A I E I A K P S L I I A I E A F  
L G Q L Y S Y G D L K A D S D S L A A H I D S L . D L P A N S P V L V F G G Q E Y Q M L A I F V A L T K A G H P Y I P V D S H S A L D R I E A I L E I A E P S L V F A V A D F  
N D R S W T Y A E A D R T A R S W A A A L H S G . G R . A P R R V G V L A A K E E S L L G F L A A L Y A G C T A V L N P E Y P V G R N R D I A A A G L D A L I V D P A G  
D G E N Y T Y G D L H E F A L R L A G G F P L S . S K S G F D T C A I F A H R S L V A Y A A I A A C H L A R Y A V A L T P A Q I G R S Q S I I T Q S Q P A V I V D S K  
G N K O V D Y V K L Y T V A Q R W A F A L R Q S . T K . P L H R V G I F A Y R S E A Y I G I L A S L L A G A T F V P L N Y N F P L Q R T Q A M I E Q A E L D A I I V D H Q S  
E D R E Y S Y R E L F A Q A D L I A C Q I Q K L . T T . T D A V L V L S K K S F K Y A G I L S C F L A G V T Y I P L N D S F P T K L V D I I Q S S G C T L L A D P Q C  
K Q K T Y T Y Q E M T D M A L S L S D Y W H L . . . Q G V Q R V A I L S V R D L A A Y S A I W S Y L L G M T Y I P L N A R A T T E Q I Q E T L I A T Q C D S I M V D A G C

111111 22222222222

120 130 140 150 160 170 180  
D1tA 135  
P39581  
P0C397  
Q8Y8D4  
Q03AZ2  
Q183T6  
P0A398  
A3CR87  
Q53526  
Q99ZA6  
A4VU15  
Q8GGN5  
A3WUG7  
Q0LNQ7  
Q12QC2  
A3GT14

I D A V G . . . Q O I Q T V S . . . A B E L L E N E . G G S V . S Q D Q W V K E H E T F Y I I Y T S G S T G N P K G V Q I S A A N L Q S F T D W I C A D . . . F P  
F E S L E . . . G E V F T I . . . . . E D I K T S . Q D P V . I F D S Q I K D N D T V Y I T F T S G S T G E P K G V Q I S A S L V Q T E W M L E L . . . N K  
N N L T I T G . . . C P V L T Q D . . . Q L M D A L E K H F G E V P D K E A C V N N D N Y Y I I Y T S G S T G N P K G V Q I S Q N N L V S F S N W I L Q D . . . F S  
I A V P D . . . V Q I I R A P . . . A L H E A E K T . H A P I S S L Q H A V V G D D N Y Y I I F T S G T T G K P K G V Q I S H D N L L S F V N W N I S D . . . F G  
N . P G D . . . I N V I D M . . . K L N Y I I N E Y Q G K S L D K E N W W K D E N A Y I L F T S G S T G K P K G V Q I S S N D L S F S D W I S P Y . . . L N  
L E O V S . . . T P M I N L A . . . Q V Q E A F Q . G N N Y . E I T H P V K G D D N Y Y I I F T S G T T G K P K G V Q I S H D N L L S F T N W M I T D . K E F A  
L A D V A . . . A P I F S A E . . . Q V Q T A F R E . G A S Y . E L S H P V Q G D D N Y Y I I F T S G T T G K P K G V Q I S H N N L L S F T N W M I T D . K E F A  
I D M L Q . . . V P V I Q Y S . . . Q L E E I F Q . K L S Y . Q I N H A V K G D D Y Y I I F T S G T T G K P K G V Q I S H D N L L S F T N W M I N A . E A F A  
L T I E G . . . I S L V S L S . . . E I S A K L A . E M P Y . E R T H S V K G D D N Y Y I I F T S G T T G Q P K G V Q I S H D N L L S F T N W M I E D . A A F D  
L E T . S . . . L P V L G L S . . . E V E A I F E K . K S T Y . Q L I H P V S E D D Y Y I I F T S G T T G K P K G V Q I S H D N L L S F T N W M I R D . K E F A  
A Q L D E V A A A A P . P L V T . L A P R L G E W . . . S G P R P C T P V . . . L M P D P A G D R P P A G Q P D S L A Y I L F T S G S T G R P K G V P I R H G N V S A F L A A S L P R . . . Y D  
K S L P L L L Q K V R G P T V I . L P D Q D A . . . I P D W A A D P Q H Q F L N K S L L T S P G E P Q Q S G H D L A V M Y T S G S T G E P K G V I S H G N V A T V R V N V D H . . . F S  
D Q F L Q A D S L P V L P P C V I L P D C L R . . . A P . L L D T M I Y T Q A E L A E L P T D H E P V T G H P A Y L L F T S G S T G N P K G V I S H N A V H F L Q N A R . . . Y Q  
E S L L K L L Q L P D P L T V . L V S G K T I Z D N E T V N E Y A A S Y A N G H L Q S T G S T K H C A K A E P V S N S H L A Y L M F T S G S T G K P K G V P V S T Q N L L S Y L E H I R K L . . . Y D  
S R L S S L E T C I D R L H I Y A L P D V D V E P L R Q O Y P Q H T P H T V Q I T E Q D . V E L L V V K Y H L D N E H E Y A I M Q T S G S T G K P K R I A V S Y S N L H C Y I S Q I D K L . . . F P

3333333333333333

190 200 210 220 230 240 250 260 270  
D1tA 178  
P39581  
P0C397  
Q8Y8D4  
Q03AZ2  
Q183T6  
P0A398  
A3CR87  
Q53526  
Q99ZA6  
A4VU15  
Q8GGN5  
A3WUG7  
Q0LNQ7  
Q12QC2  
A3GT14

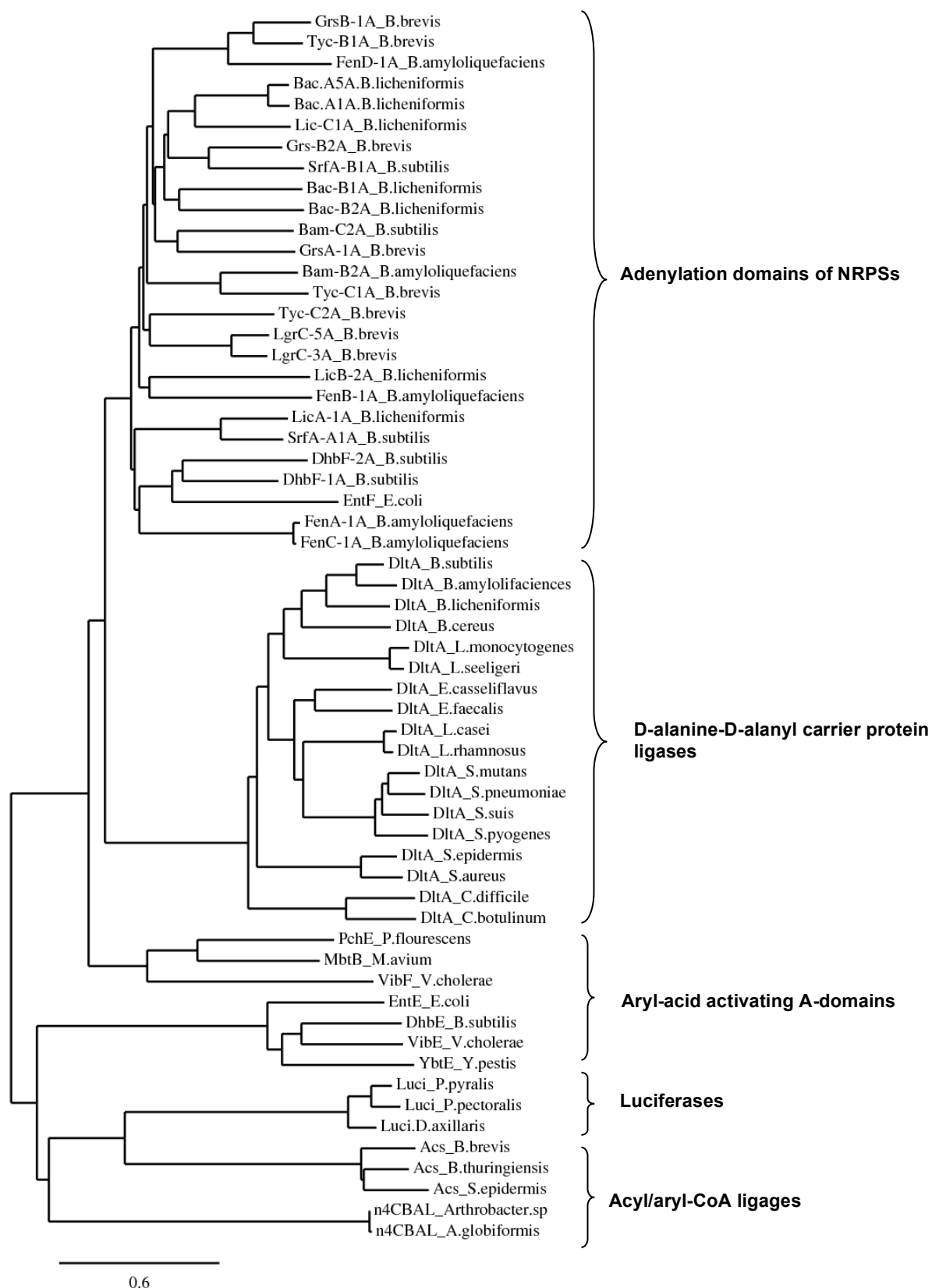
V S G G . . K I F L N Q A P F S F D L S V . M D . L Y P C L Q S G G T L H C V T K D A V N K . P K V L F E E L K K S G L N V W T S T P S F V Q M C . . . L M D P G F S Q D L L P H A D T F M F C G E V  
S G N K . . Q Q W L N Q A P F S F D L S V . M A . I Y P C L A S G G T L N L V D K M I N K . P K L L N E M L T A P I N I W S T S P S F M E M C . . . L L L P T L N E E Q Y G S L N E F F F C G E I  
L S Q G . . L R F L N Q A P F S F D L S V . M D . L Y P S L L S G G T L V P L D K I T A N . M K D L R E I P A Q N L D V W V S T P S F A D L C . . . L L D E N F N Q E N N P D L T R F L F C G E V  
L K E G . . V V A M S Q P P Y S F D L S V . M D . L Y P T L V L G S T L K L A P K E V T I D N . F K T L F A T L P K L G L N E W V S T P S F A E I A . . . L L D P N F N Q N N Y P D L T H F L F C G E E  
I D G S E . K V I M N Q P A Y S F D L S V . T T . I Y P G L I H G A T L F S I S K D V L A D . Y K E L F R Q F S I S D I A V W V S T P S F A G V C . . . I T E K E F N S K M L P N L E S M I F I G E A  
T P S R . . P Q M L A Q P P Y S F D L S V . M Y . W A P T L A L G G T L F T L P S V I T Q D . F K Q L F A A I P S L P I A I W T S T P S F A D M A . . . L L S D F F N S E K M P G I T H F Y F D G E E  
T P E R . . P Q M L A Q P P Y S F D L S V . M Y . W A P T L A L G G T L F A L P S A V T Q D . F K Q L F E T I L S L P A I W T S T P S F A D M A . . . M L S Y P S I G E M L P Q L T H F Y F D G E E  
T P H R . . P Q M L A Q P P Y S F D L S V . M Y . W A P T L A L G G T L F A L P K E I T A D . F K Q L F T T I N Q L P I G V W T S T P S F V D M A . . . M L S D D F N A Q Q L P H L T H F Y F D G E E  
V P K Q . . P Q M L A Q P P Y S F D L S V . M Y . W A P T L A L G G T L F A L P K E L V A D . F K Q L F T T I A Q L P V G I W T S T P S F A D M A . . . M L S D D F Q A K M P I W T S T P S F A D M A . . . Y D  
L P E R . . P Q M L A Q P S Y S F D L S V . M Y . W A P T L A L G G T L F A L P K E L T L D . F K T L F A A I Q T L P P K I W T S T P S F V D M A . . . L L S R E F D G E H L P D L T H F Y F D G E E  
F G P D . . D V F G Q V Y E L T F D L S M . F E . V W C A W S S G A C L T V L N R L Q A L N . P G . . . R Y I R A H G I T V W T S T P S L V A A L . . . R T R G L L G G N S L P S V R H T V F C G E P  
Y S E A . . D R F V H L P E L N F D L S V . H D . L F S A W S V G G T L Y C V P Q E V L I . P D . . . G F V K R R H L T A W T S V P S A V L I L . . . K K F N K L R T D A F E S I R A S M F C G E P  
I T P A . . D R L S Q T F D Q T F D L A I . F D . L F M A W N H G A A C V C V Q P I Q L L S . P F . . . R L I E Q G I T I W F S V P S V A A L L . . . R K Q K L L K P N L P N L R L S L F C G E A  
F T P Y . . D R H S Q F F E F T F D L S M . H D . I M V C W T S G G C L Y A A D G F A K L M . P L . . . H F A S K H K L T V W F S V P S Q V S A A Q A . . . V L K N K F P Q N L P N L R Y S L F C G E A  
L N A Q . . D R V G Q Y S D L T F D L S V . H D . I F Y S L I S G A C L Y V V E L A K L S . P A . . . E F I R H Q L T V W L S V P T V I E L A . . . L Q R Q T L T P H S L P S L R L S F C G Q A

280 290 300 310 320 330 340 350  
D1tA 268  
P39581  
P0C397  
Q8Y8D4  
Q03AZ2  
Q183T6  
P0A398  
A3CR87  
Q53526  
Q99ZA6  
A4VU15  
Q8GGN5  
A3WUG7  
Q0LNQ7  
Q12QC2  
A3GT14

L P V S V A K A L L E R F . . . P K A K I F N T Y G P T E A T V A V T S V E I T N D V I S R S E S . . . . . L P V G F A K . P D M N I F I M D E . E G . . . Q P L P E G . . . E K G E I V I A . G P S V S R  
L P H R A A K A L V S R F . . . P S A T I Y N T Y G P T E A T V A V T S I Q I T Q E I L D Q Y P T . . . . . L P V G V E R . . . . . L G . . . A R L S T T . . . D D G E L V I E . G Q S V S L  
L A K T A S E L L D R F . . . P D A V I Y N T Y G P T E A T V A V T Q V K T R E I I D A Y P S . . . . . L P L G V I K . P D M R L H I V D Q E T G . . . E V L P S E G . . . E K G E I V I L I . G A S V S K  
L V N K T A Q B L I T R F . . . P K A T V Y N T Y G P T E T T V A V T G M A I T Q D I V D Q Y P R . . . . . L P I G F A K . P D T E I F V V D E . Q G . . . N Q V S A G . . . T E G E L M V I . G P S V S K  
L S K N L T K E L M S R F . . . P N T R I I N G Y G P T E A T V G S V N D M T Q A I D D E K S . . . . . L P V G Y P M . S N E I K I L D E . D G . . . N E L K E N . . . E K G E I I I I . G P S V S K  
L T V K T A Q K L R E R F . . . P N A R I I N A Y G P T E A T V A L S A V A V T D E M L A T L K R . . . . . L P I G Y T K . A D S P T F I I D E . E G . . . N K L P N G . . . E Q G E I I V S . G P A V S K  
L T V K T A Q K L R D R F . . . P Q A R I I N A Y G P T E A T V A L S A V A V T D E M L Q N C K R . . . . . L P I G Y T K . A D S P T F I I D E . E G . . . Q K V P N G . . . Q Q G E I I V C . G P A V S K  
L T V K T A K L R Q R F . . . P Q A R I V N A Y G P T E A T V A L S A V A V T D E M L E T C K R . . . . . L P I G Y T K . P D S P T F I I D E . S G . . . H K L A N G . . . Q Q G E I I V S . G P A V S K  
L T V S T A R K L F E R F . . . P S A K I I N A Y G P T E A T V A L S A I T E M R M D N Y T R . . . . . L P I G Y P K . P D S P T Y I I D E . D G . . . K E L S N G . . . E Q G E I I V T . G P A V S K  
L T V K T A Q K L R E R F . . . P K A V I V N A Y G P T E A T V A L S A I T E M L V N Y K R . . . . . L P I G Y T K . P D S P T F I I D E . D G . . . Q V V P N G . . . Q Q G E I I V S . G P A V S K  
L P E E S A A Y W S A A A . . . P G T S I D N L Y G P T E L T I A C T A Y H W P S P . . . G S G T . . . . . V P I G E P N . T G L R Y V L L D . D G . . . V G A . . . D T G E L C V T . G P Q M F D  
F P G A L A A E W M V A A . . . P N S R V D N F Y G P T E A T V A V T A Y P C S R G T . . . D F S V . . . . . L P I G E P F . A Q E A V V C D A . D L . . . R P V S D G . . . E T G E L L V E . G T Q V S A  
L P K A T A E A W Q L A A . . . P N S I I D N L Y G P T E L T I A C A V Y R W N S L T S P A E C L N E V . . . . . V P I G K L Y . P G L T A V V V D A . N D . . . N P V P A G . . . E T E L L C V A . G P Q T F Q  
L P A S L T R D W A S V A . . . P H S I D N L Y G P T E A T I A Y T W L R V D P D K H G E C G I . . . . . I P I G K P F . G D N K V M V L D E . Q Q . . . R L V A S G . . . E I G Q L Y L V . G P Q V A  
L L H D L A E Q W Q A T . . . Q Q P V I N L Y G P T E A T I A V T Y H R F A V H S . . . G M A S . . . . . V P I G R A F . E B E C L A I N E Q G E L M F A P E G Y R G E L L L S . G K Q L V K

55555 66666666666





**Fig. 35.** Phylogenetic comparison of Acetyl-CoA synthetases (ACS), 4-chlorobenzoyl-CoA ligases (4CBAL), Luciferases (Luci), Aryl acid activating domains of NRPSs (DhbE, EntE, VibE, YbtE), D-alanine-D-alanyl carrier protein ligases (DltA), and the Adenylation domains of NRPSs (GrsA, GrsB, TycB, TycC, SrfA-A, SrfA-C, LgrC, BacA, BacB, BamC, FenA, FenB, FenC, LicA, LicB, LicC, DhbF etc)

### **6.3 The DltA structure allows modelling of the different functional states of NRPS adenylation domains**

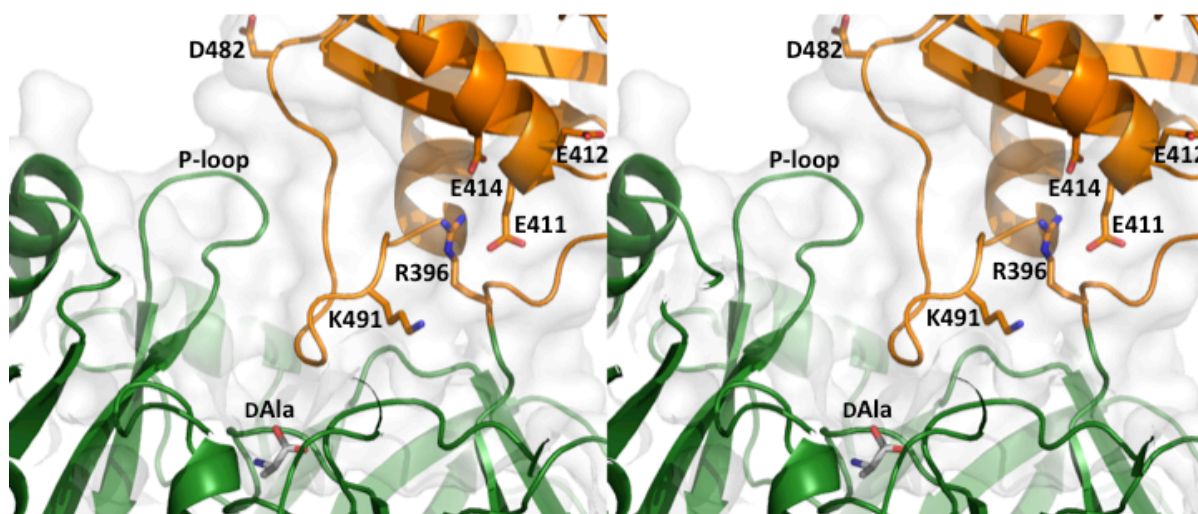
The DltA structure presented here, as well as sequence and structural conservation with other NRPS A-domains, confirm that those enzymes that transfer their activated amino acids to the phosphopantheynyl prosthetic groups of peptidyl carrier proteins (PCPs) are capable of adopting at least two major structural arrangements, as first shown for ACS (Gulick et al., 2003). The first arrangement i.e. “adenylation conformation” was until recently the only conformation seen for NRPS A-domain structures, both in the absence and presence of substrates (May et al., 2002). The second arrangement i.e. “thiolation conformation” originally seen in the ACS, is now observed in the DltA structure in an NRPS A-domain for the first time. Interestingly DltA revealed this conformation in the absence of any thiol acceptor (i.e. DltC or CoA), in contrast to earlier studies in which the binding of CoA was proposed to trigger the thiolation conformation (Gulick et al., 2003).

In the meantime, the crystal structure of a termination module SrfA-C has revealed the structure of an A-domain in an “open conformation” for the first time (Tanovic et al., 2008). DltA has also been crystallized both in the presence of ATP (Osman et al., 2009) as well as the adenylylated intermediate (Du et al., 2008), showing that the enzyme indeed adopts the „adenylation“ conformation. Thus DltA is a good model system for understanding the different catalytic stages for the superfamily of enzymes. In order to understand how the different conformations are attained, modelling of the various reaction steps in the catalytic cycle was performed for the DltA structure (by Prof. Dr. Milton Stubbs) (Yonus et al., 2008). The changes in interactions occurring along each of the different conformations are outlined below.

#### **6.3.1 Open conformation**

As mentioned previously this conformation represents an open active site accessible to solvent. By analogy to firefly luciferase and LC-FACS, it was assumed that such a state might exist in the absence of nucleotides, verified by the structural determination of SrfA-C (Tanovic et al., 2008). Although it should be noted that neither DhbE nor CBAL, both crystallized in the absence of substrates, exhibited an open conformation, it can be assumed that the reaction starts with the open conformation as originally seen for firefly luciferase (Franks et al., 1998). The large and the small domains show only few direct contacts (Fig. 36). In the open conformation the strictly invariant residue in the hinge region i.e. R396 is sequestered from the active site and interacts with two oppositely charged conserved residues

of the small domain i.e. E411 and E414. E414 is a strictly invariant residue of the A8 region while E411 is conserved among NRPS A-domains. Invariant residue, K491 (A10 motif) of the small domain is in the vicinity of the active site, poised for the binding of substrate. This residue has been seen in previous structures to interact both with the AMP as well as the substrate and has been associated with catalysis of adenylation reaction only (Conti et al., 1997; May et al., 2002). The side chain of this residue is disordered in firefly luciferase structure probably due to absence of substrates. The P-loop adopts an open conformation, reaching towards the small domain where it could be stabilised by interactions with conserved D482.



**Fig. 36. DltA structure modelled into the open conformation.**

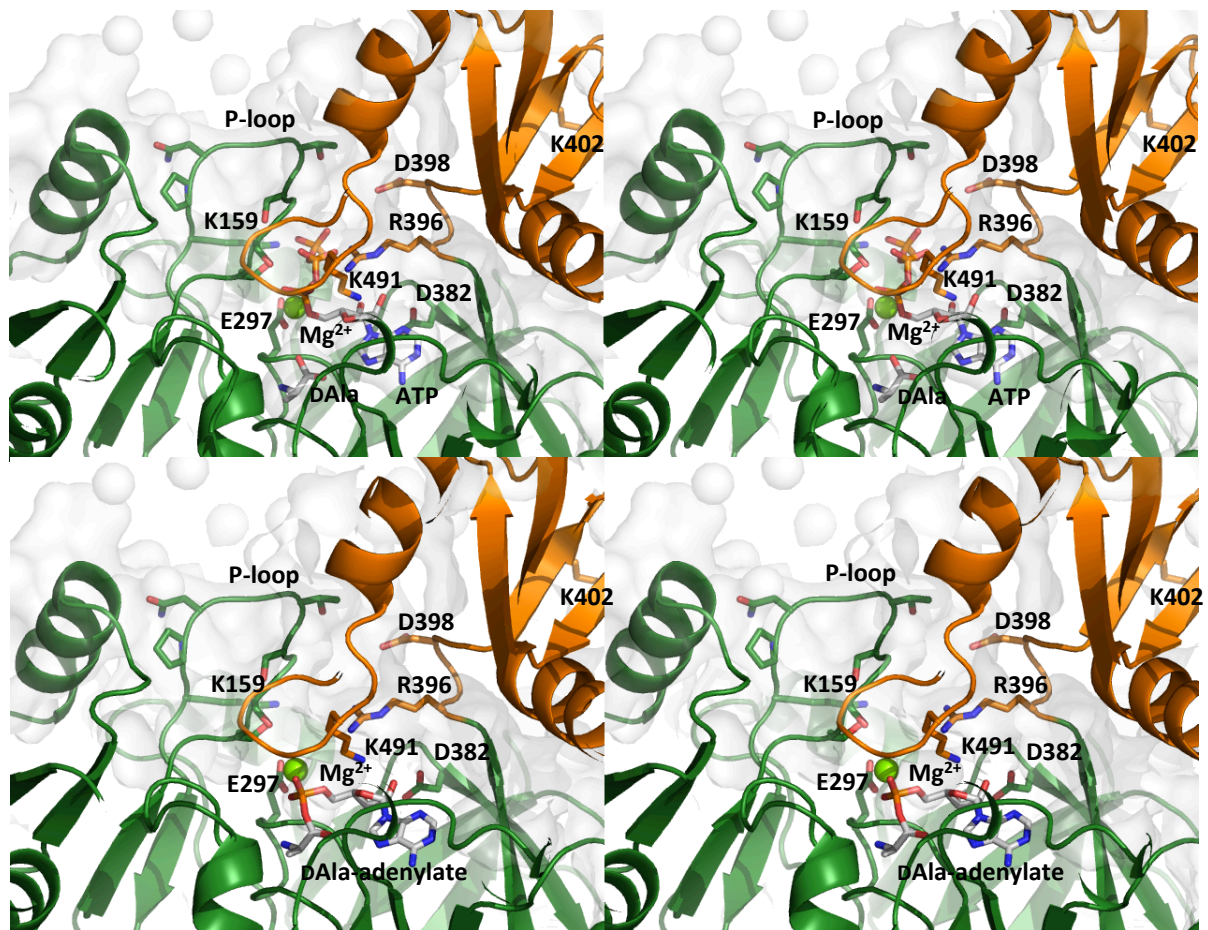
A DAla substrate has been modelled into the active site. The invariant hinge residue, R396 interacts with two highly conserved, oppositely charged residues of the small domain i.e. E414 and E411.

### 6.3.2 Adenylation conformation

Such a state should be attained upon binding of the substrates i.e. DAla,  $Mg^{2+}$  and ATP. The triphosphate moiety of ATP is expected to bind in an extended conformation to the side chain of highly conserved K159 as well as the backbone amides of the P-loop (Fig. 37). This extended mode of ATP binding had been suggested previously in one crystal form of DhbE (1md9) where the ligand was proposed to be an ATP molecule with a disordered pyrophosphate (May et al., 2002). The subsequently determined crystal structures of *B. cereus* DltA (Osman et al., 2009) and a human medium-chain ACS in complex with ATP (ACSM2A) (Kochan et al., 2009) further confirmed the proposed extended mode of ATP binding. On the other hand, a different binding mode is observed in LC-FACS (Hisanaga et al., 2004), where the AMP-PNP doesn't engage in any interactions with the highly conserved P-loop well known for ATP and GTP binding proteins. Moreover, the nucleotide fails to utilise any conserved residues of the A-domains, and the  $\beta$  and  $\gamma$  phosphates aren't oriented

properly for an in-line attack of the fatty acid substrate and the  $\alpha$  phosphate of ATP. This might be because the non-hydrolysable ATP analogue AMP-PNP may be bound in an unproductive manner.

It is plausible that the highly negatively charge of ATP may attract the basic side chain of hinge residue, R396, weakening interactions between the small domain and the P-loop. The P-loop undergoes a conformational change (remodelled), which should allow the rotation of the small domain towards the ‘adenylation conformation’. As a consequence of this movement the active site is sealed from the bulk of the solvent and catalytic K491 is brought into proximity of the scissile  $\alpha$ - $\beta$  phosphate bond, opposite to the  $Mg^{2+}$  ion bound to E297. Cleavage of the phosphoester bond with concomitant DAla-adenylate formation and pyrophosphate release would lead to a further reorganisation of the P-loop.



**Fig. 37. DltA structure modelled into the ‘adenylation conformation’.**

(A) This model represents the “pre-adenylation” stage of the adenylation reaction i.e. when ATP and DAla substrates bind. The ATP molecule binds in an extended conformation forming interactions with the P-loop. R396 is also seen in an extended conformation and may form hydrogen bonds with the  $\beta$ -phosphate of ATP thus facilitating rotation of the small domain towards the adenylation conformation. B) This model represents the “post-adenylation” stage where the positioning of K491 close to the  $\alpha$ - $\beta$  phosphate bond of ATP should facilitate the adenylation reaction i.e. cleavage of the phosphoester bond with the subsequent formation of the DAla-adenylate intermediate and pyrophosphate release

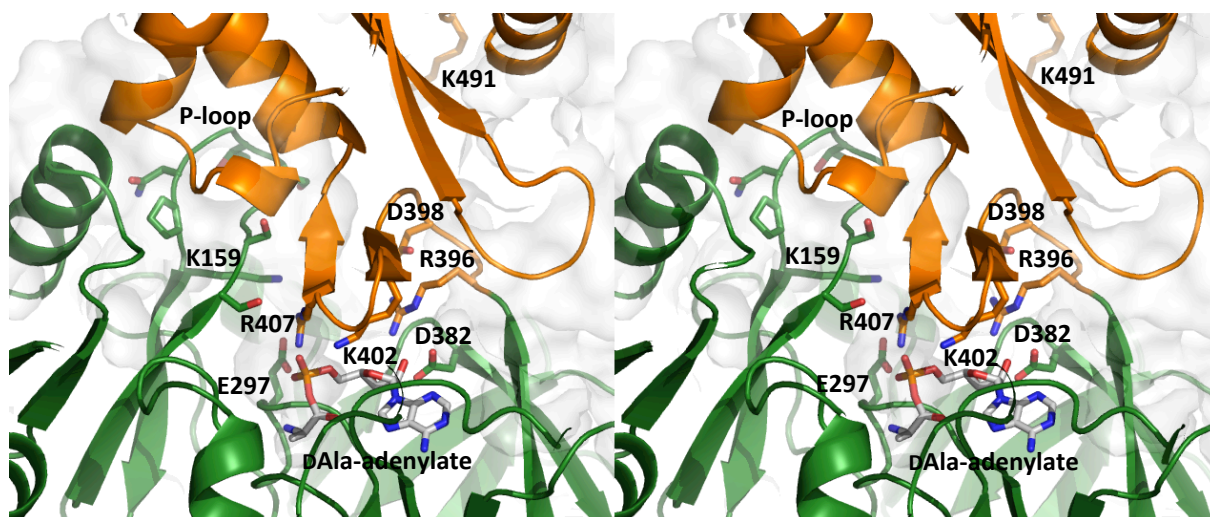
The determined ATP bound structures (Kochan et al., 2009; Osman et al., 2009) confirmed DltA modelling studies where the location of the ATP molecule as well as the orientation of the P-loop and hinge residue R396 are almost identical as predicted. Both structures displayed the ‘adenylation’ conformation when bound to ATP and revealed for the first time interactions made with both the  $\beta$  and  $\gamma$ -phosphate of ATP. In the human medium chain ACS (Kochan et al., 2009) the P-loop is highly ordered and lies in close proximity to the bound ATP thus shielding the bound ATP from solvent. This loop forms at least one hydrogen bond with each phosphate group of ATP including a direct hydrogen bond between the strictly invariant K159 (DltA numbering) and the  $\gamma$ -phosphate of ATP. The hinge residue R396 is also seen in an extended conformation and its guanidinyll group forms bidentate hydrogen bonds with the  $\beta$ -phosphate of ATP in both these structures (Kochan et al., 2009; Osman et al., 2009). Such an interaction may be one of the driving forces leading to the ‘adenylation’ state rearrangement. Interestingly K491 forms a hydrogen bond with the  $\beta$ -phosphate of ATP while in the adenylate complex structures this residue is hydrogen bonded to the carbonyl group of the substrate and/or the  $\alpha$ -phosphate of AMP. These two conformations of K491 probably play a significant role in the catalysis of the adenylation half reaction as will be discussed later. The  $Mg^{2+}$  also coordinates to two oxygens from the  $\beta$ - and  $\gamma$ -phosphate of ATP and to four water molecules, two of which interact with the highly conserved E297.

### 6.3.3 Thiolation conformation

After the cleavage of PPi, the loss of negative charge could allow the side chain of R396 to rotate towards D382. The carboxylate of hinge residue D398 would approach the basic side chain of R396, facilitating rotation of the small domain to the ‘thiolation conformation’ (Fig. 38).

The thiolation conformation reconfigures the active site for thiol transfer by correct positioning of K402. K402 is conserved in DltA, PheA, Luciferase as well as LC-FACS. In LC-FACS, the location of this residue close to the carbonyl carbon of the substrate suggested its possible role in catalysis of the thiolation reaction (Hisanaga et al., 2004). Other studies (Wu et al., 2008) suggest that the electropositive N-terminus of  $\alpha$ -helix (D220-S232) directed into the reaction centre may also reduce the pKa of the thiol group thus increasing its nucleophilicity for attack on the carboxylate in the adenylate intermediate.

A salt bridge formed between R407 (A8 loop) and E297 should effectively displace the active site  $Mg^{2+}$  ion and thus prevents the adenylation reaction (or the back reaction) from taking place in this conformation. The crystal structure of ACS (Gulick et al., 2003) as well as ACSM2A (Kochan et al., 2009) demonstrated a similar salt bridge interaction. E297 is a strictly invariant residue of the A5 motif, while R407 is conserved among the A-domains with PheA bearing an equivalent lysine residue at this position (Conti et al., 1997). This salt bridge interaction thus seems crucial for stabilizing the thiolation conformation. In the case for LC-FACS, the residue equivalent to R407 is a tryptophan, therefore it is possible that the adenylation reaction can still occur in the ‘thiolation conformation’ as proposed (Hisanaga et al., 2004). Once the activated amino acid is transferred to the phosphopantheynyl moiety of DltC by an as yet unexplained mechanism, AMP would dissociate and DltA would return to the open state, ready for binding the next round of substrates.



**Fig. 38. DltA structure modelled into the ‘thiolation conformation’.**

Hinge residue, R396 rotates towards strictly invariant, D382, while D398 approaches the basic side chain of R396. These movements facilitate the rotation of the small domain towards the ‘thiolation conformation’. Close positioning of K402 in proximity to the DAla-adenylate seems crucial for catalysis of the thiolation reaction.

#### 6.4 Domain reorganisation as a dynamic feature of the superfamily

DltA has now been crystallized in both the adenylation (Du et al., 2008; Osman et al., 2009) as well as thiolation conformation (Yonus et al., 2008). Most of the NRPS A-domains so far have been crystallized in the adenylation conformation state. The recent structural determination of a complete NRPS module i.e. SrfA-C revealed the A-domain in the open conformation (Tanovic et al). Structures from members of the acyl-CoA synthase sub-family have also been crystallized in both conformational states, including ACS (Gulick et al., 2003; Jogl and Tong, 2004), 4-CBAL (Reger et al., 2008; Wu et al., 2008) and Human medium



chain acyl-coenzyme A synthetase (ACSM2A)(Kochan et al., 2009). All these structures thus suggest that domain rotation is a dynamic feature in these enzymes and is not due to differences in the organisation of tertiary structure among different enzymes. DltA displayed a thiolation conformation in spite of the fact that it was crystallized in the absence of any thiol acceptor like DltC or CoA. This observation is in contrast to the earlier studies (Gulick et al., 2003; Jogl and Tong, 2004) where the binding of CoA has been proposed to be the trigger for attaining the thiolation conformation. The possibility of crystal contacts influencing the observed conformation is not likely as both crystal forms DltA (I222 and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>), displayed an identical conformation. Human ACSM2A also crystallized in multiple space groups, allowing both conformations to be observed in more than one crystal form. Thus it seems that it is the protein conformation that dictates its crystallization rather than vice versa.

Among the different structures of members of the superfamily, a few of them have shown deviations from the general rule (where the adenylation conformation is attained with binding of ATP, AMP or adenylate and thiolation conformation by binding of CoA). These structures include: DhbE and CBAL, where an adenylation state was seen even in the absence of any substrate (Gulick et al., 2004; May et al., 2002); LC-FACS, which displayed a thiolation conformation in the presence of an ATP analogue (AMPPNP) (Hisanaga et al., 2004) and human ACSM2A where a thiolation state is seen in the unliganded state, with AMP as well as a ternary product complex with butyryl-coA and AMP (Kochan et al., 2009). Our DltA modelling studies as well as the existing structures suggests that these enzymes may exist in diverse metastable states where only small energy differences separate the individual states. In this way, moderate changes at the active site brought about by substrate binding and product formation along with concerted movements of the hinge region and P-loop would allow re-organisation of the large and small domains, thereby driving the enzyme through the reaction cycle. Considering the above mentioned exceptions (DhbE and ACSM2A), it is also possible that in the absence of substrates each protein has a preferred state where it is more likely to crystallize as proposed by Reger *et al* (Reger et al., 2008).

The structure of human ACSM2A determined in a series of ligand bound complexes demonstrated substantial rearrangements between the two domains depending upon the identity of the bound ligand at the active site. The enzyme adopts the adenylation conformation when ATP or its nonhydrolysable homolog AMPCPP are bound, however in

the presence of AMP, substrate ibuprofen as well as the ternary product complex butyryl-CoA and AMP, the protein adopts a thiolation state. DltA (Osman et al., 2009) bound to ATP also displayed an adenylation conformation. These studies as well as our DltA structure and modelling studies clearly point to the involvement of PPi in the conformational switch. Presence of PPi in the form of ATP should induce closure of the large and small domain thereby leading to the adenylation conformation. The cleavage of PPi from ATP during the adenylation half reaction should trigger the switch to the thiolation conformation. Binding of CoA does not seem to be involved in the switch to the thiolation state (as proposed by (Gulick et al., 2003; Jogl and Tong, 2004)), supported by the structure of human ACSM2A bound to CoA (3GPC.pdb) where the protein adopts an adenylation state (Kochan et al., 2009). Although CoA binds non-productively, occupying the ATP/AMP nucleotide subsite, the enzyme still adopts an adenylation conformation. Examining the structures carefully reveals that pantetheine moiety of CoA (or of a PCP domain) would be obstructed by an A4 aromatic residue (F195, DltA numbering) in the adenylation conformation. In the thioester forming conformation, the side chain of this residue must rotate in order to prevent steric clashes with the incoming pantetheine group. Thus functional CoA binding can only occur after adenylate intermediate formation (cleavage of PPi) and only with the thioester conformation. Likewise ATP cannot bind to the thioester forming conformation because of the interaction between its  $\beta,\gamma$ -phosphates and the R407 and E297 salt bridge.

#### **6.4.1 Role of Hinge region and P-loop**

Hinge residue R396 is a completely conserved residue belonging to the A8 motif. This residue is seen in multiple orientations correlating to the different observed conformations, thus acting like an electrostatic conformational switch. In the open conformation, this residue is directed out of the active site to interact with 2 highly conserved C-terminal domain residues. Such an interaction may help in maintaining and stabilizing the open conformation. With the binding of ATP, DltA modelling studies predicted that this residue will adopt an extended conformation and form bidentate hydrogen bonds with the  $\beta$ -phosphate of ATP, which has since been demonstrated by the ATP-bound structures of human ACSM2A (Kochan et al., 2009) and DltA (Osman et al., 2009). Such an interaction with  $\beta$ -phosphate of ATP helps in neutralizing the negative charge of ATP as well as stabilizing the leaving pyrophosphate group. In the presence of adenylate or AMP, R396 is seen to hydrogen bond with the ribose hydroxyl groups (Conti et al., 1997; May et al., 2002). Such an interaction may help in

anchoring the adenylate for the second half reaction. In the thiolation conformation, R396 forms an ion pair with linker residue D398, which should help in stabilizing this conformation (Gulick et al., 2003; Reger et al., 2008; Yonus et al., 2008). Indeed the role of this residue in both half reactions has been supported by mutation studies (Wu et al., 2008). Another conserved residue of the hinge region is D398. This residue is most commonly an aspartate but some members of the family also bear a lysine at this position. This residue undergoes main chain torsion angle rotations that correlate with the differences between the two conformations. Mutation of this residue to a proline in 4CBAL allows the enzyme to adopt the adenylation conformation but not the thiolation conformation (Wu et al., 2009).

The P-loop, which is a signature sequence of the superfamily, has been hypothesized to be involved in ATP binding due to its similarity in amino acid composition to that of P-loops in ATPases and GTPases (Walker et al., 1982). This loop is often seen disordered. When ordered, the P-loop is seen in one of two different orientations relevant to the two conformational states. In the thiolation conformation, the P-loop makes close contacts to residues in the small domain and is stabilized by hydrogen bonds to absolutely conserved E412. Such an interaction may also stabilize this conformation. It also packs against the  $\beta$ -hairpin in the C-terminal domain harbouring the strictly conserved G405.

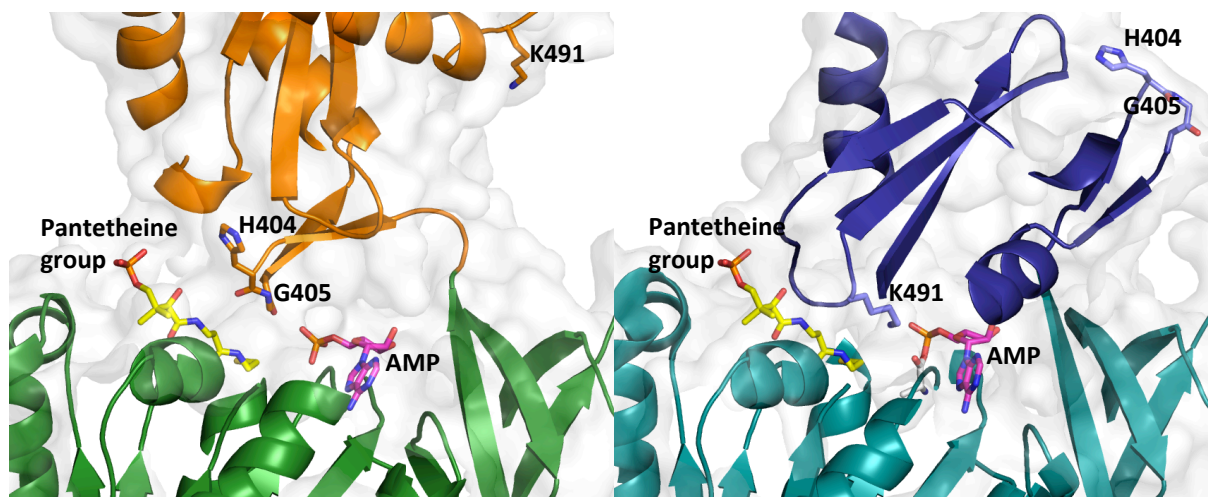
In the adenylation conformation (with bound ATP or adenylate), the P-loop is located distal from the nucleotide subsite towards the C-terminal helix of the small domain. This creates a pyrophosphate pocket and in the presence of ATP, residues T151, G153, T155 and K159 (DltA numbering) of the P-loop engage in hydrogen bonds with the  $\gamma$ -phosphate of ATP (Kochan et al., 2009). Such a wrapping mode of P-loop completely buries the ATP co-substrate and shields it from solvent. In the presence of adenylate, the P-loop interacts with the  $\alpha$ -phosphate of ATP. Such interactions of the P-loop may thus play an important role in the proper orientation of phosphate groups of ATP. The two different orientations of the P-loop might also have a role in displacement of PPi.

## **6.5 Role of domain alternation in catalysis**

Both structural and kinetic studies can now provide insight into the role of domain movement in catalysis. Domain rotation brings the two catalytically important residues into the active

site i.e. K491 and G405 (Fig. 39). K491 is a universally conserved residue of the A10 region and its role in the adenylation half reaction has been demonstrated both structurally as well as kinetically (Branchini et al., 2000; Horswill and Escalante-Semerena, 2002; Wu et al., 2008). In earlier structures this residue has been seen to engage in hydrogen bonds either with the  $\alpha$ -phosphate of AMP or/and with the bridging oxygen between the ribose and the phosphate or with the carboxylate oxygen of the substrate (Conti et al., 1997; May et al., 2002). However in the presence of ATP, K491 hydrogen bonds to the  $\beta$ -phosphate of ATP (Kochan et al., 2009; Osman et al., 2009). These different interactions of K491 illustrate its role in catalysis of the first half reaction. Its interaction with  $\beta$ -phosphate of ATP suggests that this residue helps proper positioning of the ATP phosphoryl groups for nucleophilic attack with the substrate carboxylate group. At the beginning of the reaction, the positioning of K491 between the carboxyl group of substrate and  $\alpha$ -phosphate of ATP decreases electrostatic repulsion between these two charged groups thus facilitating their approach. As the reaction proceeds K491 relocates to the  $\alpha$ -phosphate to stabilize the transition state as this group would be most negatively charged. When reaction is near completion, negative charge would accumulate at the  $\beta$ -phosphate of ATP, thus K491 relocates to the  $\beta$ -phosphate, this would favor the departure of the leaving pyrophosphate group.

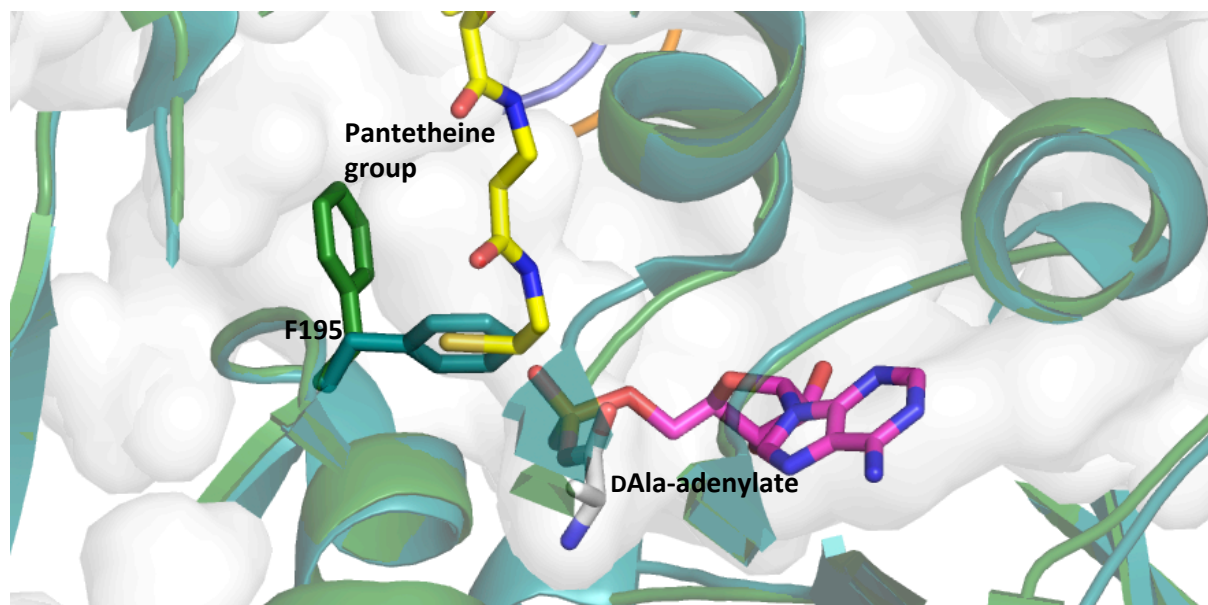
G405 is also an absolutely conserved residue of the A8 loop. In ACS, CBAL and human ACSM2A this residue as well as the residue preceding it hydrogen bonds to the two amide nitrogens of the phosphopantethiene moiety of CoA. Such an interaction further constricts the pantethiene channel thus directing the thiol group into the vicinity of the acyl-adenylate intermediate for the thiolation reaction. Mutation of G405 to a leucine in ACS (Reger et al., 2007) showed the mutant to be only active for the adenylation reaction thus emphasizing its role in the thiolation reaction. Further domain rotation brings the A8 loop into the active site thus creating a pantethiene tunnel.



**Fig. 39. Domain rotation brings two catalytic residues into the active site.**

(a) DltA in the thiolation conformation (Yonus et al., 2008), G405 and H404 of the A8 loop forms hydrogen bonds with the amide nitrogens of the pantetheine group of CoA (modelled from the ACS structure) while K491 is located far from the active site. b) DltA in the adenylation conformation (Du et al., 2008), (small domain is shown in blue while large domain in cyan) K491 of the small domain bridges the nucleophilic attack of a carboxylate group on the  $\alpha$ -phosphate of ATP while the residue of the A8 loop are located away from the active site as a consequence of domain rotation.

Domain rotation may also play a role in providing suitable environments for catalysis of both partial reactions. Critical examination of all the structures of members determined so far show a moderately conserved aromatic residue belonging to the A4 motif (F195 in DltA) in two significant side chain torsion angle rotations associated with the two conformations. In the adenylation conformation, this residue lies close to the carboxyl carbon of the substrate while in the thiolation conformation, this residue rotates out of the active site (Fig. 40). Structural as well as biochemical studies now explain this observation and also raise another possible role domain alternation plays in catalysis (Gulick, 2009; Wu et al., 2008). In the first half reaction (adenylation conformation) its placement close to the carboxyl carbon of the substrate narrows the binding pocket thus properly positioning the carboxylate for nucleophilic attack on the  $\alpha$ -phosphate of ATP. However this feature becomes an obstacle in the second half reaction (thiolation reaction) where this residue would sterically clash with the incoming pantethiene group. Domain movement brings the A8 loop into the active site that provides an environment to promote rotation of the A4 aromatic residue out of the active site thus creating a pantethiene channel whereby the pantethiene group can access the carboxylate carbon of substrate. Proteins that bear a histidine at position A4 usually contain a glutamic acid on the A8 loop which would promote hydrogen bond formation, while proteins with a phenylalanine or a tryptophan at position A4 usually bear a tyrosine or histidine on the A8 loop thus favouring hydrophobic interactions.



**Fig. 40.** F195 of DltA is seen in two side chain torsion angle rotations; the ‘down’ position (cyan) seen in the adenylation state obstructs the pantetheine group while the ‘up’ position (green) seen in thiolation state creates a pantetheine tunnel.

## 7 Afterword

Re-engineering NRPSs by module reshuffling has been a fascinating aspect of this field, but due to lack of a complete understanding of substrate specificity and transfer within an assembly line, limited success has been obtained in this regard (Strieker et al., 2010). In addition, the large number of conformational states available for NRPSs calls for a detailed understanding of the interaction interfaces accessible to the various domains. Since completion of the experimental work described in this thesis, significant progress has been made in this respect (Pfennig and Stubbs, 2012).

The structure of the complete NRPS module SrfA-C provided valuable insights into interdomain architecture (Tanovic et al., 2008). This structure demonstrated that the condensation domain and the N-terminal subdomain of the adenylation domain engage closely to form a catalytic platform on which the other domains and subdomains migrate to enable the transfer of the substrate to different active sites. Modeling of the pantetheine cofactor suggested that the carrier protein of SrfA-C was stalled to interact with the acceptor site of the condensation domain. However, significant domain reorganizations are necessary for the carrier protein domain to deliver the pantetheine thiol group to the other domains.

In order to trap the PCP-A domain interface, a chimeric construct was designed by fusing the stand-alone aryl acid activating domain EntE to the PCP-domain of EntB with a four-residue domain linker in between (Sundlov et al., 2012). The two domains were also linked together so as to mimic the intramolecular interaction seen in natural A-PCP proteins. In addition, a vinylsulfonamide mechanism based inhibitor, which reacts to result in formation of a covalent DHB-adenylate-thioether, was used to mimic the approach of the PCP-ppant arm. The structure surprisingly crystallized as a dimer with intermolecular interactions between the A-domain of one monomer and the PCP domain of the second, i.e. a trans-interaction. As expected, the A-domain displayed the thiolation conformation, with extensive contacts between PCP helices I to III and both A-domain subdomains. Although an intramolecular interaction (cis-interaction) was not possible in that crystal structure, that of a naturally occurring two-domain protein, PA1221 from *Pseudomonas aeruginosa*, appears to demonstrate the anticipated intramolecular interaction (Mitchell et al., 2012). The holo-protein, crystallized in the presence a vinylsulfonamide inhibitor, also displayed a thiolation conformation, with the PCP in a similar position relative to the A-domain. As the linker between the two domains was disordered, it was not clear whether the two domains formed inter or intramolecular interaction. Detailed kinetic analyzes with wild type and mutant

proteins suggested a preference of the A-domain to load the fused PCP intramolecularly (within a single protein chain), though the possibility of intermolecular bonding could not be ruled out. Comparison of the PA1221 structure with the EntE-B interface revealed that their main chain conformations are remarkably similar except for the C-terminal helix of the adenylation domain. This helix undergoes unfolding (melting) to facilitate a cis-transfer as postulated by Pfennig *et al* (Pfennig and Stubbs, 2012). Comparison of both structures with the complete module of SrfA-C (Tanovic *et al.*, 2008) provides a better understanding into the molecular choreography of NRPSs. The PCP domain of SrfA-C was properly positioned to interact with the condensation domain though it was too far to reach the alternate catalytic domains including the A-domain, thus the SrfA-C structure illustrated one conformation of an NRPS module. Substantial domain movements were required for the PCP to reach other catalytic domains. The EntE-B and PA1221 structures illustrated yet another conformation whereby following amino acyl adenylation, the PCP is delivered to the adenylation domain through a rotation of the C-terminal subdomain of the A-domain. Completion of thioesterification reaction and release of the loaded cofactor allows return to the open state as seen in SrfA-C enabling delivery of the substrate to the upstream condensation domain. On the basis on these studies it was proposed the domain alternation conformational change within adenylation domains of NRPSs also helps in the transport of the PCP domain for distinct steps in the NRPS catalytic cycle (Sundlov *et al.*, 2012).

The solution structure of an apo-PCP–thioesterase (T–TE) di-domain fragment of the *Escherichia coli* enterobactin synthetase, EntF subunit provided an understanding of the structural basis of communication between the PCP and TE domains in chain-release steps at the end of NRPS assembly lines (Frueh *et al.*, 2008). The globular T domain (PCP) forms the characteristic three-helix bundle that is wedged between the globular core of the thioesterase and two  $\alpha$ -helices protruding from this core. Liu *et al* subsequently used an  $\alpha$ -chloroacetylamino CoA analog to obtain a locked EntF-derived holo-PCP-TE di-domain (Liu *et al.*, 2011). When compared with the SrfA-C, the distance between the active sites of the PCP and TE domain in the SrfA-C structure (43Å) is much larger than that in PCP-TE complex (16Å). This is consistent with the fact that the PCP in the SrfA-C was in the position ready to interact with the condensation domain while in the holo-PCP-TE complex, the PCP interacts with the TE-domain.

Our DltA structure is in agreement with these studies, it demonstrated a rotation of the C-terminal domain into the thiolation conformation. Besides DltA has also been crystallized in the adenylation conformation (Du *et al.*, 2008) thus suggesting the existence of both states



among the NRPS adenylation domains. The later structures of EntE-B (Sundlov et al., 2012) and PA1221 (Mitchell et al., 2012) in the thiolation conformation further confirm our studies. All these studies demonstrate the importance of domain arrangement and conformational flexibility in the production of nonribosomal peptides. More structures of NRPSs complexes especially adenylation-PCP domains and further analysis of the structures should provide a better understanding of the interaction interface and all the steps in the choreography of the NRPS catalytic cycle.

## 8 References

- Abachin, E., Poyart, C., Pellegrini, E., Milohanic, E., Fiedler, F., Berche, P., and Trieu-Cuot, P. (2002). Formation of D-alanyl-lipoteichoic acid is required for adhesion and virulence of *Listeria monocytogenes*. *Molecular Microbiology* *43*, 1-14.
- Araki, Y., and Ito, E. (1989). Linkage Units in Cell-Walls of Gram-Positive Bacteria. *Crit Rev Microbiol* *17*, 121-135.
- Ayabe, K., Zako, T., and Ueda, H. (2005). The role of firefly luciferase C-terminal domain in efficient coupling of adenylation and oxidative steps. *FEBS Lett* *579*, 4389-4394.
- Baddiley, J. (1962). Teichoic Acids in Walls and Cells of Gram-Positive Bacteria. *Fed Proc* *21*, 1084-&.
- Baddiley, J., and Neuhaus, F.C. (1960). Enzymic Activation of D-Alanine. *Biochem J* *75*, 579-587.
- Black, P.N., and DiRusso, C.C. (1994). Molecular and biochemical analyses of fatty acid transport, metabolism, and gene regulation in *Escherichia coli*. *Biochim Biophys Acta* *1210*, 123-145.
- Boyd, D.A., Cvitkovitch, D.G., Bleiweis, A.S., Kiriukhin, M.Y., Debabov, D.V., Neuhaus, F.C., and Hamilton, I.R. (2000). Defects in D-alanyl-lipoteichoic acid synthesis in *Streptococcus mutans* results in acid sensitivity. *J Bacteriol* *182*, 6055-6065.
- Branchini, B.R., Murtiashaw, M.H., Magyar, R.A., and Anderson, S.M. (2000). The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase. *Biochemistry* *39*, 5433-5440.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* *54*, 905-921.
- Byford, M.F., Baldwin, J.E., Shiao, C.Y., and Schofield, C.J. (1997). The Mechanism of ACV Synthetase. *Chem Rev* *97*, 2631-2650.
- Challis, G.L., and Naismith, J.H. (2004). Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol* *14*, 748-756.
- Chang, K.H., Xiang, H., and Dunaway-Mariano, D. (1997). Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the *Pseudomonas* sp. strain CBS3 4-chlorobenzoate:coenzyme A ligase. *Biochemistry* *36*, 15650-15659.
- Clemans, D.L., Kolenbrander, P.E., Debabov, D.V., Zhang, Q.Y., Lunsford, R.D., Sakone, H., Whittaker, C.J., Heaton, M.P., and Neuhaus, F.C. (1999). Insertional inactivation of genes

responsible for the D-alanylation of lipoteichoic acid in *Streptococcus gordonii* DL1 (Challis) affects intrageneric coaggregations. *Infect Immun* 67, 2464-2474.

COLLABORATIVE COMPUTATIONAL PROJECT, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760-763.

Conti, E., Franks, N.P., and Brick, P. (1996). Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure* 4, 287-298.

Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *Embo J* 16, 4174-4183.

Cowie, D.B., and Cohen, G.N. (1957). Biosynthesis by *Escherichia coli* of active altered proteins containing selenium instead of sulfur. *Biochim Biophys Acta* 26, 252-261.

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

Davie, J.M., and Brock, T.D. (1966). Effect of Teichoic Acid on Resistance to Membrane-Lytic Agent of *Streptococcus Zymogenes*. *J Bacteriol* 92, 1623-&.

Debabov, D.V., Heaton, M.P., Zhang, Q.Y., Stewart, K.D., Lambalot, R.H., and Neuhaus, F.C. (1996). The D-alanyl carrier protein in *Lactobacillus casei*: Cloning, sequencing, and expression of *dltC*. *J Bacteriol* 178, 3869-3876.

Debabov, D.V., Kiriukhin, M.Y., and Neuhaus, F.C. (2000). Biosynthesis of lipoteichoic acid in *Lactobacillus rhamnosus*: Role of DltD in D-alanylation. *J Bacteriol* 182, 2855-2864.

Dieckmann, R., Lee, Y.O., van Liempt, H., von Dohren, H., and Kleinkauf, H. (1995). Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. *FEBS Lett* 357, 212-216.

Doublet, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol* 276, 523-530.

Drake, E.J., Nicolai, D.A., and Gulick, A.M. (2006). Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chem Biol* 13, 409-419.

Du, L., He, Y., and Luo, Y. (2008). Crystal structure and enantiomer selection by D-alanyl carrier protein ligase DltA from *Bacillus cereus*. *Biochemistry* 47, 11473-11480.

Duitman, E.H., Hamoen, L.W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F., Reinhardt, R., Schmidt, M., Ullrich, C., *et al.* (1999). The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc Natl Acad Sci U S A* 96, 13294-13299.

Eriani, G., Dirheimer, G., and Gangloff, J. (1990). Aspartyl-tRNA synthetase from *Escherichia coli*: cloning and characterisation of the gene, homologies of its translated amino

acid sequence with asparaginyl- and lysyl-tRNA synthetases. *Nucleic Acids Res* 18, 7109-7118.

Faulkner, D.J. (1998). Marine natural products. *Nat Prod Rep* 15, 113-158.

Finking, R., and Marahiel, M.A. (2004). Biosynthesis of nonribosomal peptides1. *Annu Rev Microbiol* 58, 453-488.

Finking, R., Neumuller, A., Solsbacher, J., Konz, D., Kretzschmar, G., Schweitzer, M., Krumm, T., and Marahiel, M.A. (2003). Aminoacyl adenylate substrate analogues for the inhibition of adenylation domains of nonribosomal peptide synthetases. *Chembiochem* 4, 903-906.

Franks, N.P., Jenkins, A., Conti, E., Lieb, W.R., and Brick, P. (1998). Structural basis for the inhibition of firefly luciferase by a general anesthetic. *Biophys J* 75, 2205-2211.

Frueh, D.P., Arthanari, H., Koglin, A., Vosburg, D.A., Bennett, A.E., Walsh, C.T., and Wagner, G. (2008). Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* 454, 903-906.

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.

Gross, M., Cramton, S.E., Gotz, F., and Peschel, A. (2001). Key role of teichoic acid net charge in *Staphylococcus aureus* colonization of artificial surfaces. *Infect Immun* 69, 3423-3426.

Gulick, A.M. (2009). Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem Biol* 4, 811-827.

Gulick, A.M., Lu, X.F., and Dunaway-Mariano, D. (2004). Crystal structure of 4-chlorobenzoate : CoA ligase/synthetase in the unliganded and aryl substrate-bound states. *Biochemistry* 43, 8670-8679.

Gulick, A.M., Starai, V.J., Horswill, A.R., Homick, K.M., and Escalante-Semerena, J.C. (2003). The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry* 42, 2866-2873.

Haslinger, K., Redfield, C., and Cryle, M.J. (2015). Structure of the terminal PCP domain of the non-ribosomal peptide synthetase in teicoplanin biosynthesis. *Proteins* 83, 711-721.

Heaton, M.P., and Neuhaus, F.C. (1994). Role of the D-Alanyl Carrier Protein in the Biosynthesis of D-Alanyl-Lipoteichoic Acid. *J Bacteriol* 176, 681-690.

Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *Embo J* 9, 1665-1672.

- Herzberg, O., and Moulton, J. (1991). Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins* *11*, 223-229.
- Hisanaga, Y., Ago, H., Nakagawa, N., Hamada, K., Ida, K., Yamamoto, M., Hori, T., Arii, Y., Sugahara, M., Kuramitsu, S., *et al.* (2004). Structural basis of the substrate-specific two-step catalysis of long chain fatty acyl-CoA synthetase dimer. *J Biol Chem* *279*, 31717-31726.
- Holak, T.A., Kearsley, S.K., Kim, Y., and Prestegard, J.H. (1988). Three-dimensional structure of acyl carrier protein determined by NMR pseudoenergy and distance geometry calculations. *Biochemistry* *27*, 6135-6142.
- Hori, K., Yamamoto, Y., Minetoki, T., Kurotsu, T., Kanda, M., Miura, S., Okamura, K., Furuyama, J., and Saito, Y. (1989). Molecular cloning and nucleotide sequence of the gramicidin S synthetase 1 gene. *J Biochem (Tokyo)* *106*, 639-645.
- Horswill, A.R., and Escalante-Semerena, J.C. (2002). Characterization of the propionyl-CoA synthetase (PrpE) enzyme of *Salmonella enterica*: residue Lys592 is required for propionyl-AMP synthesis. *Biochemistry* *41*, 2379-2387.
- Hutchison, M.L., and Gross, D.C. (1997). Lipopeptide phytotoxins produced by *Pseudomonas syringae* pv. *syringae*: comparison of the biosurfactant and ion channel-forming activities of syringopeptin and syringomycin. *Mol Plant Microbe Interact* *10*, 347-354.
- Jancarik, J., and Kim, S.H. (1991). Sparse-Matrix Sampling - a Screening Method for Crystallization of Proteins. *J Appl Crystallogr* *24*, 409-411.
- Jogl, G., and Tong, L. (2004). Crystal structure of yeast acetyl-coenzyme A synthetase in complex with AMP. *Biochemistry* *43*, 1425-1431.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* *47 (Pt 2)*, 110-119.
- Joshi, A.K., Rangan, V.S., Witkowski, A., and Smith, S. (2003). Engineering of an active animal fatty acid synthase dimer with only one competent subunit. *Chem Biol* *10*, 169-173.
- Kabsch, W. (1993). Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J Appl Crystallogr* *26*, 795-800.
- Kasahara, T., and Kato, T. (2003). A new redox-cofactor vitamin for mammals. *Nature* *422*, 832-832.
- Keating, T.A., Marshall, C.G., and Walsh, C.T. (2000). Reconstitution and characterization of the *Vibrio cholerae* vibriobactin synthetase from VibB, VibE, VibF, and VibH. *Biochemistry* *39*, 15522-15530.
- 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. *Nat Struct Biol* *9*, 522-526.

- Kim, Y., and Prestegard, J.H. (1989). A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry* 28, 8792-8797.
- Kiriukhin, M.Y., and Neuhaus, F.C. (2001). D-alanylation of lipoteichoic acid: Role of the D-alanyl carrier protein in acylation. *J Bacteriol* 183, 2051-2058.
- Kleywegt, G.J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr D Biol Crystallogr* 52, 842-857.
- Kochan, G., Pilka, E.S., von Delft, F., Oppermann, U., and Yue, W.W. (2009). Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A. *J Mol Biol* 388, 997-1008.
- Koetsier, M.J., Jekel, P.A., Wijma, H.J., Bovenberg, R.A., and Janssen, D.B. (2011). Aminoacyl-coenzyme A synthesis catalyzed by a CoA ligase from *Penicillium chrysogenum*. *FEBS Lett* 585, 893-898.
- Koglin, A., Mofid, M.R., Lohr, F., Schafer, B., Rogov, V.V., Blum, M.M., Mittag, T., Marahiel, M.A., Bernhard, F., and Dotsch, V. (2006). Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* 312, 273-276.
- Konz, D., Klens, A., Schorgendorfer, K., and Marahiel, M.A. (1997). The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol* 4, 927-937.
- Koprivnjak, T., Peschel, A., Gelb, M.H., Liang, N.S., and Weiss, J.P. (2002). Role of charge properties of bacterial envelope in bactericidal action of human group IIA phospholipase A(2) against *Staphylococcus aureus*. *Journal of Biological Chemistry* 277, 47636-47644.
- Ladbury, J.E., and Chowdhry, B.Z. (1996). Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem Biol* 3, 791-801.
- 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.
- Linne, U., Schafer, A., Stubbs, M.T., and Marahiel, M.A. (2007). Aminoacyl-coenzyme A synthesis catalyzed by adenylation domains. *FEBS Lett* 581, 905-910.
- Liu, Y., Zheng, T., and Bruner, S.D. (2011). Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chemistry & biology* 18, 1482-1488.
- Lomakin, I.B., Xiong, Y., and Steitz, T.A. (2007). The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell* 129, 319-332.
- Maier, T., Leibundgut, M., and Ban, N. (2008). The crystal structure of a mammalian fatty acid synthase. *Science* 321, 1315-1322.

- Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev* 97, 2651-2674.
- 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.
- May, J.J., Finking, R., Wiegeshoff, F., Weber, T.T., Bandur, N., Koert, U., and Marahiel, M.A. (2005). Inhibition of the D-alanine : D-alanyl carrier protein ligase from *Bacillus subtilis* increases the bacterium's susceptibility to antibiotics that target the cell wall. *Febs J* 272, 2993-3003.
- 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. *Proc Natl Acad Sci U S A* 99, 12120-12125.
- May, J.J., Wendrich, T.M., and Marahiel, M.A. (2001). The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J Biol Chem* 276, 7209-7217.
- Mirelman, D., Beck, B.D., and Shaw, D.R.D. (1970). Location of D-Alanyl Ester in Ribitol Teichoic Acid of *Staphylococcus-Aureus*. *Biochem Bioph Res Co* 39, 712-&.
- 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.
- Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2002). Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem* 3, 490-504.
- Nakatsu, T., Ichiyama, S., Hiratake, J., Saldanha, A., Kobashi, N., Sakata, K., and Kato, H. (2006). Structural basis for the spectral difference in luciferase bioluminescence. *Nature* 440, 372-376.
- Neuhaus, F.C., and Baddiley, J. (2003). A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiol Mol Biol Rev* 67, 686-723.
- Neuhaus, F.C., Heaton, M.P., Debabov, D.V., and Zhang, Q.Y. (1996). The *dlt* operon in the biosynthesis of D-alanyl-lipoteichoic acid in *Lactobacillus casei*. *Microb Drug Resist* 2, 77-84.
- Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302, 205-217.
- Osman, K.T., Du, L., He, Y., and Luo, Y. (2009). Crystal structure of *Bacillus cereus* D-alanyl carrier protein ligase (DltA) in complex with ATP. *J Mol Biol* 388, 345-355.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol* 276, 307-326.

- Pai, E.F., Krengel, U., Petsko, G.A., Goody, R.S., Kabsch, W., and Wittinghofer, A. (1990). Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *Embo J* 9, 2351-2359.
- Parris, K.D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C.C., Seehra, J., and Somers, W.S. (2000). Crystal structures of substrate binding to *Bacillus subtilis* holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* 8, 883-895.
- Perego, M., Glaser, P., Minutello, A., Strauch, M.A., Leopold, K., and Fischer, W. (1995). Incorporation of D-Alanine into Lipoteichoic Acid and Wall Teichoic-Acid in *Bacillus-Subtilis* - Identification of Genes and Regulation. *Journal of Biological Chemistry* 270, 15598-15606.
- Peschel, A., Otto, M., Jack, R.W., Kalbacher, H., Jung, G., and Gotz, F. (1999). Inactivation of the *dlt* operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. *Journal of Biological Chemistry* 274, 8405-8410.
- Peschel, A., Vuong, C., Otto, M., and Gotz, F. (2000). The D-alanine residues of *Staphylococcus aureus* teichoic acids alter the susceptibility to vancomycin and the activity of autolytic enzymes. *Antimicrob Agents Ch* 44, 2845-2847.
- Peypoux, F., Bonmatin, J.M., and Wallach, J. (1999). Recent trends in the biochemistry of surfactin. *Appl Microbiol Biotechnol* 51, 553-563.
- Pfennig, S., and Stubbs, M.T. (2012). Flexing and stretching in nonribosomal Peptide synthetases. *Chemistry & biology* 19, 167-169.
- Poyart, C., Lamy, M.C., Boumaila, C., Fiedler, F., and Trieu-Cuot, P. (2001). Regulation of D-alanyl-lipoteichoic acid biosynthesis in *Streptococcus agalactiae* involves a novel two-component regulatory system. *J Bacteriol* 183, 6324-6334.
- Quadri, L.E.N., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., and Walsh, C.T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* 37, 1585-1595.
- Radaev, S., and Sun, P.D. (2002). Crystallization of protein-protein complexes. *J Appl Crystallogr* 35, 674-676.
- Ramakrishnan, C., and Ramachandran, G.N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys J* 5, 909-933.
- 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.
- Reger, A.S., Wu, R., Dunaway-Mariano, D., and Gulick, A.M. (2008). Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry* 47, 8016-8025.



- Richardt, A., Kemme, T., Wagner, S., Schwarzer, D., Marahiel, M.A., and Hovemann, B.T. (2003). Ebony, a novel nonribosomal peptide synthetase for beta-alanine conjugation with biogenic amines in *Drosophila*. *J Biol Chem* *278*, 41160-41166.
- Roujeinikova, A., Baldock, C., Simon, W.J., Gilroy, J., Baker, P.J., Stuitje, A.R., Rice, D.W., Slabas, A.R., and Rafferty, J.B. (2002). X-ray crystallographic studies on butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site. *Structure* *10*, 825-835.
- Saraste, M., Sibbald, P.R., and Wittinghofer, A. (1990). The P-Loop - a Common Motif in Atp-Binding and Gtp-Binding Proteins. *Trends Biochem Sci* *15*, 430-434.
- Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* *58*, 1772-1779.
- Schwarzer, D., Finking, R., and Marahiel, M.A. (2003). Nonribosomal peptides: from genes to products. *Nat Prod Rep* *20*, 275-287.
- Schwarzer, D., and Marahiel, M.A. (2001). Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* *88*, 93-101.
- Schweizer, E., and Hofmann, J. (2004). Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol Mol Biol Rev* *68*, 501-517.
- Sieber, S.A., and Marahiel, M.A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* *105*, 715-738.
- Smith, S., and Tsai, S.C. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep* *24*, 1041-1072.
- Smith, S., Witkowski, A., and Joshi, A.K. (2003). Structural and functional organization of the animal fatty acid synthase. *Prog Lipid Res* *42*, 289-317.
- Stachelhaus, T., and Marahiel, M.A. (1995). Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiol Lett* *125*, 3-14.
- Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* *6*, 493-505.
- 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.
- Starai, V.J., Celic, I., Cole, R.N., Boeke, J.D., and Escalante-Semerena, J.C. (2002). Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine. *Science* *298*, 2390-2392.
- Stein, T., Vater, J., Kruft, V., Otto, A., WittmannLiebold, B., Franke, P., Panico, M., McDowell, R., and Morris, H.R. (1996). The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *Journal of Biological Chemistry* *271*, 15428-15435.

Strieker, M., Tanovic, A., and Marahiel, M.A. (2010). Nonribosomal peptide synthetases: structures and dynamics. *Current opinion in structural biology* *20*, 234-240.

Sundlov, J.A., Shi, C., Wilson, D.J., Aldrich, C.C., and Gulick, A.M. (2012). Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chemistry & biology* *19*, 188-198.

Tang, G.L., Cheng, Y.Q., and Shen, B. (2007). Chain initiation in the leinamycin-producing hybrid nonribosomal peptide/polyketide synthetase from *Streptomyces atroolivaceus* S-140. Discrete, monofunctional adenylation enzyme and peptidyl carrier protein that directly load D-alanine. *J Biol Chem* *282*, 20273-20282.

Tang, Y., Kim, C.Y., Mathews, II, Cane, D.E., and Khosla, C. (2006). The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci U S A* *103*, 11124-11129.

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* *321*, 659-663.

Thern, B., Rudolph, J., and Jung, G. (2002). Total synthesis of the nematocidal cyclododecapeptide omphalotin A by using racemization-free triphosgene-mediated couplings in the solid phase. *Angew Chem Int Ed Engl* *41*, 2307-2309.

Tufar, P., Rahighi, S., Kraas, F.I., Kirchner, D.K., Lohr, F., Henrich, E., Kopke, J., Dikic, I., Guntert, P., Marahiel, M.A., *et al.* (2014). Crystal structure of a PCP/Sfp complex reveals the structural basis for carrier protein posttranslational modification. *Chemistry & biology* *21*, 552-562.

Turgay, K., Krause, M., and Marahiel, M.A. (1992). Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol* *6*, 529-546.

Ueda, H., Shoku, Y., Hayashi, N., Mitsunaga, J., In, Y., Doi, M., Inoue, M., and Ishida, T. (1991). X-ray crystallographic conformational study of 5'-O-[N-(L-alanyl)-sulfamoyl]adenosine, a substrate analogue for alanyl-tRNA synthetase. *Biochim Biophys Acta* *1080*, 126-134.

Volkman, B.F., Zhang, Q., Debabov, D.V., Rivera, E., Kresheck, G.C., and Neuhaus, F.C. (2001). Biosynthesis of D-alanyl-lipoteichoic acid: the tertiary structure of apo-D-alanyl carrier protein. *Biochemistry* *40*, 7964-7972.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J* *1*, 945-951.

Weber, P.C. (1997). Overview of protein crystallization methods. *Method Enzymol* *276*, 13-22.

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* *8*, 407-418.

- Weber, T., and Marahiel, M.A. (2001). Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure* 9, R3-9.
- Wecke, J., Madela, K., and Fischer, W. (1997). The absence of D-alanine from lipoteichoic acid and wall teichoic acid alters surface charge, enhances autolysis and increases susceptibility to methicillin in *Bacillus subtilis*. *Microbiol-Uk* 143, 2953-2960.
- White, S.W., Zheng, J., Zhang, Y.M., and Rock (2005). The structural biology of type II fatty acid biosynthesis. *Annu Rev Biochem* 74, 791-831.
- Wicken, A.J., and Baddiley, J. (1963). Structure of Intracellular Teichoic Acids from Group D Streptococci. *Biochem J* 87, 54-&.
- Wu, N., Tsuji, S.Y., Cane, D.E., and Khosla, C. (2001). Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J Am Chem Soc* 123, 6465-6474.
- Wu, R., Cao, J., Lu, X., Reger, A.S., Gulick, A.M., and Dunaway-Mariano, D. (2008). Mechanism of 4-chlorobenzoate:coenzyme a ligase catalysis. *Biochemistry* 47, 8026-8039.
- Wu, R., Reger, A.S., Lu, X., Gulick, A.M., and Dunaway-Mariano, D. (2009). The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate: coenzyme A ligase. *Biochemistry* 48, 4115-4125.
- 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. *J Biol Chem* 283, 32484-32491.
- Zimmermann, S. (2011). Strukturelle Untersuchungen der Wechselwirkungen zwischen DltA und DltC. PhD Thesis *Martin-Luther-Universität Halle-Wittenberg*.
- Zimmermann, S., Pfennig, S., Neumann, P., Yonus, H., Weininger, U., Kovermann, M., Balbach, J., and Stubbs, M.T. (2015). High-resolution structures of the D-alanyl carrier protein (Dcp) DltC from *Bacillus Subtilis* reveal equivalent conformations of apo- and holo-forms. *FEBS Lett Manuscript communicated*.

## 9 Appendix

### 9.1 Structure of DltC

Although a data set of holo-DltC crystals that grew from 0.1 M sodium acetate, 0.2 M ammonium acetate, 30 % PEG 4000, pH 5.7 was collected at the BESSY, Berlin (beam line BESSY-MX BL14.1), attempts at solving the structure by molecular replacement as well as multiple isomorphous replacement failed. Later the structure was solved in the work of Stephan Zimmermann (Zimmermann, 2011). Here a Ser36Ala apo-DltC mutant (lacking the ability to be phosphothienylated by Sfp) was produced as selenomethionine labelled protein for selenium MAD phasing. This refined apo-DltC structure was then used as a molecular replacement template for the structure solution of the above mentioned holo-DltC crystal (i.e. form I) as well as a second crystal form of holo-DltC (form II, crystallized by Stephan Zimmermann).

All the three DltC structures showed the helix-bundle fold characteristic of ACPs and PCPs. Most side chains are highly ordered except around the cofactor. Holo-DltC form I doesn't show density for the full cofactor due to high flexibility of this moiety, however form II displays well-defined density for the phosphothienine group. Dimerization could also be observed in both crystal forms of holo-DltC. However dimerization was likely as a result of an artefact from purification and crystallization. As mentioned in results (section 5.1.2), a putative dimer of holo-DltC could also be seen before the main peak of the monomeric DltC during gel filtration chromatography.

Interestingly all the three DltC structures adopted only one conformation ie AH-state (Koglin et al., 2006). Prior studies showed that carrier proteins can adopt multiple different conformations states. (Kim and Prestegard, 1989; Koglin et al., 2006). Our studies did not show any obvious structural differences between the carrier protein in its apo- or holo-forms (Zimmermann et al., 2015). Solution NMR spectroscopy additionally confirmed these findings and the two forms exhibit similar backbone dynamics. The absence of other states has been observed in other carrier protein structures (Drake et al., 2006; Haslinger et al., 2015; Tanovic et al., 2008; Tufar et al., 2014). These studies thus raise the question on the role of carrier protein plasticity in natural product biosynthesis.

Sequence and structural alignments of DltCs with ACPs and PCPs suggest that DltCs are closer to ACPs than PCPs and thus might have evolved from them. A close relationship between DltCs and ACPs is supported by the fact that the D-alanylation of LTA is highly

conserved throughout all gram-positive bacteria, indicating that DltA and DltC have evolved very early from proteins of the fatty-acid synthesis. Sequence alignments of several adenylate-forming enzymes additionally revealed, that DltA has a higher similarity to ACS than to the adenylation domains of NRPSs (Linne et al., 2007). This also points to a closer relationship between the proteins of the *dlt* operon and the fatty acid synthesis than *dlt* proteins and NRPSs. A similar pattern was observed in the phylogenetic comparison of DltA with other members of the adenylate forming family (section 6.2) where DltAs cluster closer to the acyl/aryl-CoA ligases while NRPS A-domains lie further away. The overall sequence of DltCs are more conserved to one another than when compared to the sequences of PCPs. Therefore, DltCs and PCPs do not seem as closely related as their common function i.e. the transport of amino acids. Whether DltCs evolved separately from PCPs or are their ancestor is unclear.

## 9.2 List of figures

**Fig 1:** Cell wall of gram-positive bacteria showing the locations of lipoteichoic acid (LTA) in green and wall teichoic acid (WTA) in blue

**Fig 2:** Monomers of teichoic acid showing the modifications with substituents

**Fig 3:** Structural diversity of non-ribosomally produced peptides

**Fig 4:** Reaction cycle of an NRPS showing synthesis of the siderophore, bacillibactin from *Bacillus subtilis*.

**Fig 5:** A dimodular PKS.

**Fig 6:** The reaction catalyzed by DltA proceeds in two steps

**Fig 7:** Model for the incorporation of DAla into LTA by the enzymes of dlt operon

**Fig 8:** The DltA inhibitor, 5'-O-[N-(D-Alanyl)-sulfamoyl]-adenosine.

**Fig 9:** Two-step reaction catalyzed by adenylate forming enzymes.

**Fig 10:** Crystal structure of PheA from *Bacillus brevis*.

**Fig 11:** Domain rearrangements observed in the adenylate forming enzymes

**Fig 12:** Conversion of the carrier protein from the inactive apo-form to the active holo-form through transfer of the 4'PP cofactor of CoA by PPTases

**Fig 13:** Peptide bond formation catalyzed by C-domains between 2 two adjacent PCPs loaded with their respective amino acids.

**Fig 14:** Coomassie Blue-stained 10% SDS-polyacrylamide gel

**Fig 15:** Coomassie Blue-stained 15% SDS-PAGE gel after Ni-NTA affinity chromatography (lane 2) and after gel filtration (lanes 3-6)

**Fig 16:** HPLC analysis for estimating the amount of holo-DltC formed after in vivo post translational modification from two separate expression batches.

**Fig 17:** Amino acid specificity of wild type DltA determined by the ATP-PPi exchange assay.

**Fig 18:** Differential binding isotherm for the binding of 5'-O-[N-(D-Alanyl)-sulfamoyl]-adenosine to the DltA.

**Fig 19:** Raw ITC data for the binding reaction of DltC to the DltA-inhibitor complex at 25 °C.

**Fig 20:** Native gel shift assay to analyze the DltA-DltC complex formation.

**Fig 21:** Crystals of DltA

**Fig 22:** Crystals of DltC grown by hanging drop method.

**Fig 23:** Ramachandran plot for DltA.

**Fig 24:** Ribbon diagram of DltA bound to AMP (stereo representation).

**Fig 25:** Electron density for the AMP ligand

**Fig 26:** Schematic representation showing interactions between the protein (black) and the AMP ligand (red).

**Fig 27:** Specificity pocket of DltA.

**Fig 28:** Structure based sequence alignments of DltA with members of the adenylate forming family for which structure are available

**Fig 29:** ATP-PPi exchange assay for wt-DltA (grey) and C292A mutant (black).

**Fig 30:** Comparative overlay of the different domain rearrangements among the adenylate forming family.

**Fig 31:** Stereo representation of the P-loop and hinge region.

**Fig 32:** Crystal packing in the I222 (A) and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> (B) crystal form of DltA.

**Fig 33:** DltA in the I222 space group surrounded by close symmetry related molecules.

**Fig 34:** Sequence alignments of DltAs, LnmQ and related A-domains.

**Fig 35:** Phylogenetic comparison of among members of the adenylate forming family

**Fig 36:** DltA structure modelled into the open conformation.

**Fig 37:** DltA structure modelled into the ‘adenylation conformation’.

**Fig 38:** DltA structure modelled into the ‘thiolation conformation’.

**Fig 39:** Domain rotation brings two catalytic residues into the active site

**Fig 40:** F195 of DltA is seen in two side chain torsion angle rotations

### 9.3 List of tables

**Table 1:** Core motifs of the A-domain

**Table 2:** Instruments and accessories used during the study

**Table 3:** Kits

**Table 4:** Enzymes and chemicals used for the study.

**Table 5:** Media

**Table 6:** PCR programme for site directed mutagenesis

**Table 7:** Crystallographic statistics of DltA

**Table 8:** Data collection statistics for holo-DltC.

**Table 9:** Interactions formed by the small domain with symmetry related molecules in the both the I222 and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> crystal forms of DltA (Distance cut off of 2.0-4.0 Å).



## 9.4 Abbreviations

$\mu\text{Ci}$	Microcurie
4'-PPTases	4'-Phosphopantetheinyl transferases
4'-PP	4'-phosphopantetheine
Å	Angstrom
ACP	Acyl carrier protein
ACS	Acetyl-CoA synthetase
ACSM2A	Human medium-chain acetyl-CoA synthetase
A-domain	Adenylation-domain
AMP	Adenosine-monophosphate
AMP-PNP	Adenosine 5'-( $\beta,\gamma$ -imido)triphosphate
ATP	Adenosine-Triphosphate
ATR	Acid tolerance response
CBAL	4-chlorobenzoate:CoA ligase
C-domain	Condensation-domain
CP	Carrier protein
DAbu	D-amino butyric acid
DAla	D-alanine
DHB	2,3-dihydroxybenzoate
DhbE	2,3-Dihydroxybenzoic acid activating domain
DSer	D-serine
DTT	Dithiothreitol
E-domain	Epimerisation-domain
EDTA	Ethylenediaminetetraacetic acid
FAS	Fatty acid synthase
Gro-P	Glycerol-phosphate
GTP	Guanosine-Triphosphate
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Kd	Dissociation constant
kDa	Kilodalton
Ki	Inhibition constant
LABu	L-amino butyric acid

LAla	L-alanine
LB-medium	Luria-Bertani-medium
LC-FACS	Long chain fatty acyl-CoA synthetase
LTA	Lipoteichoic acids
Luci	Luciferase
M9-medium	Minimal medium
MAD	Multiwavelength anomalous dispersion
MIR	Multiple isomorphous replacement
mRNA	Messenger RNA
MWCO	Molecular Weight Cut Off
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthetases
NS	Norspermidin
OD600	Optical density measured at a wavelength of 600 nm
PCP-domain	Peptidyl carrier protein-domain
PEG	Polyethylene glycol
PheA	Phenylalanine activating A domain
PKS	Polyketide synthase
pmol	Picomole
PPi	Pyrophosphate
Rbo-P	Ribitol-phosphate
Rmsd	Root-mean-square deviation
RP-HPLC	Reverse phase- high performance liquid chromatography
Rpm	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Se-Met	Selenomethionine
TE-domain	Thioesterase-domain
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFA	Trifluoroacetic acid
tRNA	Transfer RNA
WTA	Wall teichoic acids

## Curriculum Vitae

**Name:** Huma Yonus  
**Date of Birth:** 21.05.1979  
**Nationality:** Indian  
**Present address:** P.O Box 270984  
Riyadh-11352  
Saudi Arabia  
[huma.y21@gmail.com](mailto:huma.y21@gmail.com)

### Education:

- High School, University of London (GCE), 1995
- Senior Secondary School, Aligarh Muslim University, India, 1995-1997
- B.Sc. (Hons) Botany, Aligarh Muslim University, India, 1997-2000
- M.Sc. Microbiology, Aligarh Muslim University, India, 2000-2002.
- Ph.D student, Martin Luther University, Germany, 2003-2007
- Homemaker, 2007 till present.

### Publications:

Huma Yonus, Piotr Neumann, Stephan Zimmermann, Jürgen J. May, Mohamed A. Marahiel and Milton T. Stubbs (2008) “Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains”, The Journal of Biological Chemistry, 283(47): 32484-32491.

Zimmermann, S., Pfennig, S., Neumann, P., Yonus, H., Weininger, U., Kovermann, M., Balbach, J., and Stubbs, M.T. (2015). High-resolution structures of the D-alanyl carrier protein (Dcp) DltC from *Bacillus Subtilis* reveal equivalent conformations of apo- and holo-forms. FEBS Lett *Manuscript communicated*.

### Presentations:

- 7<sup>th</sup> “Heart of Europe Bio-Crystallography meeting”, Poland. 30 Sept - 2 Oct 2004. “Structural studies of DltA, a D-alanine activating Enzyme”.
- SPP1152 “Evolution metabolischer Diversität“, Martin Luther University, Halle, 7-8 june 2006. “Structural studies of DltA, a D-alanine activating Enzyme”

## **Declaration**

I, the undersigned, declare that this dissertation is solely my own work and no part of it has been submitted to other universities or higher learning institutions. In addition, all sources of materials used in this dissertation have been duly acknowledged.

Huma Yonus  
Riyadh, Saudi Arabia.