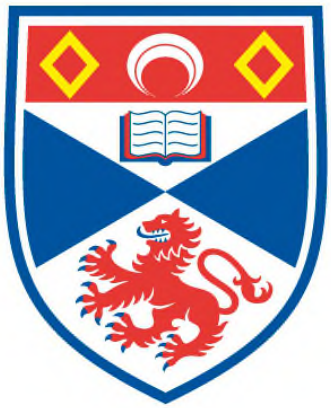
ADENYLATE FORMING ENZYMES INVOLVED IN NRPS-INDEPENDENT SIDEROPHORE BIOSYNTHESIS

Stefan Schmelz

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2010

Adenylate forming enzymes involved in NRPS-independent siderophore biosynthesis

Stefan Schmelz

A thesis submitted to

The School of Chemistry of

The University of St. Andrews for

The Degree of Doctor of Philosophy

November 2009

Supervisor – Prof. James H. Naismith
School of Chemistry & Centre for Biomolecular Sciences, The University of St. Andrews

I. Abstract

Activation of otherwise unreactive substrates is a common strategy in chemistry and in nature. Adenylate-forming enzymes use adenosine monophosphate to activate the hydroxyl of their carboxylic substrate, creating a better leaving group. In a second step this reactive group is replaced in a nucleophilic elimination reaction to form esters, amides or thioesters. Recent studies have revealed that NRPS-independent siderophore (NIS) synthetases are also adenylate-forming enzymes, but are not included in the current superfamily description. NIS enzymes are involved in biosynthesis of high-affinity iron chelators which are used for iron acquisition by many pathogenic microorganisms. This is an important area of study, not only for potential therapeutic intervention, but also to illuminate new enzyme chemistries.

Here the structural and biochemical studies of AcsD from *Pectobacterium chrysanthemi* are reported. AcsD is a NIS synthetase involved in achromobactin biosynthesis. The co-complex structures of ATP and citrate provide a mechanism for the stereospecific formation of an enzyme-bound citryl-adenylate. This intermediate reacts with L-serine to form a likely achromobactin precursor. A detailed characterization of AcsD nucleophile profile showed that it can not only catalyze ester formation, but also amide and possibly thioester formation, creating new stereospecific citric acid derivatives. The structure of a *N*-citryl-ethylenediamine product co-complex identifies the residues that are important for both recognition of L-serine and for catalyzing ester formation.

The structural studies on the processive enzyme AlcC, which is involved in the final step of alcaligin biosynthesis of *Bordetella pertussis*, show that it has a similar topology to AcsD. It also shows that ATP is coordinated in a manner similar to that seen in AcsD. Biochemical studies of a substrate analogue establish that AlcC is not only capable of synthesizing substrate dimers and trimers, but also able to assemble the respective dimer and trimer macrocycles. A series of docked binding models have been developed to illustrate the likely substrate coordination and the steps along dimerization and macrocyclization formation.

Structural and mechanistic comparison of NIS enzymes with other adenylate-forming enzymes highlights the diversity of the fold, active site architecture, and metal coordination that has evolved. Hence, a new classification scheme for adenylate forming enzymes is proposed.

II. Table of Contents

III. List of publications

Most of the results shown and discussed in this thesis are published or will be published soon.

Publications

Schmelz, S., N. Kadi, et al. (2009). "AcsD catalyzes enantioselective citrate desymmetrization in siderophore biosynthesis." Nature Chemical Biology **5**(3): 174-182.

Schmelz, S. and J. H. Naismith "Adenylate-forming enzymes." Current Opinion in Structural Biology **In Press, Corrected Proof**.

Schmelz, S. et al. (2010). "Structural basis for acyl acceptor specificity in the achromobactin biosynthetic enzyme AcsD." (writing in progress).

Other publication/presentation

Talk at the 14th CCP4 Northern Protein Structure Workshop, Carlisle 2007

Title:

AcsD Co-complex structure of an NRPS-independent siderophore synthetase

Award: student prize for best oral presentation

Poster presentation at the American crystallographic association (ACA) meeting, Toronto 2009

Poster title:

Characterization of a new adenylating enzyme class involved in bacterial siderophore biosynthesis

IV. Declarations

I, Stefan Schmelz, hereby certify that this thesis, which is approximately 35'000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in January 2007 and as a candidate for the degree of Doctor of Philosophy (PhD) in January 2008; the higher study for which this is a record was carried out in the University of St Andrews between 2007 and 2009.

Date:_____ Signed: _____

Stefan Schmelz

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy at the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date:_____ Signed: _____

James H. Naismith

V. Copyright

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Embargo on printed copy and electronic copy for the same fixed period of one year on the following ground:

Publication would preclude future publication.

Date:_____ Signed: _____

Stefan Schmelz

Date:_____ Signed: _____

James H. Naismith

VI. Acknowledgments

I am extremely thankful to my supervisor, Professor James H. Naismith, for giving me the opportunity to carry out this work and also for his encouragement, guidance and support from the initial to the final stages of each project and throughout the writing up period. His enthusiasm for crystallography and molecular mechanisms clearly helped a crystallographic "rookie" to get fascinated by a scientific field often described as boring.

I would also like to show my gratitude to Professor Gregory L. Challis and his post doc, Dr Nadia Kadi, from the University of Warwick, who collaborated in these projects, and welcomed me for one week in their laboratory. I much appreciate their time in synthesizing substrates as well as the many helpful scientific discussions. My special thanks go to Dr Catherine H Botting, who did a great job in acquiring MS and MS/MS data for the timeline and fragmentation experiments, and her assistance and expertise during data analysis.

I am indebted to my many colleagues for providing a nice work environment and a lot of support during my PhD. I am especially grateful to Gregor Hagelüken, Srikanathasan Velupillai (better known as Kannan), Lester Carter, Hexian Huang, Kerstin Steiner, Lucile Moynie, Simon Bushell, Kenneth Johnson, Stephen McMahon, Changjiang Dong and the rest of Lab 224. All of them helped and supported me in numerous ways to make this thesis possible. I am especially grateful to Lester Carter and Geethanjali Selvaretnam (better known as Geetha) for proof reading.

I extend my gratitude to my parents, sisters, little nephew Noah and the rest of my family who supported me through the last three years abroad in Scotland. Special thanks to Verena Herbst, for her endless support, understanding and her love.

Thanks all,

Stefan Schmelz

VII. Dedications

To my family...

VIII. List of abbreviations

%	percent
°	degree
∞	infinite

A

α-KG	α- ketoglutaric acid
Å	angstrom
A_{600}	Absorbance at 600 nm
aaRS(s)	aminoacyl-tRNA synthetase(s)
Acs	achromobactin biosynthesis protein
AcsA	Achromobactin biosynthesis protein A
AcsD	Achromobactin biosynthesis protein D
AcsE	Achromobactin biosynthesis protein E
ADN	adenosine
ADP	adenosine diphosphate
AlcC	alcaligin biosynthesis protein C
AMP	adenosine monophosphate
Amp	ampicillin
ATP	adenosine triphosphate

B

B. bronchiseptica	*Bordetella bronchiseptica*
B. pertussis	*Bordetella pertussis*

C

C	Celsius
cAPK	cyclic AMP-dependent protein kinase
CCP4	Collaborative Computational Project No. 4
CD	Circular Dichroism (Spectroscopy)
CID	collisionally induced dissociation
CoA	Coenzyme A
CV	column volume

D

Da	Dalton
DABA	L-2,4-diaminobutyric acid
DesD	desferrioxamine biosynthesis protein D

DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNAseI	deoxyribonuclease I
dNTP	deoxyribonucleotide triphosphate
DPIA	Division of Plant Industry Archive
*Dpn*I	*Diplococcus pneumonia* G41 restriction endonuclease I (NEB)
DTT	1,4-dithiothreitol

E

E.coli	*Escherichia coli*
e.g.	for example
EA	ethanolamine
EDA	ethylenediamine
EDTA	ethylenediaminetetraacetic acid
Em	emission
ESI-MS	electrospray ionization – mass spectroscopy
ESRF	European Synchrotron Radiation Facility
Ex	excitation

F

FDACS	Florida Department of Agriculture and Consumer Services
FU	fluorescence unit(s)

G

g	gram
GE	General electric
GF	gel filtration

H

h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	penta histidine
HP	high-performance

I

IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside

J

JCSG	Joint Center for Structural Genomics

K

kDa	kilo Dalton
keV	kilo electron volt
KG	ketoglutaric acid
kg	kilogram
K_M	Michaelis-Menten constant

L

L	liter
LB	Luria-Bertani
LCT	liquid chromatography time-of-flight
LDH	lactate dehydrogenase
Ltd	limited company

M

M	mol
m	meter
MES	2-(*N*-morpholino) ethanesulfonic acid
M/z	mass charge ratio
mg	milligram
Mg	magnesium
min	minutes
MK	myokinase
mL	milliliter
mM	millimole
mm	millimeter
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MT	methyltransferase
µg	microgram
µL	microliter
µM	micromole
µm	micrometer

N

NADH/NAD+	nicotinamide adenine dinucleotide (reduced/oxidized)
NADPH/NADP+	nicotinamide adenine dinucleotide phosphate (reduced/oxidized)
NCBI	National Center for Biotechnology Information
NCS	non-crystallographic symmetry
NIS	NRPS-independent siderophore
nm	nanometer
NRPS	non-ribosomal peptide synthetases
NTP	ribonucleoside triphosphate

O

O/N	over night

P

PAGE	polyacrylamide gel electrophoresis
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
PDB	Protein Data Bank (database also refers to file type)
PEG	polyethylene glycol
PEP	phosphoenolpyruvate
PFU	DNA polymerase of *Pyrococcus furiosus*
PISA	Protein Interfaces, Surfaces and Assemblies (web service)
PK	pyruvate kinase
PLP	pyridoxal-5'- phosphate
Pol	DNA polymerase
PP_i	pyrophosphate

R

rmsd	root mean square deviation
rpm	revolutions per minute
RT	room temperature

S

[S]	substrate concentration
Sub204	*N*-hydroxy-*N*-succinylputrescine
Sub220	N-(4-Amino-3-hydroxy-butyl)-N-hydroxy-succinamic acid
S/N	supernatant
SDS	sodium dodecyl sulfate
sec	seconds
SeMet	selenomethionine

SER	surface entropy reduction
SPoRT	Structural Proteomics of Rational Targets Initiative
SSPF	Scottish Structural Proteomics Facility

T

TE	thioesterase
TEV	tobacco etch virus
TLS	Translation/Libration/Screw rotation
TOF	time of flight
TRIS	tris(hydroxymethyl)aminomethane
TRIS-HCl	tris(hydroxymethyl)aminomethane Hydrochloride
tRNA	transfer ribonucleic acid

U

U	unit(s)
UV	ultra violet (radiation)

V

v	enzyme reaction rate
(v/v)	volume per volume
v_{max}	enzyme's maximum reaction rate

W

(w/v)	weight per volume
wt	wild type

IX. List of Figures

Following figures were drawn with PyMOL and labeled with CorelDraw:
Fig. 1.3, 2.1, 2.4, 2.5, 2.16b-c, 2.17, 2.18, 2.19, 2.20, 2.21, 2.23a, 2.32, 2.33, 2.35, 2.36a-b, 2.37a-b, 3.1, 3.3, 3.9, 3.10, 3.11a-b, 3.11d, 3.12, 3.13, 3.24, 3.25, 3.26c, 3.28, 3.29, 4.3, 4.4, 4.5, 4.6

Following figures were drawn with ChemBioDraw and labeled with CorelDraw:
1.2b, 1.4, 1.6, 1.8, 1.10, 1.13, 2.6, 2.26, 2.27b-c, 2.30, 2.36d, 2.37c, 2.38, 2.39, 3.2, 3.8, 3.15, 3.16, 3.17, 3.18, 3.19, 3.20, 3.21, 3.22, 3.23, 3.30, 4.2

Following graphs were generated in Origin (OrignLabs) and labeled in CorelDraw:
Fig. 2.2, 2.9, 2.10, 2.11, 2.12, 2.13, 2.22, 2.23b, 2.27a, 2.34, 3.14,

Following figures were drawn or labeled with Corel Draw:
Fig. 1.1, 1.2a, 1.5, 2.3, 2.6, 3.4, 3.6, 3.26a, 3.27,

Following figures were drawn with Microsoft Powerpoint:
Fig. 2.36c, 4.1,

X. List of Tables

Chapter 1 - Introduction

1.1 Siderophore-mediated iron uptake in bacteria

Iron plays an essential role in the cell metabolism of all living organisms. It is, for instance, incorporated into the heme complexes of cytochrome proteins which mediate redox reactions, and the oxygen carrier proteins such as hemoglobin, myoglobin and leghemoglobin (Kiese 1966; Chance 1967; Beri and Chandra 1993). Iron is also essential in the iron-sulfur clusters of nitrogenases (nitrogen fixation) and hydrogenases (anaerobic hydrogen metabolism) (Rees 2002). Other non-heme containing iron proteins are ribonucleotide reductases and methane monooxygenases, which reduce ribose to deoxyribose in DNA biosynthesis (Atkin et al. 1973) or oxidize methane to methanol in methanotrophic bacteria (Higgins et al. 1981), respectively.

Although iron is one of the most abundant elements on earth, all living organisms have limited access to it (Köster 2001; Rouault 2003) due to the insoluble nature of iron(III) (the ferric form; Fe^{3+}) at physiological conditions. Fe^{3+} is the predominant form of Fe ions in an oxygen atmosphere. Iron acquisition from the environment is problematic for saprophytic organisms. Pathogens of higher eukaryotes compete for iron in the host that is tightly bound to transferrin, a transport and storage glyco protein (Rouault 2003). Many microorganisms achieve this by the biosynthesis and excretion of low molecular weight iron chelating compounds, the so called siderophores (Drechsel and Jung 1998; Challis 2005).

Siderophores gather iron(III) from the environment or their host. The resulting iron-chelate complexes are re-assimilated by an iron uptake system (Figure 1.1). In Gram positive bacteria siderophore-iron(III) complexes are recognized by membrane-

associated lipoprotein receptors and transported into the cytoplasm by ATP-dependent ABC transporters (Köster 2001). There iron is reduced to iron(II) and the liberated siderophore molecules are recycled (re-used or degraded). In Gram negative microorganism the receptors for siderophore-iron(III) complex are located in the outer membrane. The complex is recognized and then transported by a TonB proton pump into the periplasm, where a soluble transport and receptor protein binds the complex and directs it to an ABC transporter analogous to the one found in Gram positive bacteria (Figure 1.1) (Köster 2005).

Synthesis of siderophores in bacteria is primarily accomplished by two main pathways: the well studied non-ribosomal peptide synthetases (NRPS) pathway and the less well described NRPS-independent siderophore (NIS) pathway. Enzymes of both pathways are members of the adenylate-forming enzyme family.

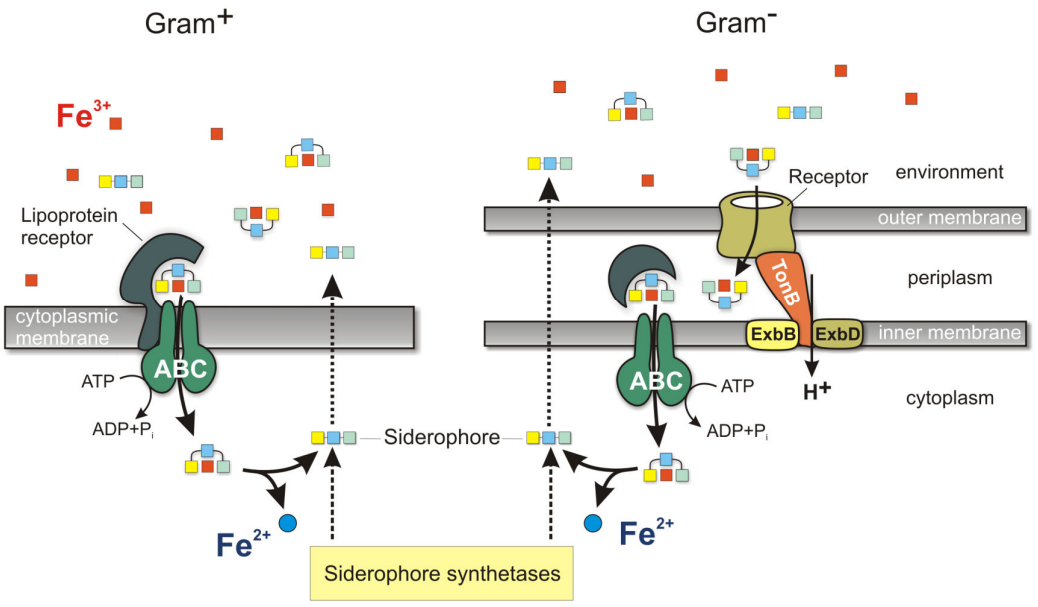


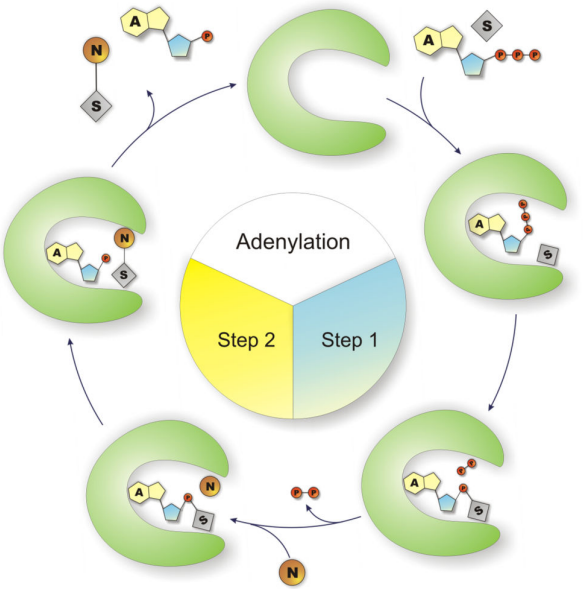
Figure 1.1: Siderophore mediated iron acquisition in prokaryotes.

1.2 Adenylate-forming enzymes

Adenylation is an elegant biochemical process that is used to carboxylate substrates for a second nucleophilic attack. The enzymes catalyze a two step reaction (Figure 1.2a). In the first step, condensation of a carboxylate substrate and ATP is performed to generate a carboxylated adenylate. In the second step the highly reactive adenylate intermediate reacts with a second substrate forming the final carboxylated product. The mechanism proceeds through a five coordinated phosphorus transition state stabilized by the enzyme (Figure 1.2b).

Adenylating enzymes are involved in various metabolic pathways such as ribosomal and non-ribosomal peptide synthesis, fatty acid oxidation and enzyme regulation. Based on a sequence analysis, a superfamily with four sub-families of adenylate-forming enzymes was proposed (Fulda et al. 1994). Aminoacyl-tRNA synthetases, which perform a similar adenylation reaction (Perona and Hou 2007), but are a structurally distinct family, have not been integrated into the scheme. Recent studies suggest that NRPS-independent siderophore (NIS) synthetases also activate their substrate by adenylation (Kadi et al. 2008; Kadi and Challis 2009), but show no sequence identity to other adenylate forming enzymes. This suggests a new classification scheme may be needed (Schmelz and Naismith 2009).

a



b

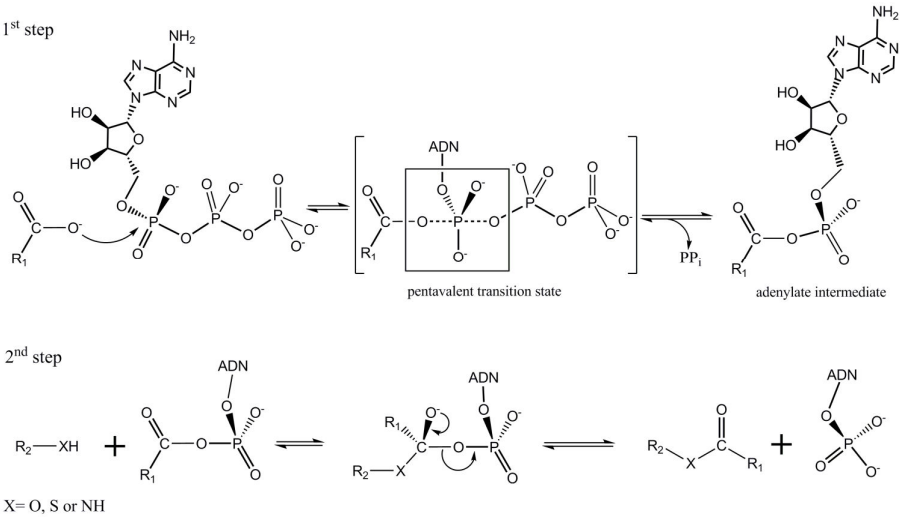


Figure 1.2: Adenylation is a biochemical important process **(a)** Schematic representation of the two step process of adenylation. In the first reaction step adenylate-forming enzymes activate their substrate with ATP (A) by generating a reactive substrate-adenylate intermediate and pyrophosphate (PP_i), which is instantly released. In the second part of the reaction a nucleophile (N) reacts with the activated intermediate and liberates the final product and adenosine monophosphate. **(b)** Chemical representation of the two step adenylation reaction catalyzed by adenylate-forming enzymes to activate an alcohol, amide or sulfhydryl moiety of the nucleophile forming a pentavalent transition state.

1.2.1 Adenylate-forming enzymes

Adenylate-forming enzymes were subdivided into four separate, but closely related sub-families (Fulda et al. 1994), comprising the adenylation domains of NRPS (May et al. 2002; Challis and Naismith 2004), acyl- or aryl CoA synthetases (Jogl and Tong 2004), the class of luciferase oxidoreductases (Conti et al. 1997) and eukaryotic long chain acyl-CoA ligases. In many publications, adenylate-forming enzymes are just divided into three families, combining acyl-/aryl CoA synthetases and eukaryotic long chain acyl-CoA ligases.

Although the superfamily of adenylate forming enzymes has been known for several decades, it was only thirteen years ago that the first structure of a member (firefly luciferase from *P. pyralis)* was solved (Conti et al. 1996). So far eleven other structures have been reported comprising the characteristic two-domain architecture, described as “hammer-and-anvil” topology with a large N-terminal (anvil) and a much smaller C-terminal (hammer) domain (Baldwin 1996; Conti et al. 1996; Kochan et al. 2009) (Figure 1.3a).

The N-terminal fold is subdivided into three sub-domains with an α/β topology, topped by a αβαβα sandwich and a β-barrel (Conti et al. 1996). The C-terminal domain on top is made up of five β-strands and three α-helices. The interface of the N- and C-terminal domain forms the catalytic site. A hinge region provides the C-terminal domain with flexibility, and allows switching from the adenylate-forming to the thioester-forming conformation. The rotation of the C-terminal domain shifts a conserved catalytic lysine residue in and out of the active site, and also controls its solvent accessibility (Kochan et al. 2009; Osman et al. 2009). In contrast to other

adenylate-forming enzymes, such as aminoacyl-tRNA synthetases, this domain movement during catalysis is unique.

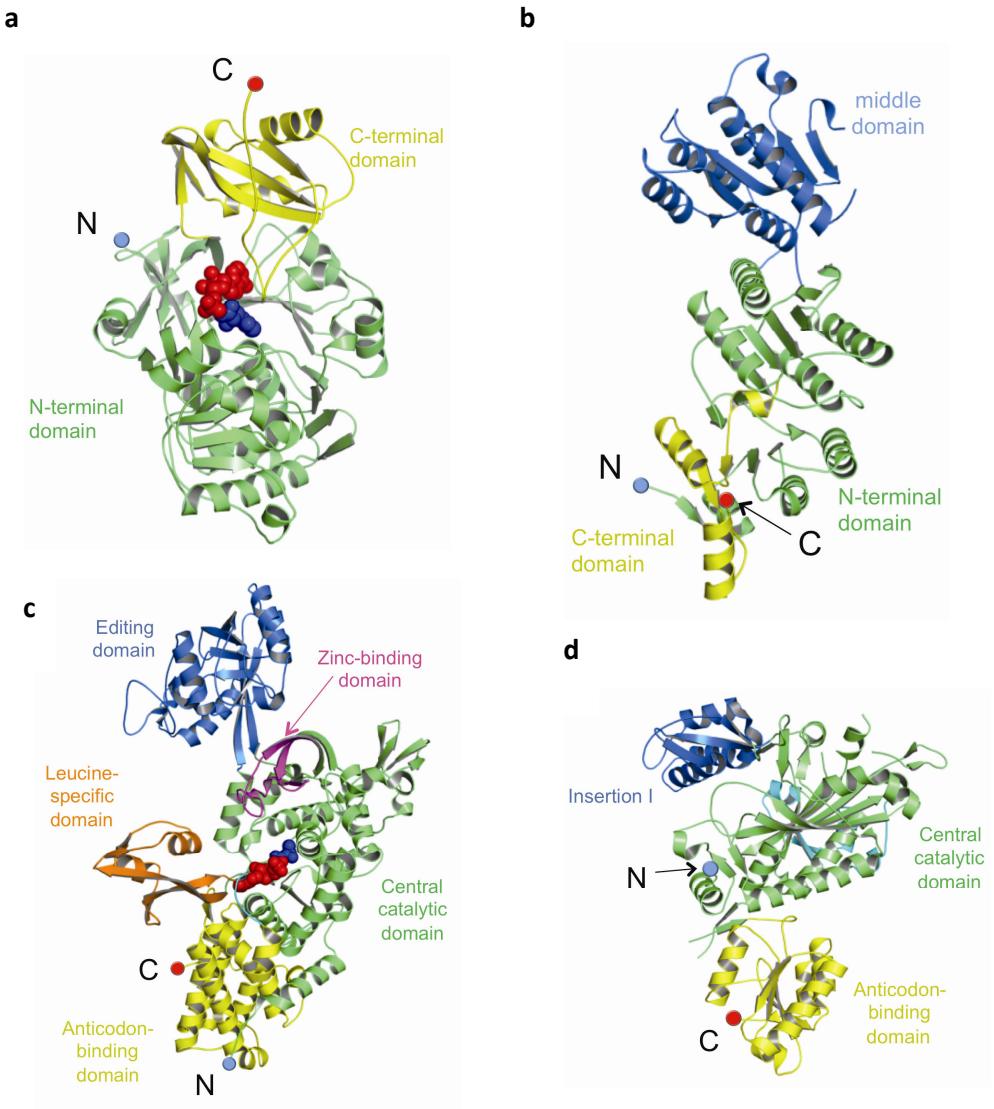


Figure 1.3: Adenylate-forming enzymes. N- and C-terminus are marked with blue and red spheres, respectively. The N-terminal and C-terminal domain are colored in green and yellow, respectively. **(a)** non-ribosomal independent Gramicidin synthetase N-terminal adenylation domain in complex with AMP (red spheres) and phenylalanine (blue spheres, PDB code: 1amu). **(b)** Crystal structure of a predicted acyl-CoA synthetase from *E. coli*, with an unusual three domain architecture (3dmy). **(c)** LeuRS form *T. thermophilus* (1obc) a class I aminoacyl tRNA synthetase in complex with a sulfamoyl analogue of leucyl-adenylate (red spheres) and leucine (blue spheres). **(d)** monomer of human GlyRS (2pme) a subclass class II aminoacyl-tRNA synthetase.

In the superfamily of adenylate-forming enzymes ATP is coordinated in a U-shaped formation through a highly conserved glycine-rich loop containing the consensus sequence YTSG(S/T)TGxPKG (Marahiel et al. 1997). This loop shows similarities with the ATP-binding region in some aminoacyl-tRNA synthetases.

1.2.1.1 Adenylation domain of NRPS

Non-ribosomal peptide synthetases (NRPS) synthesize small peptides with biological activities ranging from antibiotics to immune suppressors (Challis and Naismith 2004). NRPS are modular enzymes comprising three core domains, an adenylation (A) domain, a peptidyl carrier protein (PCP) domain and a condensation (C) domain. NRPS catalyze the formation of peptide bonds by activation of their amino acid substrates using an adenylation step which generates a highly reactive amino- or iminoacyl adenylate intermediate (see section 1.2, Figure 1.5). This adenylate intermediate reacts with a phosphopantetheinyl prosthetic group of the PCP domain forming a reactive thioester. The C domain subsequently catalyzes an amide bond formation with the following aminoacyl thioester attached to an upstream PCP domain. The A domain of NRPS belongs to the adenylate-forming enzyme super-family and is responsible for the selection of a specific substrate and its activation. So far three structures have been reported: the N-terminal phenylalanine-activating A domain of gramicidin synthetase I, GrsA (Figure 1.3a) (Conti et al. 1997), 2,3-dihydroxybenzoate-activating E domain DhbE (May et al. 2002) and D-alanyl carrier protein ligase DltA (Yonus et al. 2008).

1.2.1.2 Acyl- or aryl CoA synthetases

Acyl-/aryl CoA-synthetases, also known as thiokinases, are the second sub-family of adenylate-forming enzymes. They catalyze formation of acyl-/aryl-CoA from short or medium chain fatty acids and coenzyme A in a two step adenylation reaction. First the fatty acid substrates are adenylated and then transferred onto coenzyme A. The activation of carboxylic acids with coenzyme A is an important metabolic key reaction and is involved in fatty acid synthesis, fatty acid β-oxidation and metabolic citric acid cycle. To date, four structures in this sub-family have been published: Acetyl-CoA of *S. enteric* (Gulick et al. 2003) and *S. cerevisiae* (Jogl and Tong 2004), human acyl-coenzyme A (Kochan et al. 2009) and long chain fatty acid-CoA ligase from *T. thermophilus* (Hisanaga et al. 2004).

While these four structures have the characteristic two domain architecture, another recently solved structure of FdrA, a "predicted Acyl-CoA synthetase" from *E. coli* (3dmy), comprises three domains: a large N-terminal domain, a bigger middle domain substituting the usual C-terminal domain of adenylate-forming enzymes, and a small C-terminal domain, wrapping parts of the N-terminal domain (Figure 1.3b). However, the enzyme's function is unclear.

1.2.1.3 Luciferase and other oxidoreductases

Oxidoreductases are exemplified by firefly luciferase, which drives bioluminescence, in this case enzymatic generated light for mating. The light is emitted by luciferin, a heterocyclic compound, which is oxidized by molecular oxygen. First luciferin is adenylated by luciferase forming luciferyl-adenylate. Thereafter molecular oxygen

oxidizes the activated luciferin. This reaction emits light and generates oxyluciferin and adenosine monophosphate. A second structure from *L. cruciata* was solved recently by the group of Nakatsu et al (Nakatsu et al. 2006). Both structures show the characteristic N-terminal and C-terminal domains and share basic sequence similarities with NRPS adenylation domains and Acyl-CoA synthetases, but not with aminoacyl-tRNAses (Nakatsu et al. 2006). The oxidoreductase 4-chlorobenzoyl-CoA ligase (Wu et al. 2009) and the benzoate CoA ligases from *Burkholderia* species (Bains and Boulanger 2007) have been determined recently. While 4-chlorobenzoyl-CoA ligase catalyzes the initial dehalogenation step of 4-chlorobenzoate, benzoate CoA ligase performs the coenzyme A dependent thioesterification of benzoate.

1.2.2 Aminoacyl-tRNA synthetases

Aminoacyl-tRNA synthetases (aaRSs) are involved in recognition, activation and covalently binding of amino acids to their associated tRNA. This transfer occurs in a similar two-step mechanism as described for adenylate-forming enzymes. After initial activation of the amino acid with ATP the amino acid adenylate is transferred to the 3’ or 2’ hydroxyl group of tRNA forming a covalent ester bond. The group of aaRSs was subdivided into two evolutionarily distinct classes with several subclasses, for full details see reviews of (O'Donoghue and Luthey-Schulten 2003) and (Woese et al. 2000). Aminoacyl-tRNA synthetases are well-characterized. The first structural work was published in the mid 70’s (Irwin et al. 1976), and the first co-complex with tRNA was reported four years later (Giege et al. 1980; Giege et al. 1982; Lorber et al. 1983). These studies showed that the two main families are not structurally related and have evolved convergently (Figure 1.3c,d) (Cusack et al. 1990; Eriani et al. 1990).

While ATP in family I is bound in a Rossmann fold along with two conserved sequence signatures HIGH and KMSKS, ATP in family II enzymes is coordinated by a seven strand anti-parallel β-sheet with three conserved motifs (Cusack 1995; Crepin et al. 2006).

1.2.3 NRPS-independent siderophore synthetases

NRPS-independent siderophore (NIS) synthetases are a new enzyme class discovered less than 20 years ago. 88 different enzymes were identified and subdivided into three phylogenetic groups called type A, B and C. The enzymes are involved in different steps in biosynthesis of siderophores (Challis 2005). Siderophore biosynthesis pathways are often activated by pathogenic bacteria during host invasion and hence are potential drug targets. In contrast to NRPS enzymes, which are also involved in siderophore synthesis, NIS enzymes are not well studied. All enzymes, AcsD, AcsA, AlcC and DesD, reported in this thesis are NIS enzymes. AcsD and AcsA are involved in achromobactin biosynthesis of *P. chrysanthemi*, a plant pathogen. AlcC and DesD on the other hand catalyze the final steps in alcaligin and desferrioxamine biosynthesis, respectively.

1.3 NRPS pathway

Usually peptide bond formation, for instance in protein biosynthesis, is catalyzed by ribosomes. However, non-ribosomal peptide synthetases (NRPS) were found to catalyze peptide bond formation in a diverse group of smaller molecules (Figure 1.4).

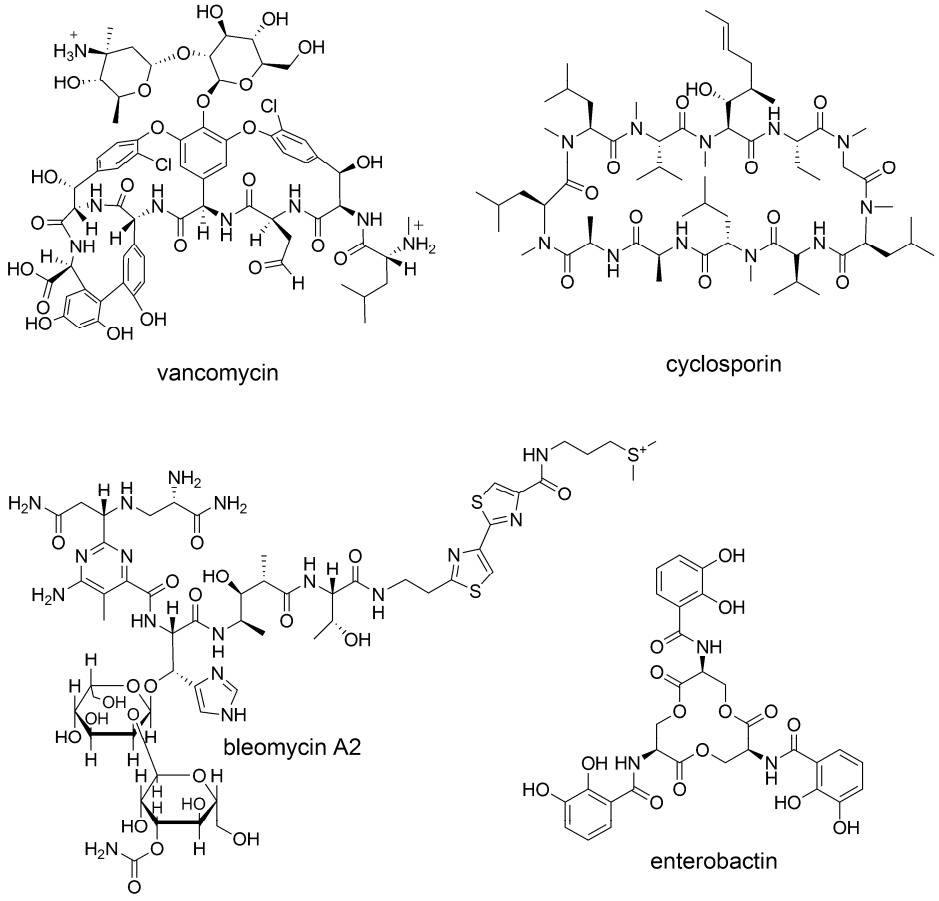


Figure 1.4: Diversity of molecules synthesized by NRPS pathways (modified from Challis and Naismith 2004).

The diversity reaches from peptide-based antibiotics such as vancomycin (van Wageningen et al. 1998) to immune suppressors like cyclosporine (e.g. used in treatment after organ transplants) or bleomycin, which is an anti tumor drug (Weber et al. 1994; Du et al. 2000). Furthermore, NRPS are intensively studied and well

characterized and are also involved in siderophore biosynthesis, such as in enterobactin biosynthesis from *E. coli (Crosa and Walsh 2002)*. Most NRPS assembled molecules are cyclic and contain non-proteinogenic amino acids. NRPS are large protein complexes of iterative enzyme domains (Figure 1.5). The number of modules and domains in the enzyme is dependent on the final number of incorporated and utilized substrate molecules. NRPS have been reported in several siderophore assembling systems (Gehring et al. 1998; Yu et al. 1998; Miller et al. 2002). NRPS comprise three types of modules: the initiation, elongation and termination modules (Figure 1.5a).

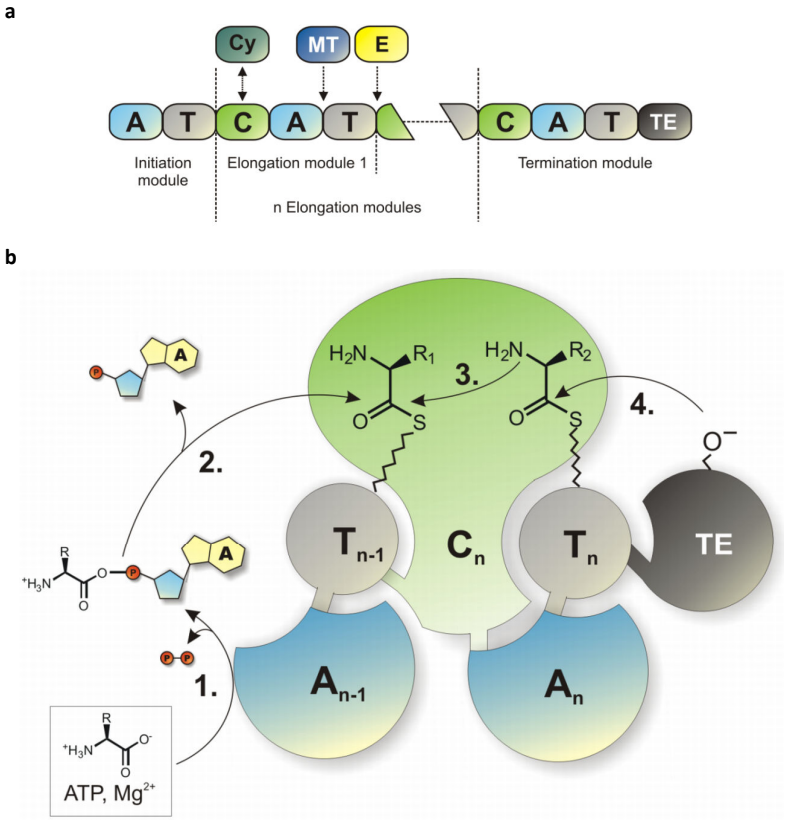


Figure 1.5: Non-ribosomal peptide biosynthesis. **(a)** Modules and domains involved in NRPS biosynthesis (modified from Challis and Naismith 2004). A minimal elongation module comprises three core domains: condensation (C), adenylation (A) and thiolation (T) domain. While the initiation module lacks a condensation domain the termination module has an additional thioesterase (TE) domain. Some NRPS comprise additional epimerization (E), methyltransferase (MT) or special condensation domains (Cy), able to catalyze intramolecular heterocyclizations. **(b)** Schematic representation of the NRPS reaction and the transfer of intermediates between different domains (modified from Marahiel et al. 2009).

A minimal elongation module consists of three core domains: condensation (C), adenylation (A) and a peptidyl carrier (PCP) domain, respectively. The PCP domain is also known as thiolation or T domain (Lautru and Challis 2004). The initiation module usually lacks a condensation domain, while the termination module has an additional thioesterase (TE) domain to liberate the finalized peptide. At the start of each reaction domain A selectively binds and activates an amino or carboxylic acid via adenylation (Figure 1.5b). Thereafter the generated aminoacyl adenylate is passed on and covalently bound to a thiol group in the PCP domain forming an active thioester. This substrate activation occurs in every module and consumes ATP. The final bond formation (whether peptide or ester) happens between two subsequent PCP domain aminoacyl thioesters. Certain NRPS comprise additional epimerization (E) domains (racemization of C_{α}- carbons in amino acids), methyltransferase (MT) domains (methylation of amines) or special condensation (C_{y}) domains, which catalyze condensation and intramolecular heterocyclization of serine, cysteine and threonine residues. This modular complexity and flexibility explains the diversity of NRPS synthesized molecules.

1.4 NIS pathway

Several non-peptide based bacterial siderophores are synthesized by a second pathway (Figure 1.6). This so called NRPS-independent siderophores (NIS) biosynthesis pathway appears to be present in over 40 species of bacteria (Challis 2005). In 1986, initial biochemical genetic studies showed that aerobactin (Figure 1.6), an *E. coli* siderophore, is biosynthesized via an unknown NRPS-independent

pathway (de Lorenzo et al. 1986; de Lorenzo and Neilands 1986). In aerobactin the functional iron-chelating groups are linked by two amide bonds to either side of a dicarboxylic acid. These reactions are catalyzed by IucA and IucC, two NIS enzymes, and remained an exception over nearly 10 years. Since then more NIS siderophores were discovered (Figure 1.7). Their pathways involve up to three functionally related enzymes to the aerobactin pathway (Challis 2005). Several NIS-synthesized iron chelators were proven to be essential for pathogenesis and development of pathogens within their hosts. Aerobactin for instance is synthesized by many virulent strains of *E. coli* (Warner et al. 1981) while anthrachelin was confirmed to be crucial for growth of *Bacillus anthracis* in macrophages (Cendrowski et al. 2004).

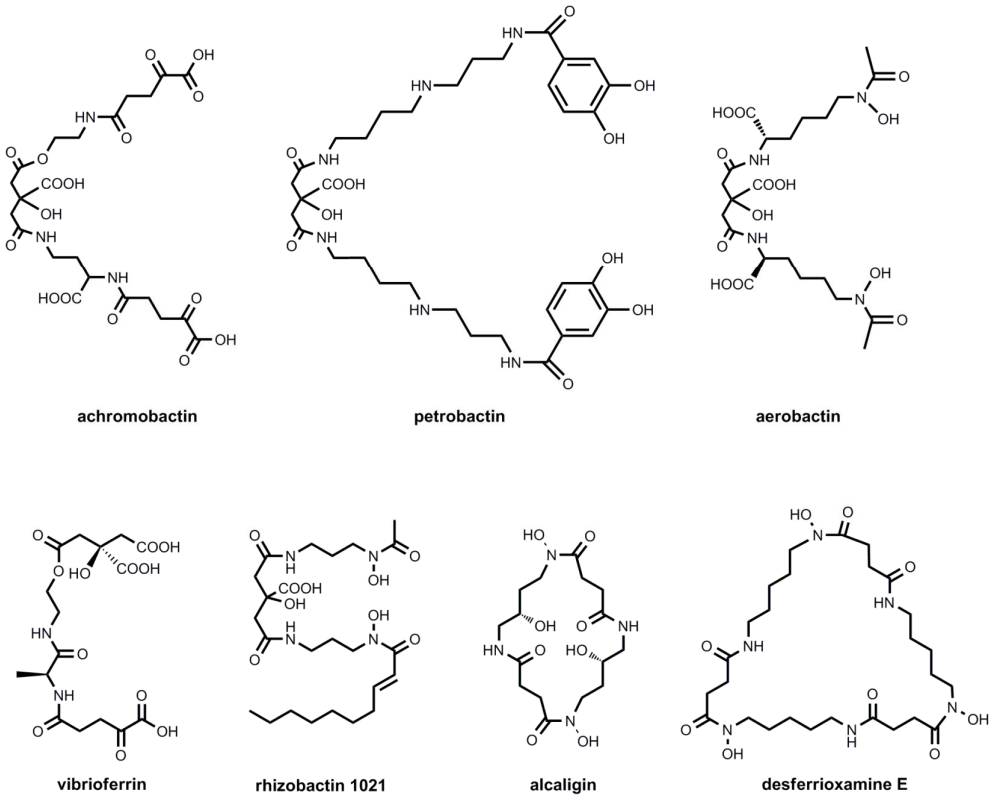


Figure 1.6: Siderophores synthesized by NIS enzymes: achromobactin, petrobactin, aerobactin, vibrioferrin, rhizobactin 1021, alcaligin and desferrioxamine E (modified from Challis 2005; Wilson et al. 2006).

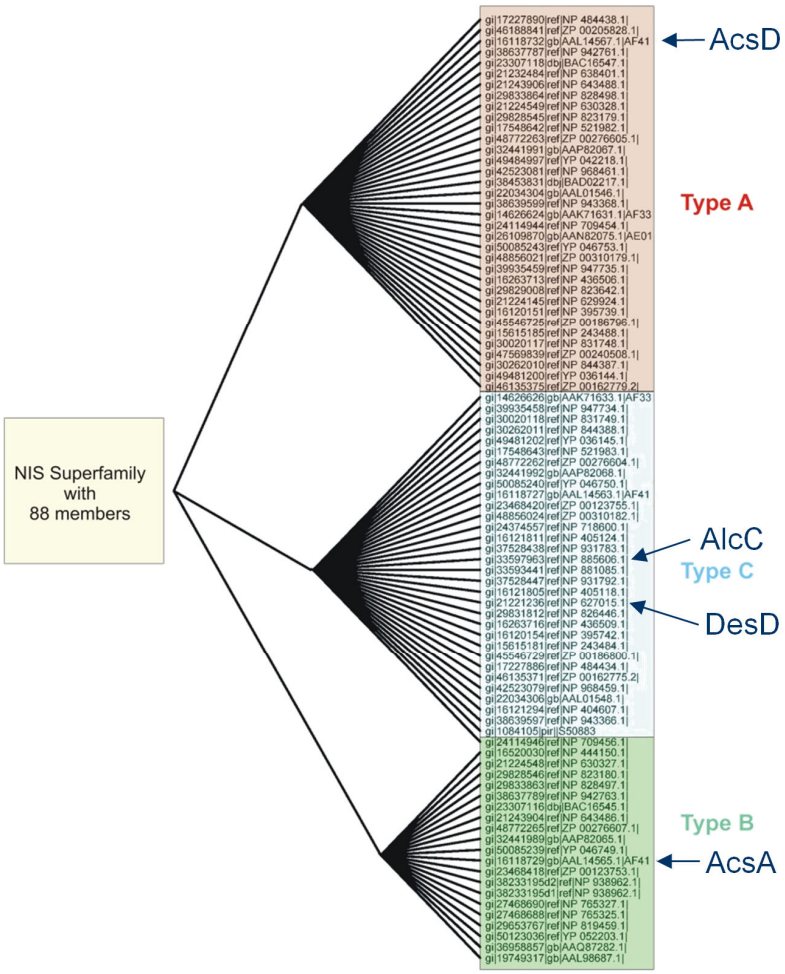


Figure 1.7: NIS superfamily: type A, B and C enzymes. Highlighted and labeled are the proteins: AcsD/AcsA, AlcC and DesD which are involved in achromobactin biosynthesis in *Pectobacterium chrysanthemi,* alcaligin biosynthesis in *Bordetella pertussis* and in desferrioxamine biosynthesis in *Streptomyces coelicolor,* respectively (modified from Challis 2005; Challis 2007).

Based on sequence similarity to enzymes of the aerobactin biosynthesis (IucA and IucC) Challis et al. proposed a new superfamily of NRPS-independent siderophore synthetases. Phylogenetic analysis showed that its 88 members are clustered into three distinct groups: Type A, B and C, respectively (Figure 1.7). Each group was found to have distinct substrate specificity (Challis 2005; Challis 2007). Type A enzymes catalyze amide or ester bond formation of amine or alcohol substrates with one of the two prochiral carboxyl groups of citric acid (Figure 1.8a). Type B on the

other hand performs amide bond formation of amino groups and the C_5-carboxyl group of α-ketoglutaric acid (α-KG) (Figure 1.8b). Type C catalyzes amide or ester bond formation of amino or alcohol groups of various substrates with monoamine or monoester derivatives of citric or succinic acid (Figure 1.8c).

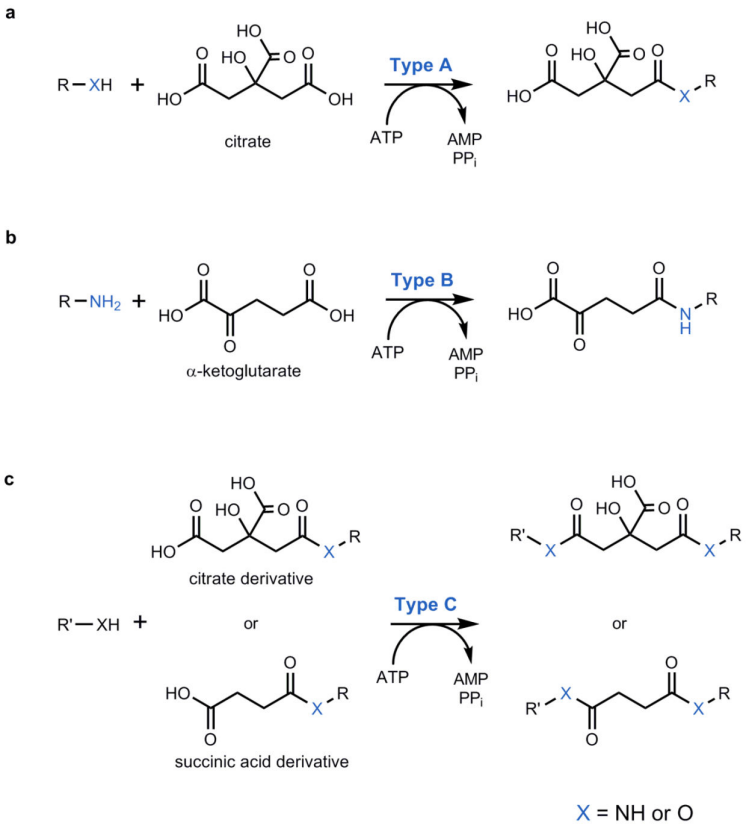


Figure 1.8: Proposed mechanisms NIS enzymes (modified from Challis 2005; Challis 2007). **(a)** Reaction of type A enzymes. **(b)** Reaction of type B enzymes. **(c)** Reaction of type C enzymes.

Unlike NRPS, which were intensively studied and are well understood, NIS enzymes have only recently drawn attention (Kadi and Challis 2009). In 2007, when I started the work on NIS enzymes no protein structure of a superfamily member was reported. NIS enzymes are interesting targets due to their chemistry and potential use in biotransformation. Therefore we decided to obtain structural data for each superfamily member and examine their chemical mechanisms.

1.5 Achromobactin biosynthesis pathway

Achromobactin, a tris-α-hydroxycarboxylate siderophore (Figure 1.6), is biosynthesized via a NIS pathway and excreted by the plant pathogen *Pectobacterium* (formerly *Erwinia*) *chrysanthemi* (Münzinger et al. 2000; Franza et al. 2005). It plays an important role as a virulence factor during plant infection (Franza et al. 2005). *P. chrysanthemi* leads to worldwide economic losses by inducing soft-rotting, wilting, dwarfing and other plant dysfunctions (Figure 1.9) in a wide range of crops including potatoes, bananas, date palms, maize and greenhouse crops (Perombelon and Kelman 1980; Abdalla 2001; Palacio-Bielsa et al. 2006; Slawiak et al. 2009). This is primarily caused by secretion of cell wall degenerating enzymes, such as pectin and pectate lyases, methylesterases and polygalacturonases (Hugouvieux-Cotte-Pattat et al. 1996). Expression of these enzymes is important for pathogenesis of *P. chrysanthemi* and is co-regulated by several environmental conditions such as iron limitation. Depending on iron availability, *P. chrysanthemi* is able to express two siderophore-mediated iron uptake systems (Franza et al. 2002). Under adequate conditions, the siderophore chrysobactin is synthesized and used for iron acquisition, but in response to iron depletion expression of achromobactin is initiated instead (Franza et al. 2005). Among the acs genes that direct achromobactin biosynthesis, three NIS gens, *acsD*, *acsA* and *acsC*, were located. AcsD, AcsA and AcsC are type A, B and C NIS synthetases, respectively and comprise a characteristic IucA/IucC domain, found in aerobactin biosynthesis enzymes (Figure 1.11). Gene

inactivation studies confirmed that Acs proteins are required for siderophore biosynthesis in *P. chrysanthemi* (Franza et al. 2005).

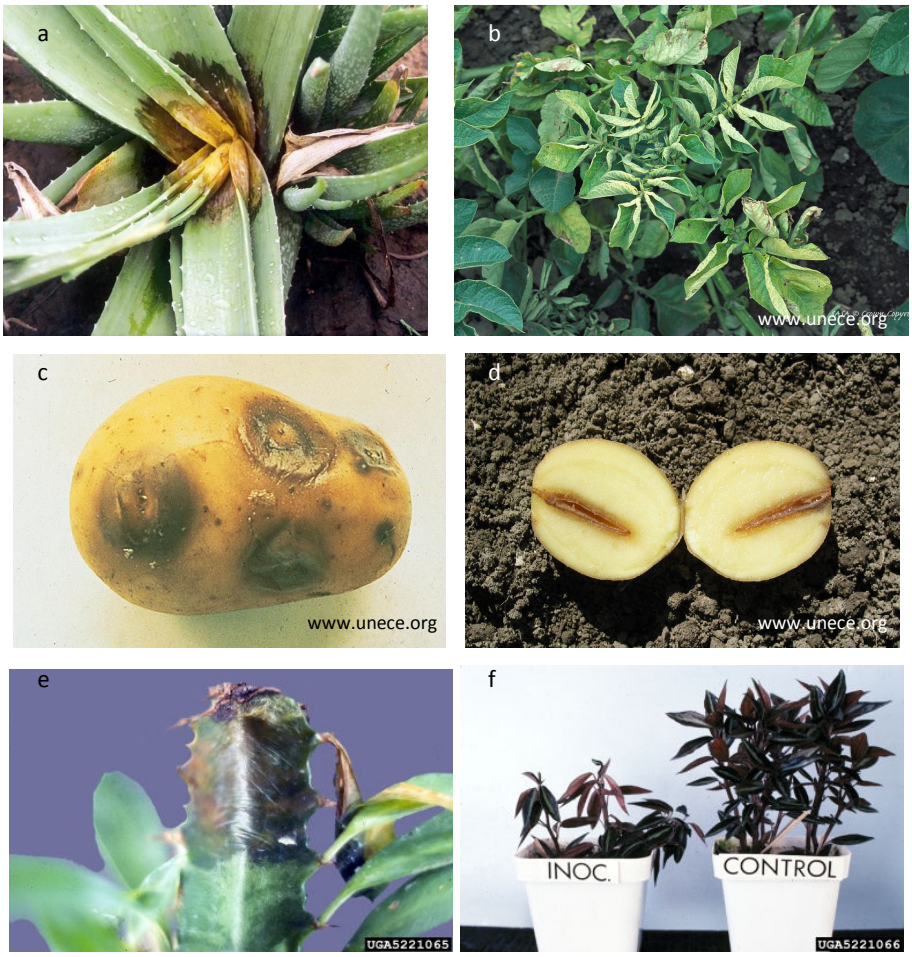


Figure 1.9: Examples of diseases and malfunctions in plants caused by *Pectobacteria* **(a)** Soft rot disease in aloe (Mandal and Maiti 2005) **(b)** Black leg disease in potato, plants yellow, rolled upper leaves, potato tuber: watery, soft rot of disintegrating flesh **(c** and **d)** (b-d source: www.unece.org); **(e)** Bacterial soft rot in Euphorbia spp. (D. Davison DPIA/FDACS Bugwood.org) **(f)** Dwarfing caused by *P. chrysanthemi* (D. Davison DPIA/FDACS Bugwood.org).

A model of achromobactin biosynthesis (Figure 1.10) suggests that L-2,4-diaminobutyric acid (DABA) or ethanolamine (derived from decarboxylation of serine by AcsE) is bound via an amide or an ester bond to one of the prochiral carboxyl groups of citric acid (Challis 2005). This condensation reaction is performed by AcsD, a type A enzyme. Depending on the utilized substrate, the type C siderophore AcsC

catalyzes the linkage of the remaining substrate onto the other prochiral carboxyl group of citric acid. Finally, the resulting citric acid derivative is twice acylated with α-ketoglutaric acid on either site by AcsA to yield achromobactin (Challis 2005; Challis 2007) (Figure 1.10). While AcsD and AcsC are expected to catalyze only one condensation reaction, AcsA is proposed to be a processive enzyme performing two condensation reactions. Hence achromobactin biosynthesis is an excellent model system for studying the substrate specificity, catalytic mechanism and structural biology of NIS synthetases, due to its involvement in all three NIS types.

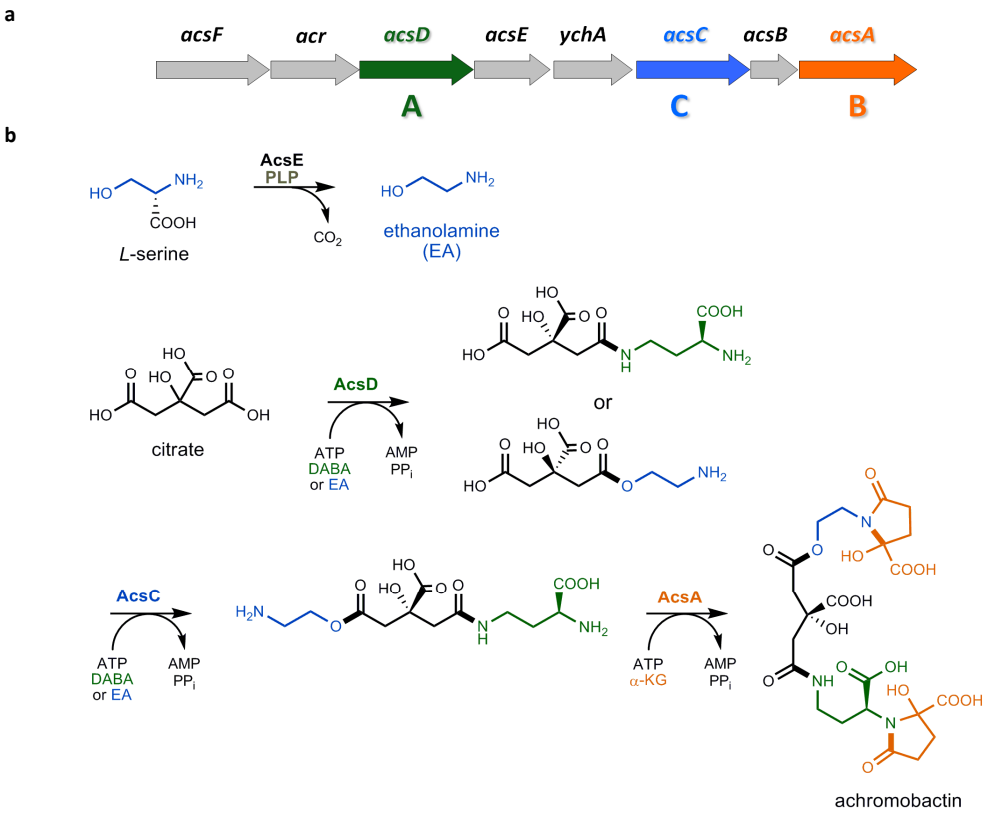


Figure 1.10: Achromobactin biosynthesis. **(a)** Gene cluster directing achromobactin biosynthesis in *P. chrysanthemi*. **(b)** Proposed model of achromobactin biosynthesis (Challis 2005; Challis 2007). AcsE, a pyridoxal phosphate (PLP) dependent enzyme, presumably decarboxylates L-serine generating ethanolamine (EA). AcsD than might utilizes EA or 1,4-diaminobutyric acid (DABA) forming an ester or amide bond with citric acid. Depending on AcsD, subsequent AcsC consumes the remaining substrate binding it to the second prochiral carboxyl of citric acid. AcsA, the last enzyme, acylates the two amino groups on either site with α-ketoglutarate in the resulting citric acid derivative to form the final product achromobactin. Embedded substrates are highlighted in distinct colors for clarity.

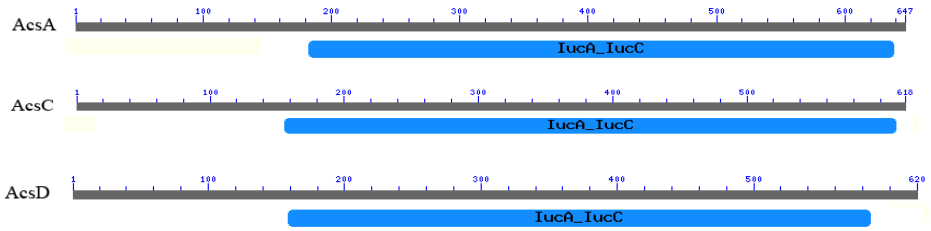


Figure 1.11: Domain annotation for achromobactin synthetase proteins A, C and D (NCBI Conserved Domain Search; http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

1.6 Alcaligin biosynthesis pathway

Alcaligin, a macrocyclic dihydroxamate siderophore (Nishio et al. 1988), is produced and excreted by *Bordetella bronchiseptica* and *Bordetella pertussis*. While *B. bronchiseptica* is the causative agent for bronchitis in smaller mammals such as cats, dogs and rabbits, *B. pertussis* is an obligate human pathogen causing whooping cough. Known as a common infant and young children's disease, it is a persistent respiratory illness and can lead to pneumonia, seizures and encephalopathy and eventually death (Galanis et al. 2006).

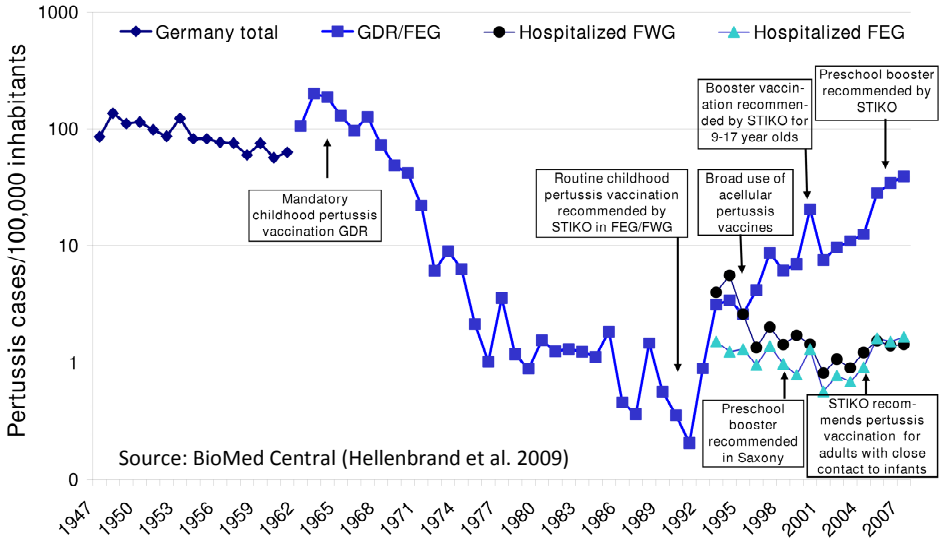


Figure 1.12: Reported cases of pertussis (1947–2007) and hospitalizations (1993–2007) in Germany (Hellenbrand et al. 2009). Abbreviations: FEG: Former East Germany, FWG: Former West Germany, GDR: German Democratic Republic.

To fight whooping cough an active immunization was introduced in many countries over 50 years ago. Though vaccination was successful for the first 30 years, pertussis cases started rising worldwide in the past 20 years. Many studies were carried out searching for explanations and strategies to solve this problem. A recent study in

Germany demonstrates that after implementation of vaccination in the early 1960s, pertussis cases initially dropped from over 200 to less than 1 out of 100'000 inhabitants by 1992. Ever since, increasing number of pertussis cases are reported (Figure 1.12). One of the many reasons for this is the refusal of parents to vaccinate their children and themselves. This behavior definitely increases the risk of *B. pertussis* related infections as an American study demonstrates (Glanz et al. 2009). However, it was also notified that older children, adolescents and adults in countries with high vaccination coverage are now more affected than previously, which is surprising. It was also found that an asymptomatic history (no distinctive cough) in older patients often leads to misdiagnosis causing delayed or incorrect treatment (Tan et al. 2005; Galanis et al. 2006; Hellenbrand et al. 2009; Melo et al. 2009), whereas an early detection is critical and decisive (Melo et al. 2009).

Pertussis, a disease thought to be "defeated" or under control, has made a come-back. Hence, searching for advanced antibiotics or *B. pertussis* drug targets still remains an important field of study. Such a target might be the biosynthesis of siderophores and their uptake systems. Alcaligin and enterobactin siderophore biosynthesis for instance, are triggered by iron limitation and shown to be vital for *B. pertussis* pathogenicity during host invasion and its manifestation (Brickman et al. 2008; Brickman and Armstrong 2009). Although alcaligin was discovered more than 20 years ago (Nishio et al. 1988), its biosynthesis is still unclear. The involvement of two gene clusters, *alcABC* and *alcR,* has been verified (Beaumont et al. 1998; Kang and Armstrong 1998) and a likely biosynthesis pathway was proposed (Figure 1.13) (Challis 2005; Brickman et al. 2007).

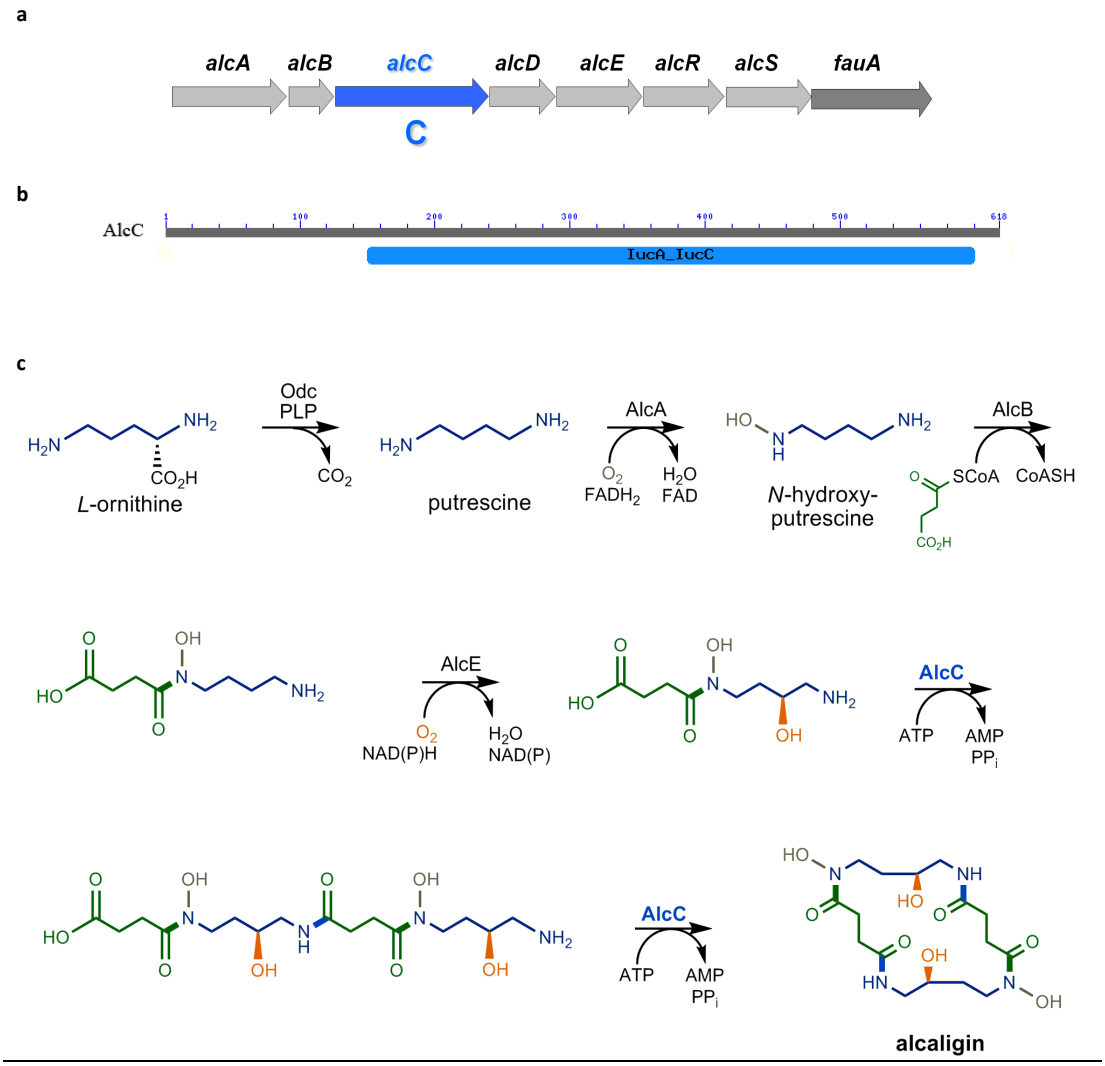


Figure 1.13: Alcaligin biosynthesis. **(a)** Proposed gene cluster directing alcaligin biosynthesis in *B. pertussis* and *B. bronchiseptica*. **(b)** Domain annotation for alcaligin synthetase protein C using NCBI's conserved domain search tools (see also Figure 1.11). **(c)** Proposed model of alcaligin biosynthesis (Challis 2005; Brickman et al. 2007). AlcC, the only NIS enzyme in alcaligin biosynthesis, is a type C enzyme (processive enzyme) catalyzing the last two steps, dimerization and macrocyclization, to finalize the macrocyclic dihydroxamate siderophore.

The gene product of *alcC* is AlcC, a type C NIS enzyme (Figure 1.13a) and comprises, like other NIS synthetases, a characteristic IucA/IucC domain (Figure 1.13b). AlcC is a processive enzyme and performs the final two steps in alcaligin biosynthesis (Figure 1.13b). First it catalyzes the dimerization of two molecules of *N*-(4-amino-3-hydroxy-butyl)-*N*-hydroxy-succinamic acid, a likely product of AlcE. Then it generates a macrocycle of the dimer to yield alcaligin (Figure 1.13b) (Challis 2005). Each reaction

is catalyzed in a similar two step adenylation reaction, as discussed in section 1.2, by consuming ATP (Challis et al, not published). In contrast to achromobactin biosynthesis, AlcC is the only NIS enzyme involved in alcaligin formation. Studying siderophore-mediated iron uptake systems might discover new ways to harm or at least slow down bacterial growth in mammals. This was impressively demonstrated by Register et al. In their study, neonatal pigs were treated with a *B. bronchiseptica* alcaligin synthesis mutant. Although the pigs were infected, they developed milder clinical symptoms than a control group (Register et al. 2001).

Chapter 2 - AcsD

2.1 Introduction

2.1.1 Summary

In this chapter, the structural and biochemical studies of AcsD from *Pectobacterium* (formerly *Erwinia*) *chrysanthemi* are reported. This enzyme is part of a biosynthetic pathway that produces the siderophore achromobactin. Siderophore production is a known target for new antibacterial agents. AcsD is proposed to esterify the symmetric molecule citric acid in a stereochemical controlled two step reaction. Co-complex structures of ATP, citrate and *N*-citryl-ethylenediamine product complex provide a mechanistic rationale for stereospecific formation of an enzyme-bound (*3R*)-citryl-adenylate, which disintegrates by a nucleophilic attack with L-serine forming *O*-citryl-L-serine, a likely achromobactin precursor. Site directed mutations and biochemical studies assign roles to the catalytic residues.

A detailed characterization of the substrate profile for the second reaction step demonstrated that AcsD is capable of assembling enantioselective citrate amides, esters and possibly thioesters. The product co-complex identifies important residues for recognition of L-serine. One of the isolated mutants shows a reversal of specificity with respect to non-natural substrates. This suggests that AcsD (and other members of the superfamily) may have biotransformation potential. Structural and biochemical data suggests that AcsD is a novel adenylate-forming enzyme with a new fold and chemical catalysis strategy.

2.1.2 AcsD apo-structure

The apo-structure of AcsD was solved and refined in the Structural Proteomics of Rational Targets Initiative (SPoRT) in St Andrews when I began this project. Therefore the overall structure of AcsD will be introduced briefly. AcsD comprises 620 amino acids with a total weight of about 70 kDa. The analysis of the structural data with PISA (Krissinel and Henrick 2007) identified a homo dimer with a small interface across 20 residues (making 4 hydrogen bonds) which buries 821 $Å^2$ per monomer (Figure 2.1a). Analytic gel filtration was used to confirm the dimeric state of AcsD (Figure 2.2).

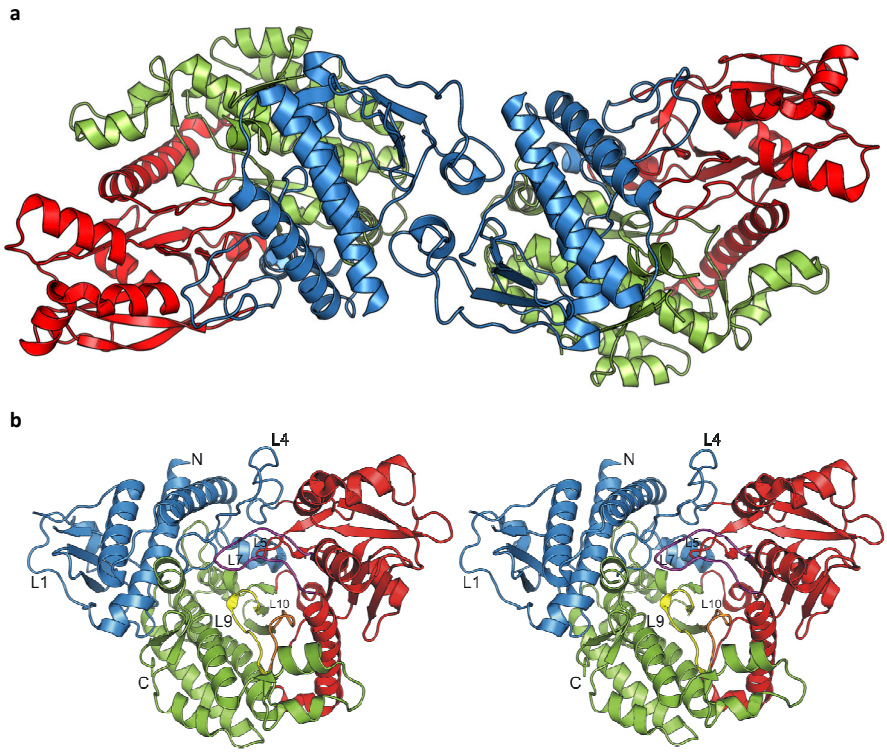


Figure 2.1: AcsD apo-structure **(a)** Dimer of AcsD, facing the interface. The C-terminal domain is colored in blue, middle domain in red and N-terminal domain in green. **(b)** AcsD monomer in stereo view, facing the possible binding pocket. Coloring as above. Secondary structure elements are labeled from the N-terminus. The NIS synthetase conserved loop L7 covers the possible binding pocket.

AcsD monomers are composed of three domains which together are best visualized as a cupped hand. The N-terminal domain (residues 7 – 147, the fingers of the hand) forms a three helix bundle flanked by a four stranded anti-parallel β-sheet. A 28 residue loop (L1) interrupted by a four residue helix lies outside to the strand (Figure 2.1b). This domain forms the dimer interface. Four hydrogen bonds located in this domain are involved in dimerization. Surprisingly, none of these hydrogen bonding residues are conserved among the superfamily. The middle domain (wrist of hand) is located opposite the N-terminal domain and is linked to it by a disordered large loop (residues 164-197, shown in blue). The 31 residue loop is stabilized by a network of intramolecular interactions and E193, which is conserved throughout the superfamily and makes 3 salt bridges with K177 and R179. Later two residues are also well conserved throughout NIS enzymes. The middle domain includes 182 residues folded in a twisted 8 stranded anti-parallel β-sheet interspersed with 5 α-helices and an extended loop structure. Seven residues, H242, S279, R281, T282, K293, T301 and R305, conserved throughout the NIS superfamily are within this domain and three, namely S279, R281 and T282, are situated on L5. A high level of sequence conservation is also illustrated in L7, which with L4 seals the cavity created by the arrangement of the three domains, throughout type A enzymes of the NIS superfamily which is not preserved in type B or C.

The C-terminal domain (residues 381 – 587) is predominantly α-helical and leads back to the N-terminal, what creates a cavity, the active site, in the middle of the structure (Figure 2.1b). A large 4 helix bundle is flanked by a three stranded anti-parallel β-sheet on the inside and by a small two stranded anti-parallel β-sheet on

the outside. As in the other domains, the secondary structure elements are interspersed with large loops, two of which, L9 and L10, are on the inner face of the domain and are home to three superfamily conserved residues, H444, N447 and D464 and N509 which is conserved in type A enzymes only.

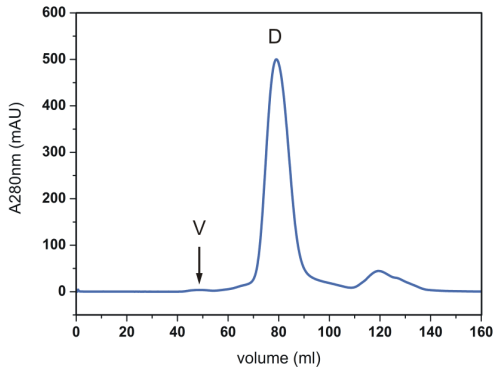


Figure 2.2: Analytic gel filtration of AcsD. Elution profile of AcsD shows a prominent peak at 80 mL. The observed retention volume indicates that AcsD is a dimer (D). Void volume is labeled with V.

2.2 Materials and methods

2.2.1 Overproduction and purification of wt AcsD

The gene encoding AcsD was amplified from cosmid pL9G1 and cloned into pET151/D-TOPO by Prof. Challis' group, University of Warwick. AcsD construct codes for the protein fused with an N-terminal His_6-TEV tag (Figure 2.3).

Figure 2.3: AcsD expression construct. Cleavable His_6-tagged protein construct of AcsD including a spacer (colored in black)

The His_6-tag and tobacco etch virus (TEV) cleaving site are used for protein purification. The AcsD plasmid was transformed in *Escherichia coli* BL21 (DE3). The transformed cells were grown from 1:200 dilutions of overnight cultures in 1 L flasks containing LB medium (Luria-Broth) with a final ampicillin concentration of 100 µg/L in an orbital shaker at 200 rpm at 37 °C. When the cells reached an $A_{600} = 0.6$, 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added and the temperature was lowered to 15 °C. High temperatures led to insoluble over-expressed protein. After a further 18 hour incubation period, cells were harvested by centrifugation 6000 rpm, 4 °C for 15 min (Beckman JLA 8.1000), and resuspended and homogenized in 100-200 mL binding buffer (20 mM sodium phosphate, 40 mM imidazole, 500 mM NaCl, 10 % (v/v) glycerol), 1 mg/mL lysozyme, 40 µg/mL DNAse I and two tablets of EDTA-free protease inhibitor cocktail (Roche Diagnostics). Cell lysis was performed either in a cell disrupter (Constant Systems Ltd) or by sonication (6-8 cycles of 30 s

and 1 min break). Cell debris and unlysed cells were separated by centrifugation at 20,000 rpm (rotor SS-34) for 30 min at 4 °C.

The first purification step of His_6-AcsD was performed by immobilized metal affinity chromatography (IMAC) with Ni-beads. Therefore 3 mL of column material (6 mL of slurry) was added to supernatant (S/N) and incubated for 15 min and gently stirred, to allow binding of His_6-tagged protein to the nickel-beads. Unbound protein was removed by washing with 50-100 column volumes (CV) of binding buffer (20 mM sodium phosphate, 40 mM imidazole, 500 mM NaCl, 10 % (v/v) glycerol). His_6-AcsD was eluted in three steps with 2-3 CV elution buffer (20 mM sodium phosphate, 500 mM imidazole, 500 mM NaCl, 10 % (v/v) glycerol). Eluted protein solutions were desalted using a HiPrep 26/10 desalting column (Amersham Bioscience) with 20 mM Tris pH 7.5, 500 mM NaCl, 10 % (v/v) glycerol. The His_6-tag was cleaved with a 1:50 dilution of 1 mg/mL TEV-protease (O/N at RT) and removed by a 5 mL HisTrap HP column (Amersham Bioscience). In a final purification step, impurities were separated from AcsD protein via gel filtration (GF), using a Superdex 200 or Sephacryl 200 column (GE Healthcare) with GF buffer (50 mM Tris pH 7.5, 500 mM NaCl, 10 % (v/v) glycerol). After tag removal AcsD comprises six additional residues at the N-terminus (GIDPFT) with a final weight of 71 140 Da. The purified protein was concentrated to 6 or 9 mg/mL using protein concentrating devices (50 kDa Vivaspin), frozen in liquid nitrogen and stored at – 80 °C. The protein yield of AcsD *wt* per liter cell culture was 1.8 mg pure protein. Purification purity and integrity was monitored by SDS gel electrophoresis using 4-12 % SDS-polyacrylamide gels from Invitrogen. The identity and integrity of protein was confirmed by mass spectrometry.

2.2.2 Co-complex crystallization of AcsD

The initial crystal condition for apo AcsD was 0.1 M TRIS–HCl pH 8.5 and 1.0 M sodium tartrate (McMahon et al. 2008). Attempts to co-crystallize in this condition failed. New crystallization conditions for AcsD co-complex crystals were identified by rescreening with a variety of commercially available crystallization solutions (The Classics, JCSG+, Pegs, Wizard I+II crystallization screens from QIAGEN) using three different final protein concentrations as described for the apo crystallization (McMahon et al. 2008). Successful conditions and optimization are discussed in subsequent sections for each co-complex individually. A crystal gallery is shown in Figure 2.15.

2.2.2.1 Attempt to obtain a citrate co-complex with ATP-γ-S

For citrate complex formation a new crystal condition (20 % (w/v) polyethylene glycol (PEG) 8000 0.1 M TRIS-HCl pH 8.5) was obtained and systematically optimized in conventional hanging drop grid screens. Prior to crystal droplet setup, AcsD was incubated with 2 mM ATP-γ-S (Sigma A1388) at room temperature for 10 min and centrifuged at 4 °C for 5 min to remove potential aggregates. This step was performed to bind ATP-γ-S and citric acid (from the optimized crystallization condition) in the active site of AcsD. Complex crystals grew in hanging drops from equal mixtures (3 µL each) of 0.1 M TRIS-HCl pH 7.8, 26 % (w/v) PEG 8000, 300 mM sodium citrate tribasic, 2 mM ATP-γ-S (Sigma) and 6 mg/mL AcsD to full size (~0.2 x 0.4 mm) in one week at 20 °C.

2.2.2.2 ATP co-complex

The complex with ATP was obtained using 0.1 M TRIS-HCl pH 7.8, 26 % (w/v) PEG 8000 and 300 mM L-serine (Sigma S4500). Protein (6 mg/mL) was pre-incubated with 5 mM of ATP (dissolved in water with adjusted pH 7.8) at room temperature for 10 min. Forming protein/ATP precipitate was removed by centrifugation in a table top centrifuge (14'000 rpm for 5 min at 4 °C). 1 µL of supernatant and equal amount of precipitant was used in hanging drop crystal trails. Co-complex crystals grew slower to full size (~0.2 x 0.4 mm) within 2-3 weeks at 20 °C.

2.2.2.3 AMP - sulfate co-complex

The complex with AMP was obtained using equal volumes of protein (9 mg/mL) and 0.1 M sodium cacodylate pH 6.5, 29 % (w/v) PEG 8000 and 200 mM $(NH_4)_2SO_4$ and 10 mM AMP. Crystals grew within days and reached full size (~0.08 x 0.11 mm) within two weeks at 20 °C.

2.2.2.4 Citrate, adenosine and sulfate co-complex

The citric acid, adenosine and sulfate co-complex crystals were obtained in hanging drops from equal mixtures (3 µL each) of AcsD (9 mg/mL) and precipitant (0.1 M sodium cacodylate pH 6.5, 29 % (w/v) PEG 8000 and 200 mM $(NH4)_2SO_4$ 10 mM sodium citrate, 5 mM adenosine). Crystals grew fast within several days and reached their full size (> 0.1 x 0.15 mm) within two weeks at 20 °C.

2.2.2.5 Citryl-ethylenediamine co-complex

For co-crystallization with *N*-citryl-ethylenediamine 9 mg/mL *wt* AcsD (stored in protein buffer: 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 10 % (v/v) glycerol) was incubated for one hour with 15 mM ethylenediamine, 10 mM $MgCl_2$, 15 mM ATP and 15 mM citrate. The supernatant was used to grow *N*-citryl-ethylenediamine co-complex crystals in hanging drops from equal mixtures (2+2 µL) with 0.1 M HEPES pH 7.2, 17 % (w/v) PEG 8000, 7.5 % (v/v) glycerol at 20 °C. Crystals started growing after a few days and grew within two weeks to large crystals with a size of approximately 0.2 x 0.6 mm.

2.2.3 Structural biology: data collection and refinement

Data collection was carried out on different beam lines at the European Synchrotron Radiation Facility (ESRF, Grenoble). The 2.25 Å native structure (3FFE) was used to phase all co-complex structures of AcsD. Structures were refined using REFMAC5 (Murshudov et al. 1997) with or without TLS (Winn et al. 2001), NCS restraints and WinCoot (Emsley and Cowtan 2004) for manual manipulation.

Electron density maps (F_oF_c and $2F_oF_c$) for each ligand are shown in Figure 2.4. All ligand structures were fully refined without ligand atoms. Water atoms were added and refined before ligand atoms were placed in the active site and refined. Data collection and refinement statistics are presented in Table 2.1. It was not possible to locate residues 1-6, 99, 572-577 and 588-626 in the experimental maps. The ATP and citrate (overlaps with ATP) data sets were solved in the same space group as the apo-structure ($P2_12_12_1$). All other structures were solved in P1. Refinement was

validated using the structure validation service MolProbity (Davis et al. 2007) to ensure that clash scores, rotamer outliers, Ramachandran outliers and C_β-deviations are within acceptable range for the resolution. The following PDB codes were deposited in Protein Data Bank (PDB): 2W02 (ATP), 2W04 (citrate artifact), 2W03 (citrate, adenosine, SO_4^{2-}).

2.2.3.1 Citrate co-complex structure

Data of the citrate complex crystal were collected on ID23-1. Crystals were cryo-protected by soaking the crystal in 20 % (v/v) PEG 400 and cryo-cooled at -173 °C. Two data sets to 2.2 Å were collected at a beam wavelength of 0.9803 Å (energy of 12.648 keV). The first data set was collected with an exposure time of 1.0 second at 5 % beam transmission over 120° starting at 0.0° phi with an oscillation range of 0.5° and 0.0° overlap. The second data set was collected from phi 60° to 120°. In both cases the detector distance was set to 321 mm. Both data sets were indexed and integrated separately using “mosflm” and combined and scaled with "Scala" from the CCP4 package (Collaborative Computational Project, number 4. 1994) to 2.8 Å. The AcsD apo-structure was used as a model for molecular replacement with “auto MolRep” (CCP4). The structure was refined in several cycles with seven NCS restraints and one TLS group per monomer which was identified by the TLS motion determination server (Painter and Merritt 2006). Only electron density for a citrate and not for ATP-γ-S was found in the active site.

2.2.3.2 ATP co-complex structure

Data of the ATP complex crystal were collected on ID14-2. Crystals were cryo-protected with 20 % (v/v) PEG 400 and cryo-cooled at -173 °C. Two data sets at 2 Å (high resolution pass) and 3 Å (low resolution pass) were collected at a beam wavelength of 0.933 Å (energy of 13.289 keV). The high resolution data set was collected with an exposure time of 5.0 s at 100 % beam transmission over 180° starting at 280.0° phi with an oscillation range of 0.5° and no overlap. The low resolution data collection was carried out in the same manner as the high resolution data set, but with 3 s exposure time. Both data sets were indexed and integrated using XDS (Kabsch 1993), and combined and scaled with "XSCALE" to 2.2 Å. The structure was refined with four NCS restraints and three TLS groups per monomer (Painter and Merritt 2006).

2.2.3.3 AMP – sulfate co-complex structure

Data of the AMP – sulfate co-complex crystal were collected on ID14-2. Crystals were cryo-protected with 20 % (v/v) PEG 400. Two data sets were collected at a beam wavelength of 0.933 Å (energy of 13.289 keV) with the detector edge at 2 Å (distance 182.1 mm) starting at a phi angle of 0° to 100° and 95° to 220°, respectively. An exposure time of 5.0 s at 100 % beam transmission, 100x100 µm beam size and an oscillation range of 0.5° (no overlap) were used. Data were indexed and integrated using XDS, and combined and scaled with "XSCALE" to 2.5 Å (Kabsch 1993).

2.2.3.4 Adenosine – citrate – sulfate co-complex structure

Data of the adenosine – citrate - sulfate co-complex was collected at 2.6 Å (d = 319.7 mm) at ID23-2. Crystals were cryo-protected with 20 % (v/v) PEG 400 and cryo-cooled at -173 °C. An exposure time of 0.8 s at 10 % beam transmission was used to collect 105° starting at 20.0° phi with an oscillation range of 0.5° and no overlap. Beam wavelength was λ = 0.8726 Å (14.209 keV) with a beam size of 200x200 µm. Data set was indexed and integrated using XDS and scaled with "XSCALE" to 2.95 Å. The structure was refined with four NCS restraints and three TLS groups per monomer (Painter and Merritt 2006).

2.2.3.5 Citryl-ethylenediamine co-complex

Two data sets for the ethylenediamine co-complex were collected on ID29 to 1.9 Å. Crystals were cryo-protected with 0.1 M HEPES pH 7.2, 19 % (w/v) PEG 8000, 20 % (v/v) glycerol and cryo-cooled at -173 °C. Both data sets were collected at a beam wavelength of 0.979 Å (energy of 12.664 keV) with an exposure time of 0.2 s at 8 % beam transmission over 130° starting at -20° phi and 200° starting at 100°, respectively. In both cases an oscillation range of 0.5° with 0.0° overlap was chosen. The detector distance was set to 271.3 mm. Data were indexed and integrated separately using HKL2000 and scaled together using "SCALEPACK" (part of HKL2000 (Otwinowski and Minor 1997)) to 2 Å.

Table 2.1: Data and refinement statistics of AcsD complex structures (including apo structure solved in SPorT)

	Apo	**Citrate (overlaps with ATP)**	**ATP**	**AMP-SO_4^{2-}**	**Adenosine–citrate –SO_4^{2-}**	**Citryl-EDA-ATP**
Data collection						
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	P1	P1	P1
Cell dimensions (Å), (°)	*a*= 80.3, *b*= 95.7, *c*= 161.1	*a*= 80.4, *b*= 95.6, *c*= 160.3	*a*= 79.9, *b*= 94.5, *c*= 57.8	*a*=57.7, *b*=70.0, *c*=96.0 *α*=95.0, β =102.0, *γ*=94.7	*a*=57.6, *b*=69.1, *c*=94.2 *α*=95.4, β =101.5, *γ*=95.1	*a* = 57.7 *b* = 71.5 *c* = 95.6 α = 97.3 β = 102.0 γ =91.0
Resolution (Å)* (high resolution)	50.0-2.25 (2.38-2.25)	81.6 -2.8 (2.95-2.8)	81.1-2.2 (2.95-2.2)	39.7-2.5 (2.6-2.5)	45.9 - 2.95 (3.11-2.95)	50.0-2.00 (2.07-2.0)
R_{merge}	9.8 (40.2)	15.7 (40.7)	11.4 (38.1)	7.5 (29.6)	14.3 (44.5)	6.4 (35.0)
I/σ(I)	11.5 (3.2)	10.6 (3.0)	15.7 (5.9)	5.6 (1.7)	6.2 (1.9)	12.8 (2.7)
Completeness (%)	94.4 (91.5)	99.4 (99.9)	99.9 (99.9)	96.58 (94.6)	98.2 (98.5)	93.8 (74.7)
Aver. redundancy	3.0 (3.0)	4.2 (4.3)	10.1 (7.3)	2.6	2.0 (2.0)	3.1 (2.7)
V_m(Å^3/Da)	2.2	2.2	2.1	2.7	2.6	2.7
Solvent (%)	43	43	42	54	53	55
Software	XDS/XSCALE	mosflm/Scala	XDS/XSCALE	XDS/XSCALE	XDS/XSCALE	HKL2000
Refinement						
Unique reflections	56200	30736	61355	46179	29165	89038
R_{work} / R_{free}	18.6 / 24.1	20.2 / 28.0	21.6 / 27.7	20.5/26.1	20.2 / 26.7	19.0 / 23.1
No. atoms						
Protein	9250	9215	9252	9192	9361	9484
Water	505	130	277	169	73	571
Citrate	0	13	0	0	13	0
ATP	0	0	62	0	0	62
AMP	0	0	0	46	0	0
Adenosine	0	0	0	0	38	0
L-serine	0	0	7	0	0	0
Mg	0	0	2	0	0	2
SO_4	0	0	0	10	10	0
glycerol	0	0	0	0	0	30
citryl-EDA	0	0	0	0	0	16
B-factors (Å^2)						
Protein	17	12	12	48	20	38
Water	25	12	12	44	9	42
Citrate	0	4	9	0	27	0
ATP	0	11	12	0	0	35
AMP	0	0	0	49	0	0
Mg	0	0	23	0	0	36
L-serine	0	0	13	0	0	0
Adenosine	0	0	0	0	39	0
SO_4	0	0	0	47	23	0
glycerol	0	0	0	0	0	32
citryl-EDA	0	0	0	0	0	47
Rmsd bonds(Å)/angles (°)	0.008/1.099	0.013/1.461	0.01/1.330	0.015/1.563	0.016/1.747	0.010/1.260
B-factor dev. Bond/angle (Å^2)						
Main chain	0.657/0.750	0.451/0.783	0.495/0.823	0.755/1.422	0.568/1.031	0.750/1.256
Side chain	1.313/2.033	1.096/1.823	1.278/1.939	2.101/3.479	1.539/2.561	1.733/2.705
MolProbity score (percentile)	98	95	95	95	83	95
PDB code	3FFE	2W02	2W04		2W03	

* Values in parentheses refer to the highest resolution shell

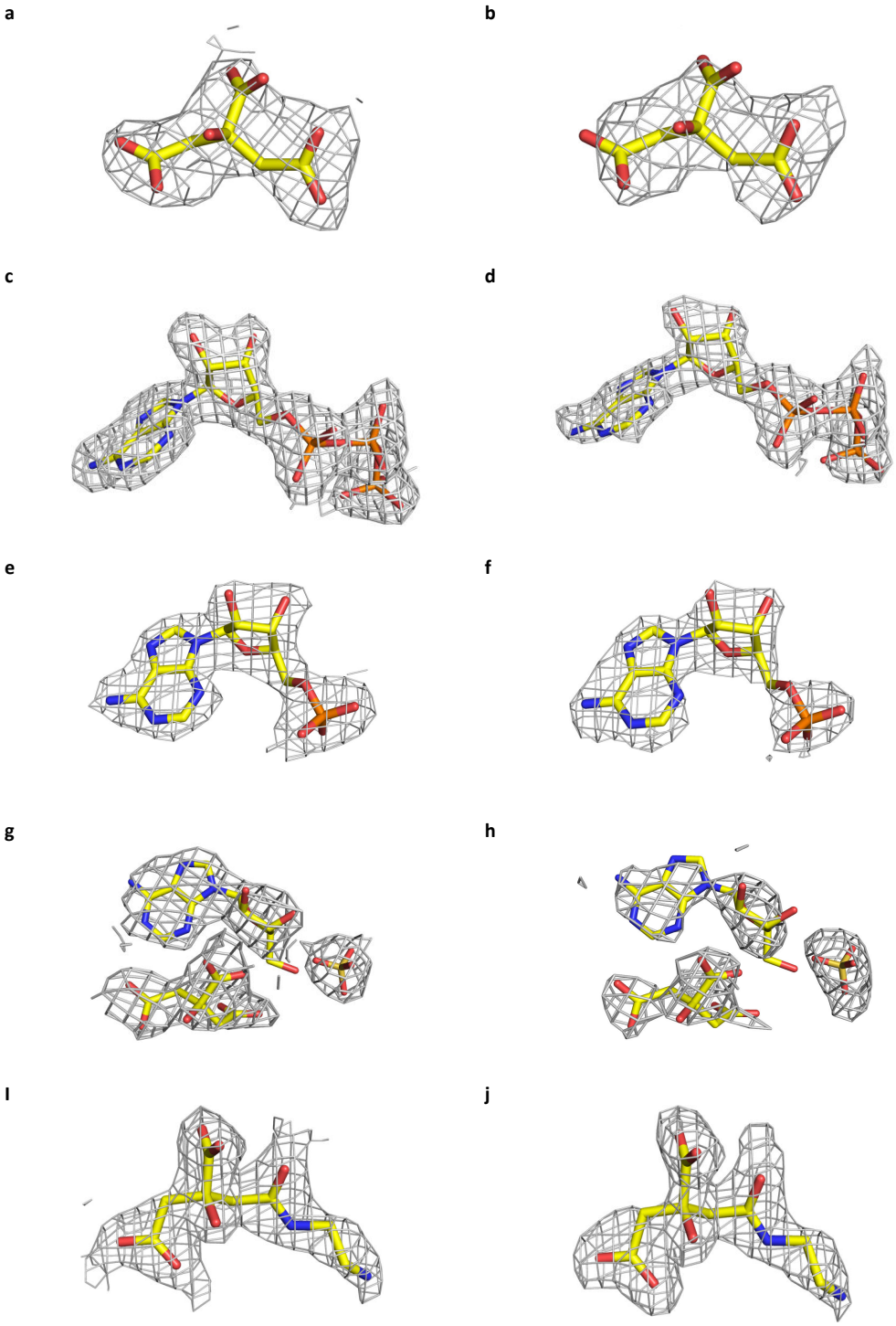


Figure 2.4: Electron density of co-complex structures. **(a)** $2F_oF_c$ map of citrate (overlaps with ATP) contoured at 1.0 σ. **(b)** F_oF_c map of citrate (overlaps with ATP) contoured at 2.5 σ. **(c)** $2F_oF_c$ map of ATP contoured at 1 σ. **(d)** F_oF_c map of ATP contoured at 2.5 σ. **(e)** $2F_oF_c$ map of AMP contoured at 1.0 σ. **(f)** F_oF_c map of AMP contoured at 2.5 σ. **(g)** $2F_oF_c$ map of adenosine, sulfate and citrate contoured at 1.0 σ. **(h)** $2F_oF_c$ map of adenosine, sulfate and citrate contoured at 2.5 σ. **(i)** $2F_oF_c$ electron density map of *N*-citryl-ethylenediamine at 1 σ. **(j)** F_oF_c electron density map of *N*-citryl-ethylenediamine at 2.5 σ.

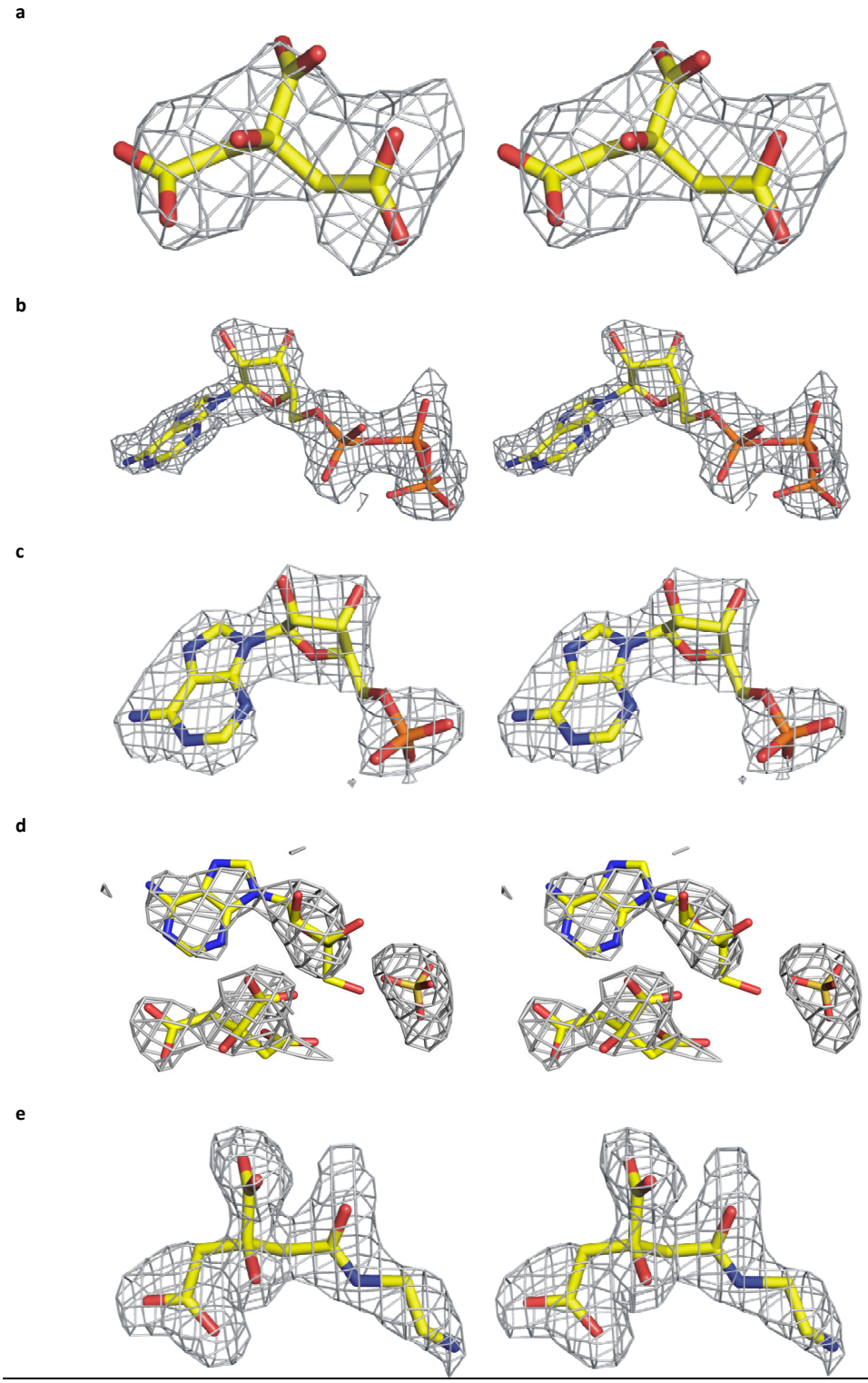


Figure 2.5: Stereo view of F_oF_c electron density maps of AcsD co-complex structures. **(a)** Citrate (overlaps with ATP) **(b)** ATP. **(c)** AMP. **(d)** Adenosine, sulfate and citrate. **(e)** *N*-citryl-ethylenediamine. All maps are contoured at 2.5 σ.

2.2.4 Single mutation of AcsD residues

2.2.4.1 Primer design and site directed mutagenesis

Site directed mutations were generated following the QuikChange (Stratagene) protocol with primers shown in Table 2.2. Mutations were inserted by 20 up to 40 bases long primers carrying a central designated mutation. Polymerase chain reaction (PCR) was used to amplify the *wt* template using these synthesized primers to insert the right mutation. In a final step the methylated *wt* template was removed by digestion with *Dpn*I, a DNAse specific for methylated DNA. All primers were synthesized by Eurogentec and stored at – 20 °C. While PFU was used as standard DNA polymerase, in certain cases the use of Phusion Hot Start High-Fidelity DNA polymerase (Finnzymes) was successful, as indicated in Table 2.2. Protocols for reaction mix and PCR programs are shown in Table 2.3 and Table 2.4, respectively. Due to problems with PCR amplification, primers for mutants 13, 14 and 21 were constructed with a modified QuikChange protocol (Liu and Naismith 2008). Here the primers are constructed with 5'-overhangs to ensure exponential amplification. PCR for these mutants was carried out with PFU DNA polymerase and a modified reaction schedule (Table 2.5). After successful amplification, the *wt* DNA template was digested with 10 units of *Dpn*I in a water bath (37 °C, 1 h). All PCR products were purified using Montage Single Sample PCR Cleanup column (Millipore) as described in the kit manual. After transformation with calcium chloride competent *E. coli* DH5α cells (45 sec at 42 °C), samples were incubated in an orbital shaker (90 min, 37 °C, 200 rpm) and plated on LB agar plates containing ampicillin. Overnight cultures of

plated clones (10 mL LB, 37 °C) were prepared and plasmid DNA extracted using the Wizard® Plus SV Minipreps DNA Purification System (Promega). Each mutation was verified by the sequencing service of the University of Dundee (http://www.dnaseq.co.uk/). Prior protein expression mutated plasmids were transformed into BL21 (DE3) *E. coli* cells.

Table 2.2: List of site-directed mutations in AcsD; PCR was carried out with DNA polymerases (pol) PFU (P) or Phusion Hot start High-Fidelity DNA Polymerase (H) (Finnzymes) as indicated. All successful steps in mutating, purifying and testing of each mutant are marked with (✓) or otherwise as failed.

position	no	mutation	codon	new codon	primer DNA sequence	mutated	pol	purified	assayed
H170	1	A	CAC	GCC	fw GATTGTGGTTCGGCGCCCCAGCCACCCCGCGC rv GCGCGGGGTGGCTGGGGGCGCCGAACCACAATC	✓	P	✓	✓
E193	2	A	GAA	GCG	fw GCAGTGGGCGCCTGCGTTTCAGGCCCGCG rv CGCGGGCCTGAAACGCAGGCGCCCACTGC	failed			
	3	Q	GAA	CAG	fw GCAGTGGGCGCCTCAGTTTCAGGCCCGCG rv CGCGGGCCTGAAACTGAGGCGCCCACTGC	✓	P	✓	✓
	4	D	GAA	GAC	fw CAGTGGGCGCCTGACTTTCAGGCCCGC rv GCGGGCCTGAAAGTCAGGCGCCCACTG	failed			
H242	5	N	CAC	AAC	fw GCCATCATCTGTATGAACCCGGTGCAGGCGCAAC rv GTTGCGCCTGCACCGGGTTCATACAGATGATGGC	✓	P	✓	✓
	6	A	CAC	GCG	fw CGCCATCATCTGTATGGCGCCGGTGCAGGCGCAAC rv GTTGCGCCTGCACCGGCGCCATACAGATGATGGCG	failed			
S279	7	A	TCT	GCG	fw GCCAGCCCGACAGCCGCGATTCGTACCTGGTTC rv GAACCAGGTACGAATCGCGGCTGTCGGGCTGGC	✓	H	✓	✓
F291	8	Y	TTC	TAT	fw GACGATCACGACTACTTCATCAAAGGTTCGCTC rv GAGCGAACCTTTGATGAAGTAGTCGTGAGCGTC	✓	P	✓	✓
T301	9	A	ACC	GCC	fw GAATGTGCGCATCGCCAACTGTGTGCGC rv GCGCACACAGTTGGCGATGCGCACATTC	✓	P	✓	✓
R305	10	A	CGC	GCC	fw CATCACCAACTGTGTGGCCAAAAACGCCTGGTATG rv CATACCAGGCGTTTTTGGGCACACAGTTGGTGATG	✓	P	✓	✓
	11	K	CGC	AAG	fw GCGCATCACCAACTGTGTGAAGAAAAACGCCTGGTATGAAC rv GTTCATACCAGGCGTTTTTCTTCACACAGTTGGTGATGCGC	✓	P	✓	✓
E442	12	A	GAG	GCG	vw CACGGCGTGGTGATGGCGCCGCACCTGCAAAAC rv GTTTTGCAGGTGCGGCGCCATCACCACGCCGTG	failed			
	13	Q	GAG	CAG	vw GCGTGGTGATGCAGCCGCACCTGCAAAACAGCGTG rv GGCTGCATCACCACGCCGTGATTGAAGAACAGCGAC	✓	P	✓	✓
	14	D	GAG	GAT	vw GCGTGGTGATGGATCCGCACCTGCAAAACAGCGTG rv GGATCCATCACCACGCCGTGATTGAAGAACAGCGAC	✓	P	✓	✓
H444	15	A	CAC	GCG	fw GTGGTGATGGAGCCGGCGCTGCAAAACAGCGTGC rv GCACGCTGTTTTGCAGCGCCGGCTCCATCACCAC	✓	H	✓	✓
	16	N	CAC	AAC	fw GTGGTGATGGAGCCGAACCTGCAAAACAGCGTGC rv GCACGCTGTTTTGCAGGTTCGGCTCCATCACCAC	✓	P	✓	✓
D464	17	A	GAT	GCG	fw CAGGTGCTGCTGCGCGCGTTCGAAGGCGTAAAACTG rv CAGTTTTACGCCTTCGAACGCGCGCAGCAGCACCTG	✓	P	✓	✓
	18	N	GAT	AAC	vw CAGGTGCTGCTGCGCAACTTCGAAGGCGTAAAACTG rv CAGTTTTACGCCTTCGAAGTTGCGCAGCAGCACCTG	✓	P	failed	
R501	19	A	CGC	GCC	fw CGAACAGGGCTGGAATGCCATCATGTACTGCCTG rv CAGGCAGTACATGATGGCATTCCAGCCCTGTTCG	✓	P	✓	✓
	20	K	CGC	AAA	fw CCGCGAACAGGGCTGGAATAAAATCATGTACTGCCTGTTC rv GAACAGGCAGTACATGATTTTATTCCAGCCCTGTTCGCGG	✓	P	✓	✓
K563	21	A	AAA	GCG	fw CGGTAGCGTGCGCGACCAACCTTAAGGTCCGGCTGG rv AGGTCGCGCACGCTACCGGATGGCCGGCGATCAGC	✓	P	✓	✓
R576	22	A	CGT	GCG	vw GCGGCCGAGGCCGATGCGCAGGCCAGCTATGTTC rv GAACATAGCTGGCCTGCGCATCGGCCTCGGCCGC	✓	P	✓	✓
	23	K	CGT	AAA	vw GGCGGCCGAGGCCGATAAACAGGCCAGCTATGTTCG rv CGAACATAGCTGGCCTGTTTATCGGCCTCGGCCGCC	✓	P	✓	✓

Table 2.3: PCR reaction mix for site directed mutations in AcsD. **(a)** Reaction mix with Phusion Hot start Fidelity DNA polymerase. **(b)** Reaction mix with PFU DNA polymerase.

a

solution	vol
5x Phusion reaction buffer	10 µL
primer forward (10 µM)	1.5 µL
primer reverse (10 µM)	1.5 µL
wt DNA (40 ng/µL)	1 µL
dNTPs (10mM)	1 µL
DMSO (100 (v/v)%	1 µL
water	34.5 µL
DNA Polymerase (3 U/µL)	1 µL
reaction volume	50 µL

b

solution	vol
10x reaction buffer	5 µL
primer forward (10 µM)	1 µL
primer reverse (10 µM)	1 µL
wt DNA (25 ng/µL)	2 µL
dNTPs (10mM)	1 µL
water	40 µL
DNA Polymerase (2.5 U/µL)	1 µL
reaction volume	50 µL

Table 2.4: PCR reaction time table for site directed mutations with Phusion Hot start Fidelity DNA polymerase **(a)** and PFU DNA polymerase **(b)**

a

step	cycle	temp.	time
1	1x	95 ℃	5 min
2	16x	95 ℃	30 sec
		50 ℃	30 sec
		72 ℃	8 min
3	1x	72 ℃	7 min
		4 ℃	∞

b

step	cycle	temp.	time
1	1x	95 ℃	2 min
2	16x	95 ℃	1 min
		57 ℃	45 sec
		72 ℃	15 min
3	1x	72 ℃	7 min
		4 ℃	∞

Table 2.5: Site directed mutations for primers 13, 14, 21 using PFU DNA polymerase. **(a)** PCR reaction mix. **(b)** Program used for PCR amplification.

a

solution	vol
5x PFU reaction buffer	10 µL
primer forward (10 µM)	1.15 µL
primer reverse (10 µM)	1.15 µL
wt DNA (21 ng/µL)	0.5 µL
dNTPs (10mM)	1 µL
water	34.5 µL
PFU (2.5 U/µL)	1 µL
reaction volume	50 µL

b

step	cycle	temp.	time
1	1x	95 ℃	5 min
2	11x	95 ℃	1 min
		58 ℃	1 min
		72 ℃	16 min
3	2x	95 ℃	1 min
		50 ℃	1 min
		72 ℃	16 min
4	1x	72 ℃	7 min
		4 ℃	∞

2.2.4.2 Expression and purification of mutated AcsD

Cultures of *E. coli* BL21 (DE3) (Novagen) containing AcsD point mutations were launched in 100 mL LB /Amp overnight at 37 °C. These were diluted 1:200 into 3x1 L flasks containing LB medium with a final ampicillin concentration of 100 µg/L and incubated in an orbital shaker at 250 rpm. When cultures reached an A_{600} = 0.8, 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added and the temperature lowered to 15 °C for a further 18 hour incubation period. Cell harvest and the first purification step were carried out as described in section 2.2.1. The His_6-tag of most mutants was not removed (enzyme activity showed no differences between tagged and non-tagged protein version). An additional desalting or dialysis step (desalting buffer: 50 mM TRIS-HCl pH 7.5, 500 mM NaCl, and 10 % (v/v) glycerol) was applied to remove imidazole, which was introduced in the IMAC purification step. Mutated AcsD and His_6-AcsD was concentrated to 2 – 9 mg/mL using protein concentrating devices (50 kDa Vivaspin); flash frozen in liquid nitrogen and stored at - 80 °C. Protein yields varied and were in the range of 1 – 2 mg pure protein per liter cell culture.

2.2.5 AMP production assay

The activity of *wt* and mutated AcsD were tested in a modified coupled enzyme assay developed on basis of Wu and Hill (Wu and Hill 1993). As described in section 1.5, AcsD liberates adenosine monophosphate by catalyzing the carboxylation of ethanolamine or L-2,4-diaminobutyric acid (DABA) with one of the prochiral carboxyl groups of citric acid (Challis 2007). Preliminary tests showed that AcsD uses ATP as

energy source and also performs a condensation reaction between citric acid and other nucleophiles like L-serine (Challis et al, unpublished data). The AMP production assay is based on coupling the liberation of AMP to the lactate dehydrogenase (LDH) oxidation of nicotinamide adenine dinucleotide (NADH). Figure 2.6 shows the enzymes and substrates used in the assay. The formed AMP (by the condensation reaction of citric acid and ethanolamine or DABA catalyzed by AcsD) is converted via myokinase (MK) into two ADP molecules using one ATP. In the following step two ATP molecules are regenerated by pyruvate kinase (PK), which catalyses dephosphorylation of phosphoenol pyruvate (PEP) to pyruvate. In the final reaction LDH oxidizes NADH by reducing pyruvate to lactate. The decrease in NADH levels can be monitored whether by absorbance spectroscopy at 340 nm or by fluorescence spectroscopy at 376 nm excitation and 462 nm emission wavelength.

The reaction contained 50 mM TRIS-HCl buffer (pH 8.0), 3 mM ATP, 15 mM $MgCl_2$, 1.5 mM phosphoenol pyruvate, 0.25 mM NADH, 2 µM *wt* or mutant His_6-AcsD, 12.6 units of lactate dehydrogenase, 8.4 units of pyruvate kinase, 4 units of myokinase, 75 mM L-serine and 2 mM citric acid at a total volume of 140 µL. The assay was carried out in triplicate using a Cary Eclipse spectrometer (Varian) and Perkin-Elmer 1 mL cuvettes. The mixture was incubated at 20 °C for 5 min and the reaction started by adding of 2 µM *wt* or mutated AcsD. Incubation times varied from 20 minutes (AcsD) up to 12 h (mutants). The decrease in NADH fluorescence (excitation 376 nm and emission 462 nm) in real time was monitored. For the ADP production assay, the same procedure was used, but myokinase was omitted from reactions. All data were processed using Excel 2003 and Origin 7 (OriginLab).

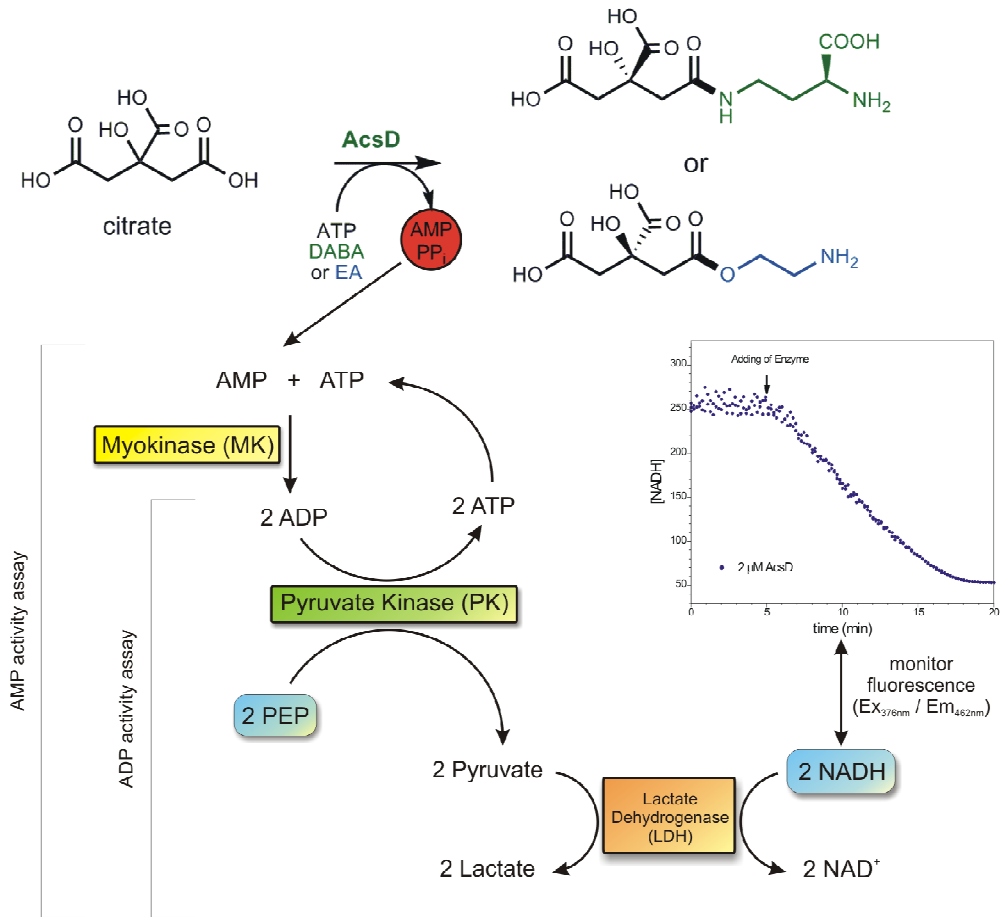


Figure 2.6: Schematics for AMP activity assay for detecting and monitoring of AcsD enzyme activity. AMP, liberated by AcsD reaction, is used as starting substrate in an enzyme coupled reaction to oxidize NADH. PEP and NADH are consumed, ATP is regenerated keeping its concentration constant. Abbreviations: DABA: L-2,4-diaminobutyric acid; EA: ethanolamine; PEP: phosphoenolpyruvate; NADH/NAD^+: reduced and oxidized form of nicotinamide adenine dinucleotide

2.2.6 Kinetic studies

To obtain kinetic data, the AMP production assay was carried out as described in the previous section. The reaction contained 50 mM TRIS-HCl buffer (pH 8.0), 3 mM ATP, 15 mM $MgCl_2$ (30 or 105 mM), 1.5 mM phosphoenol pyruvate, 0.25 mM NADH, 2 µM protein, 12.6 units of lactate dehydrogenase, 8.4 units of pyruvate kinase, 4 units of myokinase, 75 mM L-serine and various concentration of citrate (buffered at pH 8.0).

The total volume per reaction was 140 µL. Reactions were pre-incubated at 20 °C for 5 min and started by adding AcsD to a final concentration of 2 µM. Kinetic data were converted and fitted using Origin 7 (OriginLab) with the Michaelis-Menten equation $v = v_{max} \cdot [S] \cdot ([S] + K_M)^{-1}$ (see also section 2.3.2.5).

2.2.7 Activity test for nucleophilic substrates

AcsD substrate specificity (*wt* and mutated protein) for various nucleophiles were tested using the previous described AMP activity assay (see section 2.2.5). Reactions were prepared as usual, but contained 75 mM of the respective nucleophile. Samples were pre-incubated for 5 min at 20 °C and reaction started with 2 µM *wt* or mutated AcsD. Fluorescence data were acquired corresponding to reactions with L-serine. Each data set was normalized by dividing starting intensity value when enzyme was added (at 5 min). Slopes of normalized data were determined with Origin 7 (OriginLab).

2.2.8 Mass spectroscopic analysis citric acid derivatives

AcsD products with various nucleophiles were confirmed by MS and MS/MS analysis. Enzymatic reactions were carried out in ammonium acetate buffer (10 mM, pH 8.0) containing 10-30 µM *wt* AcsD, 30 mM $MgCl_2$, 20 mM citrate and 40 mM of various nucleophiles (each buffered at pH 8). The mixtures were incubated at 20 °C and the reactions were initiated by adding 20 mM ATP.

At certain time points 1 µL aliquots were removed, diluted into 50:50 acetonitrile: water (100 µL) and centrifuged at 12,000 rpm for 1 min. The samples were analyzed using continuous syringe infusion (Harvard syringe pump, Harvard Apparatus, Kent,

UK) into a LCT (Micromass, Manchester, UK) ESI MS instrument or a Q-Star XL (Applied Biosystems, Foster City, USA) ESI-QTOF instrument using the appropriate ionization mode. *M/z*s of interest were subjected to collisionally induced dissociation (CID) on the Q-Star XL instrument, with collision energy adjustment to give some intact signal. The operation of mass spectrometry instruments and acquisition of data were carried out by Dr Catherine H Botting (BMS Mass spectrometry and Proteomics Facility, University of St Andrews).

2.3 Results

2.3.1 Protein purification

2.3.1.1 Immobilized-metal affinity chromatography

Over expressing His_6-AcsD *E. coli* cells were grown, induced and lysed as described in section 2.2.1. The first purification step after cell lysis was carried out using immobilized-metal affinity chromatography (IMAC) (Figure 2.7 a). His_6-AcsD was detectable in the crude protein solution (lane 1) at the level of the 75 kDa standard protein marker (lane S). Expression trails of His_6–AcsD at 37 °C yielded only smaller amounts of soluble protein (data not shown). It was found that this effect could be minimized by decreasing the temperature to 15 °C after induction. However, still a large amount of insoluble protein was found in the cell pellet (lane 3) compared to the supernatant (lane 2). Elution with 500 mM imidazole indicates that most of His_6–AcsD was eluted in the first fraction (lane 6) and less in the later ones (lane 7 and 8). Other than crude protein solution (lane 2) eluted His_6-AcsD (lane 6-8) was relatively pure.

2.3.1.2 Desalting and tag removal

The high imidazole concentration from the IMAC purification had a negative effect on AcsD when kept at RT for several hours. Therefore protein solution was directly desalted on a HiPrep 26/10 column. The elution profile indicates a peak (10 – 15 mL) corresponding to AcsD protein (Figure 2.7c). A smaller peak at 55-70 mL contains the separated ions, such as imidazole and salts. TEV-cleavage was carried out O/N at RT

and verified by SDS-PAGE (Figure 2.7b). Cleavage was successful with a protease to protein ratio of 1:50. TEV protease and cleaved tag were removed by a second IMAC purification step using a 5 mL HisTrap column (GE Healthcare). Protein and washing buffer contained 50 mM imidazole to avoid AcsD binding to the column.

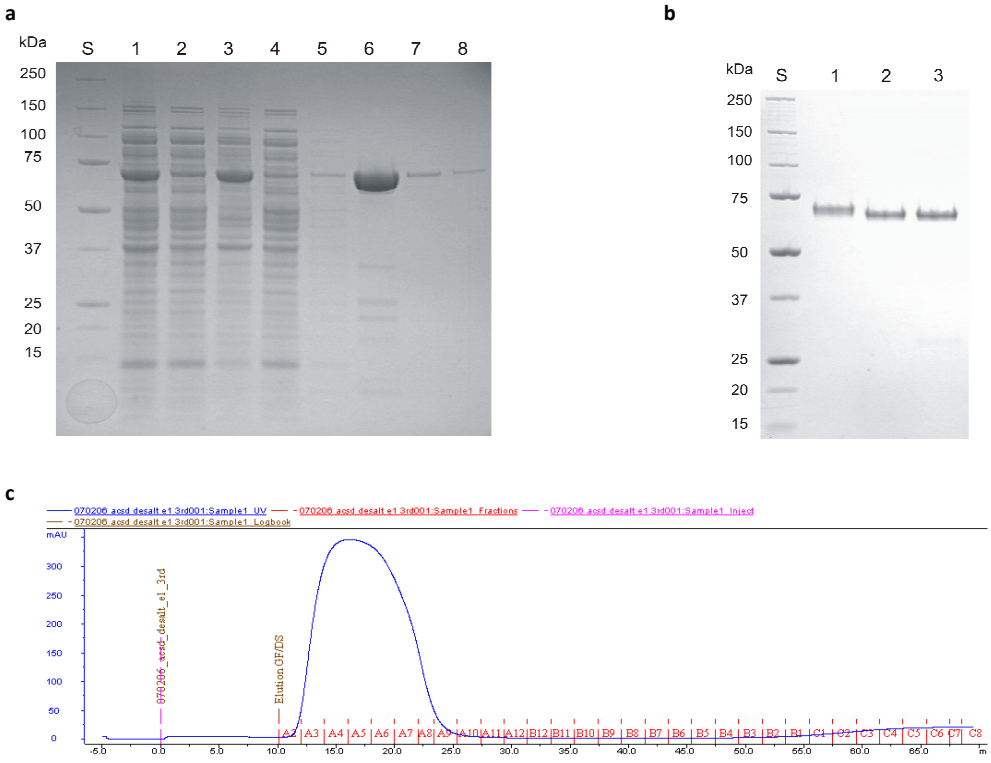
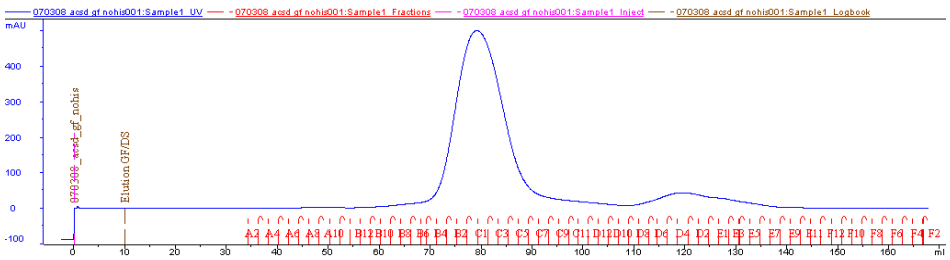


Figure 2.7: First step in His_6-AcsD purification and TEV cleavage. **(a)** First purification of AcsD with immobilized-metal affinity chromatography (IMAC): standard; **1** crude protein solution after cell breakdown; **2** supernatant (S/N); **3** pellet dissolved in 8 M urea; **4** flow through of IMAC column; **5** protein eluted by the washing step; **6** elution I; **7** elution II; **8** elution III; **(b)** Cleaving of His_6-tag with 1:50 TEV protease; **S** standard; **1** His_6-AcsD; **2** cleaved AcsD treated with 1:50 TEV protease ; **3** cleaved AcsD treated with 1:10 TEV protease. **(c)** Elution profile of the desalting step on a HiPrep 26/10 desalting column to remove imidazole used for elution in the IMAC purification step.

2.3.1.3 Gel filtration

In a final gel filtration step, AcsD was separated from remaining impurities, such as potential breakdown products or *E. coli* proteins. The elution profile shows that AcsD elutes from fraction B4 to C6 (Figure 2.8a). Protein containing fractions (10 µL each) were analyzed by SDS-PAGE to examine their purity (Figure 2.8b). Purest fractions (B4-C6) were combined and concentrated to 6 – 9 mg/mL. Fractions with impurities or degraded AcsD (B6 and B5) were discarded. A second smaller peak (D5-E4) contained no detectable protein (Figure 2.8a, b). The high molecular weight band (~140 kDa) in fraction B4-C5 was confirmed by mass spectrometry to be AcsD and is likely to be dimeric.

a



b

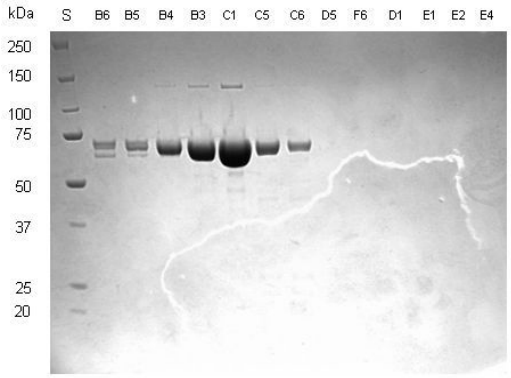


Figure 2.8: Last purification step of AcsD. **(a)** Elution profile of gel filtration step on a Superdex 200 (GE Healthcare) column, the last AcsD purification step. **(b)** SDS-polyacrylamide gels of AcsD gel filtration showing the collected fractions. Highest purity of protein is found in fraction B4-C6.

2.3.2 AcsD activity assay and kinetics

Activity of *wt* and mutated His_6-AcsD was determined in a coupled enzyme assay (see section 2.2.5). The AcsD activity assay is based on coupling the AcsD dependent formation of AMP to the LDH catalyzed oxidation of NADH to NAD^+. The coupled enzymes used in the assay are myokinase (MK), pyruvate kinase (PK) and lactate dehydrogenase (LDH), whereas later two enzymes were purchased combined in one solution. The reduced form of nicotinamide adenine dinucleotide can be excited at 376 nm, which emits at 462 nm, whereas the oxidized form does not. Hence, enzymatic conversion of NADH to NAD^+ can be monitored by fluorescence. Data were used to compare mutant with *wt* enzyme activities and calculate kinetic parameters.

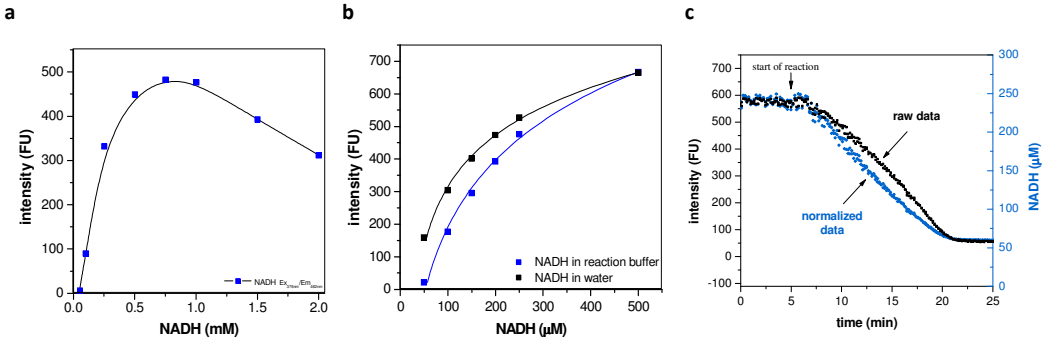


Figure 2.9: AcsD activity assay setup. **(a)** Calibration of the fluorescence based AcsD activity assay. At high NADH concentrations (>1 mM) fluorescence intensity decreases due to fluorescence quenching. **(b)** At lower NADH concentrations the relation is not precisely linear. This effect occurs in water (black) as well as in reaction buffer (minus AcsD) (blue). **(c)** Comparison of raw data (fluorescing units) with normalized data (NADH concentration) shows that normalization improves the linear relationship.

2.3.2.1 Data evaluation

In order to obtain an appropriate concentration regime for accurate NADH detection, reactions lacking AcsD were measured with NADH concentrations up to 2 mM (Figure 2.9a). Above 1 mM NADH fluorescence intensities decreased when more NADH was added, presumably due to self-quenching. The concentration range employed from 50 to 300 µM NADH (in reaction buffer and in water) still showed a non-linear relationship (Figure 2.9b). If one assumes a linear relationship, the fit of kinetic curves was acceptable, but shows deviations from the "normal" shape. Therefore the fluorescence intensity was converted to NADH concentration using a calibration curve, which resulted in a much smoother graph (Figure 2.9c). The regression is exponential and the fitting function, obtained by Origin, was used to re-calculate and convert the raw fluorescent intensity data to NADH concentration using Microsoft Excel (Figure 2.9b). Standard calibration graphs were obtained (on a daily basis before each set of experiments) to ensure reliable conversion. These derived values for NADH concentration improved the fit to standard kinetic curves. Data for kinetics were fitted with Michaelis-Menten equation using Origin software with the background rate (no AcsD) subtracted.

2.3.2.2 Stability and verification of AMP activity assay

Since the spectroscopic assay by Wu and Hill (Wu and Hill 1993) was modified (see section 2.2.5), verification of its liability and usability is important. Initial tests of AcsD reaction at 37 °C were completed within minutes. In reactions lacking AcsD background signal was observed. To obtain more stable and reproducible data, the

reaction temperature was dropped to 20 °C and a 5 min pre-incubation time added and recorded. At this temperature reactions with 2 µM *wt* AcsD were completed within 15 min. No background reaction was detected in samples were AcsD was omitted (Figure 2.10a).

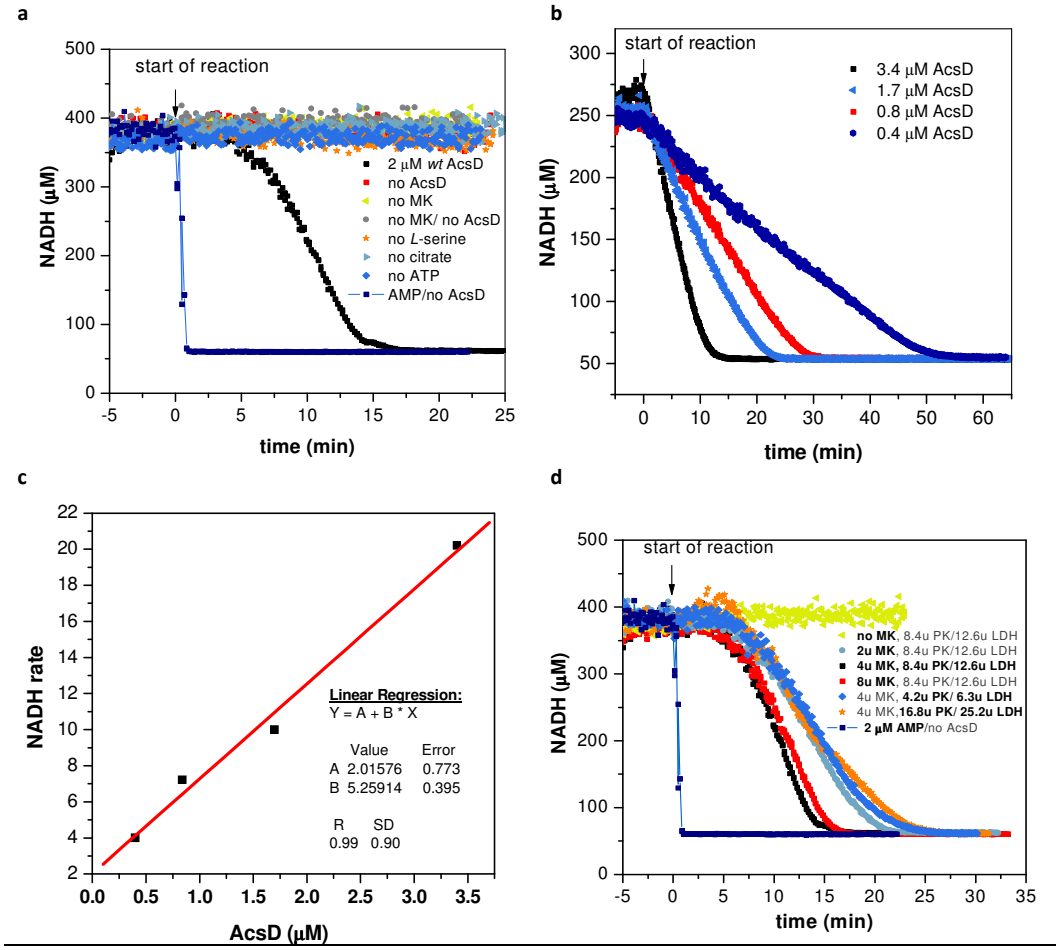


Figure 2.10: Control reactions for AMP activity assay. **(a)** Used negative and positive controls to verify stability of AMP assay. Reactions with 2 µM of AcsD wt protein (black ■) finished (consumption of all NADH) within 15 min. Reactions where L-serine (orange ★), citrate (turquoise ▶), ATP (blue ◆) , AcsD (red ■), myokinase (MK) (yellow-green ◀) or MK/AcsD (grey ●) was omitted, were not active. Samples with 2 µM AMP, but no AcsD (dark blue -■-) were 6 fold faster. **(b)** AMP assay with different concentrations of AcsD provides additional proof that AcsD is rate limiting. **(c)** Reaction rates plotted over AcsD concentration used in (b) show a linear regression. **(d)** Reactions with halved or doubled concentration of myokinase (MK) and pyruvate kinase/ lactate dehydrogenase (PK/LDH) have no significant effect on turnover rates of NADH. Reactions with 1xMK/ 1xPK/LDH (black ■) and 2xMK (red ■) show slight faster turnover rates than reactions with 0.5 MK (turquoise ●), 2x PK/LDH (orange ★) or 0.5x PK/LDH (blue ◆). A negative control without MK (yellow-green ◀) is not active, while the positive control with 2 µM AMP (where AcsD was omitted) is the fastest reaction (dark blue -■-).

When at least one AcsD substrate (either L-serine, citric acid or ATP) was omitted, no significant decrease in NADH levels was detected. Similar observations were made in reactions lacking AcsD and/or myokinase. Addition of exogenous AMP (2 µM) to reactions where AcsD was omitted resulted in a rate of NADH consumption over six fold faster than AcsD (Figure 2.10a). This indicates that the coupled enzymes are not rate limiting even at artificially high concentrations of AMP. Hence, AcsD must be the rate limiting enzyme. This was tested with increasing or decreasing protein concentrations, which reflected in increasing or decreasing NADH turnover rates, respectively (Figure 2.10b). Plotted rates over protein concentrations show a linear regression and indicate that AcsD in this concentration range (0.5-3.5 µM) is not yet saturated (Figure 2.10c). These experimental data suggested that 2 µM AcsD per reaction is reasonable and was chosen to obtain consistent data.

The influence of coupled enzymes on reaction rates by varying their protein concentration was studied and compared with the AMP reaction. No significant effect was observed by doubled or halved MK or PK/LDH concentrations (Figure 2.10d). Reactions containing 4 units of MK, 8.4 units of PK and 12.6 units of LDH showed the fastest conversion and were used in all further samples. Reactions lacking MK showed no change in NADH levels and indicate that AcsD is producing AMP and PP_i and not ADP, which could have been used by subsequent PK and hence would lead to fluorescence decrease. All these control reactions support and confirm that AcsD is the rate limiting enzyme. They also verify the stability and usability of AMP activity assay as a tool to obtain kinetic parameters.

2.3.2.3 Testing of proposed nucleophiles

As mentioned in section 1.5, AcsD is proposed to perform a condensation reaction between citric acid and L-2,4-diaminobutyric acid (DABA) or ethanolamine (EA) (Challis 2007). The activity was tested with a variety of nucleophiles, including L-serine, D-serine, ethanolamine and DABA which could in theory act as the substrate of AcsD. Reactions containing L-serine or hydroxylamine were more active than these with ethanolamine or DABA (Figure 2.11). D-serine turns over with a similar rate as ethanolamine, but twice as fast as DABA. All other tested nucleophiles have activities similar to reactions where no nucleophile (other than water) was present (Figure 2.11a). Control reactions where AcsD was omitted showed no detectable activity (Figure 2.11b).

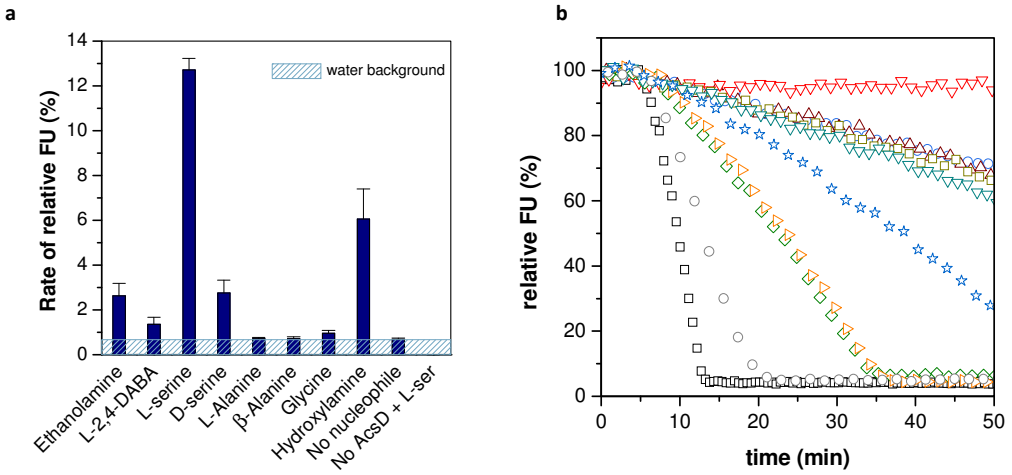


Figure 2.11: Activity assay with likely natural substrates. **(a)** L-serine is more reactive than D-serine, ethanolamine or hydroxylamine. Hydroxylamine, the second smallest nucleophile (water would be the smallest), accelerates the reaction much faster as reactions lacking any nucleophiles (except water). L-alanine, β-alanine or glycine have activity similar to reactions where no nucleophile (other than water) was present. Reactions lacking His_6-AcsD (wt) but containing L-serine showed no detectable activity. **(b)** Raw data for (a). L-serine (black ☐), D-serine (orange ▷), ethanolamine (olive ◇), L-2,4-DABA (blue ☆), hydroxylamine (grey ○), L-alanine (purple △), β-alanine (dark yellow ☐), glycine (dark cyan ▽), no nucleophile (except water) (blue ○), no His_6-AcsD (wt) but L-serine (red ▽).

L-serine, the fastest nucleophile, reacts 5 to 10 fold quicker than ethanolamine or DABA, respectively. Since the carboxylate moiety of serine is not incorporated in the final siderophore it must be removed in subsequent synthesis steps.

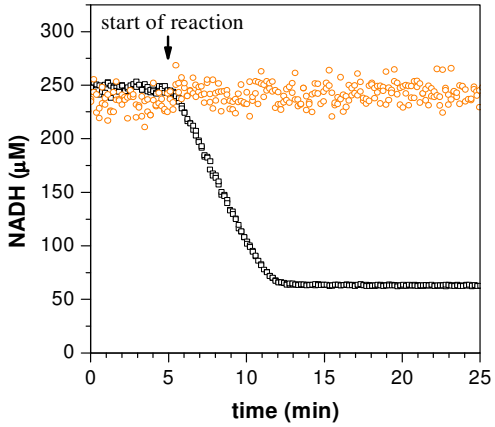


Figure 2.12: AMP production by AcsD. NADH consumption using His_6-AcsD with and without MK demonstrates AMP (and by implication PP_i) formation by AcsD. Omitting myokinase (orange circles) results in no detectable activity. Myokinase converts AMP and ATP to ADP. ATP is present in the reaction mixture and AMP must therefore be produced by AcsD.

2.3.2.4 AcsD reaction generates AMP and PP_i

ATP has recently been shown to be converted to AMP by other NIS synthetases (Kadi et al. 2007; Oves-Costales et al. 2007). The decomposition of citryl-adenylate by AcsD should therefore liberate AMP, a citrate derivative and pyrophosphate. No NADH consumption (with L-serine as nucleophile) was detected when myokinase was omitted (Figure 2.12). Thus AcsD cannot form ADP (and generate P_i) on its own, as ADP is the substrate of pyruvate kinase (the next enzyme in the coupled assay, see also section 2.2.5). Formation of ADP by AcsD would have caused NADH oxidation and led to a decrease in fluorescence signal. Myokinase converts AMP and ATP into two molecules ADP, which are then used by pyruvate kinase to recycle ATP by PEP

dephosphorylation. ATP is present in the reaction mixture and AMP must therefore be generated by AcsD (Figure 2.6).

2.3.2.5 Kinetic data

Kinetic parameters for AcsD were obtained by varying the concentration of buffered citric acid (pH 8.0) from 0 to 90 mM. To ensure Michaelis-Menten enzyme kinetics the second substrate (L-serine) was added in excess (75 mM) to the reaction, while all other parameters were kept constant as described in section 2.2.6.

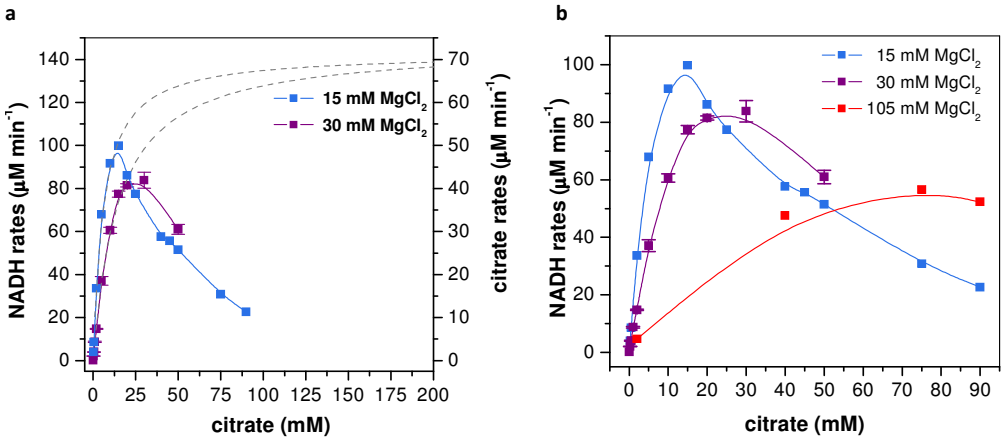


Figure 2.13: Enzyme kinetics for citrate. **(a)** Data obtained with 15 and 30 mM $MgCl_2$ were fitted with Michaelis-Menten equation (dotted lines) and corrected for background. **(b)** Same as in (a), but including data with 105 mM $MgCl_2$. At high citrate concentrations reaction rate increases when artificially high concentration of Mg^{2+} are present. Error bars were generated by three independent measurements.

Data were converted (see section 2.3.2.1) and rates (slopes) determined by fitting each data set with a linear function using Origin (Origin Labs). Rates (NADH consumption in µM min^{-1}) were plotted over citric acid concentrations (Figure 2.13a). With a Mg^{2+} concentration of 30 mM the apparent kinetic parameters for citrate are K_m = 14.7 ± 0.002 mM, k_{cat} = 2.2 x 10^3 s^{-1} and k_{cat}/K_M = 1.5 x 10^5 $M^{-1}s^{-1}$ (Table 2.6). Repeating the assay with 15 mM $MgCl_2$ shows, as citrate concentration increases

beyond 15 mM, that reaction rate decreases rapidly. This can be attributed to Mg^{2+} citrate complexation, which is supported by the fact that additional Mg^{2+} increases the reaction rate (Figure 2.13b). However, high concentrations of Mg^{2+} also significantly diminish AcsD performance at low citric acid concentrations.

Table 2.6: Derived apparent kinetic parameters for citric acid. The values were corrected by subtracting the background rate where AcsD was omitted.

$MgCl_2$	**K_M (mM)**	**V_{max} (µM/min)**	**k_{cat} (s^{-1})**	**k_{cat}/K_M (M^{-1} s^{-1})**
15 mM	5.96 ± 0.86	71.44 ± 4.06	2.14×10^3	3.6×10^5
30 mM	14.67 ± 0.02	73.24 ± 0.046	2.20×10^3	1.5×10^5

2.3.3 Screening and optimization of co-crystals

Due to problems in obtaining good diffracting co-crystals with the published sodium tartrate condition (McMahon et al. 2008), a new screening series was carried out and revealed a new crystallization condition in the JMAC Screen, well 92 (0.1 M TRIS-HCl pH 8.5, 20 % (w/v) PEG 8000). Though crystals were thin clustered plates (Figure 2.14a), optimization for each co-complex by varying salt, precipitant, buffer and substrate concentration improved clusters to rather large and well diffracting crystals (Figure 2.15). For more information on each condition see section 2.2.2.

a

b

```
 - Experiment detail -
Experiment: AcsD No HIS - JMAC
Barcode: ACSDNHIS04
Temperature: 20C
Screen: JMAc Crystal Screen DWB
Well: D12, Site 2
Fill Vol: 100uL
 - Screen well detail -
  Final     Well%     Name
 ------     -----     ----
 100.00%   100.00    JMac Crystal Screen Tubes
 100.00%   100.00    JMAC Tube 92 (100.0% pH 8.5)

 - Drop site detail -
Total Drop Vol: 0.20uL
Macromolecule: AcsD - no HIS tag(9/3/07 - 2xPCT - 6mg/ml)
Dispense Order: Precip-Prot-Adtv (0.1uL, 0.1uL, 0.0uL)
```

Figure 2.14: Initial crystal condition of AcsD co-complex. **(a)** Initial crystal clusters obtained after a couple of days from JMac Crystal Screen tube 92. **(b)** Experimental detail used for the crystallization robot (Cartesian HoneyBee) condition: 0.1 M TRIS-HCl pH 8.5, 20 % (w/v) PEG 8000

While structures of apo, citrate co-complex (overlapping with ATP) and ATP co-complex show $P2_12_12_1$ symmetry (Figure 2.15a-c), the AMP and SO_4^{2-}, adenosine-citrate-SO_4^{2} and ethylenediamine co-complex were solved in P1 space group. Other than the $P2_12_12_1$ crystals, P1 crystals have a rhombic morphology with sharp edges (Figure 2.15d-f). The data set for the ethylenediamine co-complex was collected from a very large crystal, joined with a smaller fragment as seen in Figure 2.15f. Careful centering on the top edge using a beam size of 100 x 100 µm (ID29, ESRF) helped to avoid an overlapping diffraction pattern during data collection.

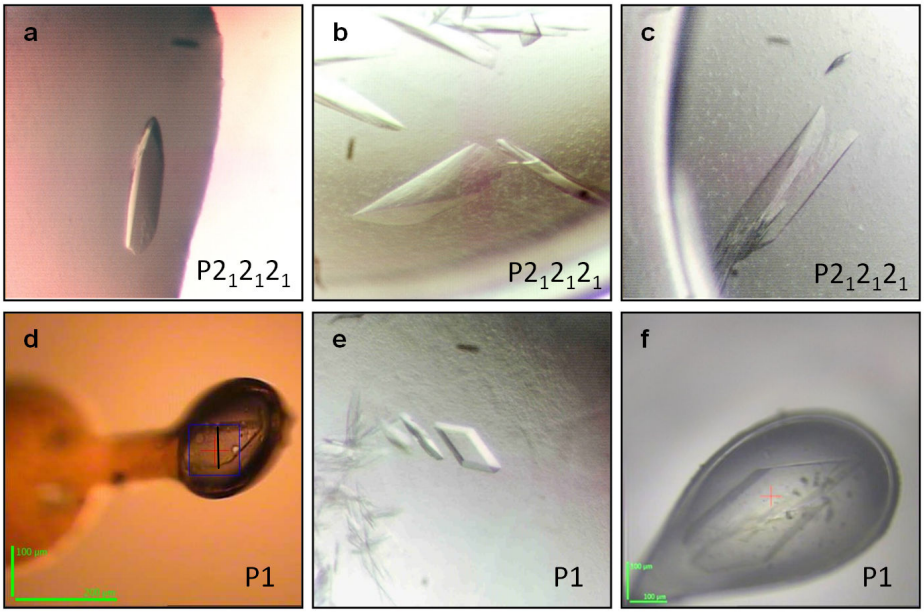


Figure 2.15: Crystal gallery of AcsD crystals. **(a)** AcsD apo crystals (modified from McMahon et al. 2008). **(b)** Co-complex with citrate (overlaps with ATP). **(c)** Co-complex with ATP. **(d)** Co-complex with AMP and SO_4^{2-}. **(e)** Co-complex with adenosine, citric acid and SO_4^{2-}. **(f)** Co-complex with ethylenediamine and ATP.

2.3.4 NIS conserved residues in active site

A ClustralW alignment of selected type A, B and C NRPS-independent siderophores shows that in comparison with the active site region of AcsD several conserved residues are found (Figure 2.16a). Residues S279, R281, K293, R305, E341, R369, H444, N447 and D464 are strongly conserved, while H170, T282, T301, E341, H444, Q446, N447, E466 and N509 are partly conserved in the superfamily of NIS siderophores. Interestingly, all but not E341 are involved in coordination of ATP or Mg^{2+} in AcsD (Figure 2.16b, c). Most of these residues (S279, R281, T282, K293, T301, R305 and R369) are located in the middle domain (red), while the remaining residues are located in the C-terminal domain (green) (Figure 2.16a).

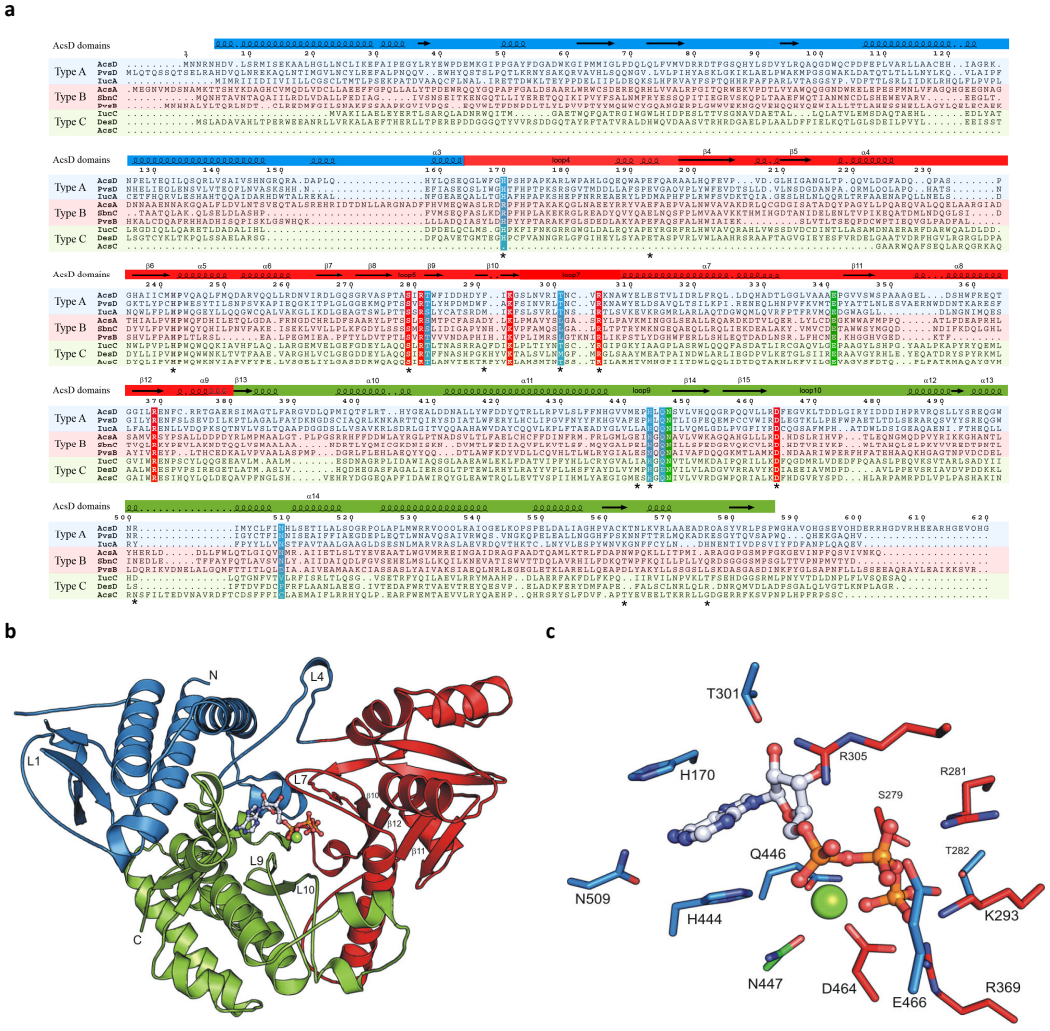


Figure 2.16: Conserved active site residues. **(a)** Alignment of type A, B and C NIS synthetases. Highlighted in red are residues absolutely conserved in the active site of AcsD, residues highlighted in blue are partly conserved within NIS. Residues highlighted in red and blue coordinate ATP in AcsD. (★: AcsD residues mutated) Protein sequences and Genbank references used in constructing the alignment: **Type A** AcsD - Pectobacterium chrysanthemi gi|16118732; PvsD - vibrioferrin biosynthesis protein *Vibrio parahaemolyticus* gi|23307118; IucA - IucA protein *E. coli* CFT073 gi|26109870. **Type B** AcsA - *P. chrysanthemi* gi|16118729; SbnC - *Staphylococcus aureus* gi:32441989; PvsB - vibrioferrin biosynthesis protein PvsB *V. parahaemolyticus* gi:23307116. **Type C** IucC - *E. fergusonii* gi|22034306; DesD - hypothetical protein SCO2785 *Streptomyces coelicolor* A3(2) gi|21221236; AcsC - *P. chrysanthem*i gi|16118727). AcsD domains are colored in blue (N-terminal domain), red (middle domain) and yellow (C-terminal domain). **(b)** AcsD monomer structure with ATP bound (balls and stick model) in cartoon representation color coded and annotated as above. **(c)** Absolutely conserved (highlighted in red) and partly conserved (highlighted in blue and green) active site residues in NIS enzymes coordinating ATP (colored in white carbons) and Mg^{2+} (colored in green).

2.3.5 Co-complex structures

The native protein and all co-complex structures are essentially identical with a root mean square deviation (rmsd) for 573 C_α atoms of 0.4 Å. No large conformational changes were observed upon ligand binding. The active site is created by the arrangement of the three domains which form a deep cavity. In this cavity ATP is coordinated between the middle, C-terminal domain and loops 4 and 7. It is cradled in a position perpendicular to β-strands 10, 11 and 12 (Figure 2.17a). AMP and sulfate are located in the same position as ATP (Figure 2.17b).

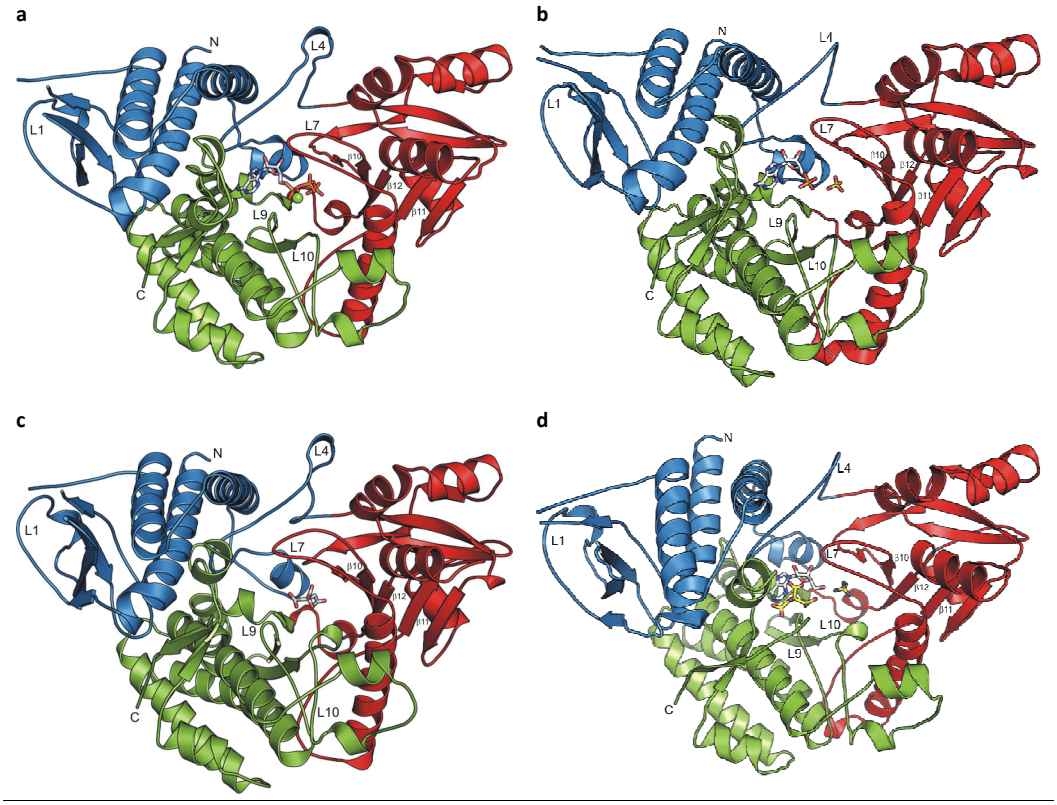


Figure 2.17: Complex structures of AcsD from *Pectobacterium chrysanthemi*. **(a)** Monomer of AcsD in cupped hand representation cradling ATP. N-terminal domain is colored in blue; middle domain in red and C-terminal domain in green. Carbons of ATP are shown in white. Key secondary structure elements are labeled from the N-terminus. **(b)** AMP sulfate co-complex, coloring as above. **(c)** Citric acid artifact (bound at high citrate concentrations) mimicking the triphosphates of ATP. Coloring as above. **(d)** Coordination of citrate in its "real" binding site. Color and labeling as above; carbons of citric acid are yellow and for adenosine in white; oxygen atoms of sulfate are colored in dark grey.

Initial attempts to co-crystallize AcsD with citrate and ATP-γ-S led to a complex in which citrate is bound in the ATP site mimicking the triphosphate (Figure 2.17c). No density was observed for the ATP analogue. As this co-complex was obtained in artificially high citric acid concentrations (300 mM) it was considered to be an artifact. AcsD was therefore co-crystallized with adenosine, sulfate and citrate. This was intended to lock the protein into the correct conformation by imitating bound ATP but still leave space (at the α phosphorous) for citrate to bind in its true location (Figure 2.17d). In this co-complex, a sulfate ion binds at the γ phosphate position and adenosine binds in the same pocket as seen in the ATP structure. A citrate molecule was found bound in subunit B adjacent to the ATP binding site.

2.3.5.1 ATP coordination

ATP is coordinated almost exclusively through interactions with conserved residues (Figure 2.16, Figure 2.18a). Residues S279, R281, T282, K293, R305, R369, H444 and D464 all interact with the phosphates of ATP or with citrate (R305, H444) in the respective structures. Only K293, R369 and E466 are located on secondary structure motifs (β10, β12 and L10 respectively). T282 (totally conserved in C, partly in A) and H444 (conserved in Type A and C but not B) are the only residues not highly conserved in all groups of NIS. A further five conserved residues, H170, T301, H444, Q446 and N509, one Mg^{2+} ion and six bridging water molecules make additional contacts to the adenosine moiety of the ATP structure. In both subunits, the Mg^{2+} ion is bound to the α and γ phosphates of ATP and the side chains of Q446, N447 and D464 in an approximate octahedral environment (Figure 2.18a). There is no Mg^{2+} in the apo or citrate bound structures. In this position the adenine purine ring is

stabilized by a π-π interaction network between H170 and H444 whilst making additional interactions with N509. Interestingly N509 is only conserved among type A enzymes. The ribose ring is twisted relative to the adenine purine ring and is stabilized in this conformation through interactions with T301 and Q446.

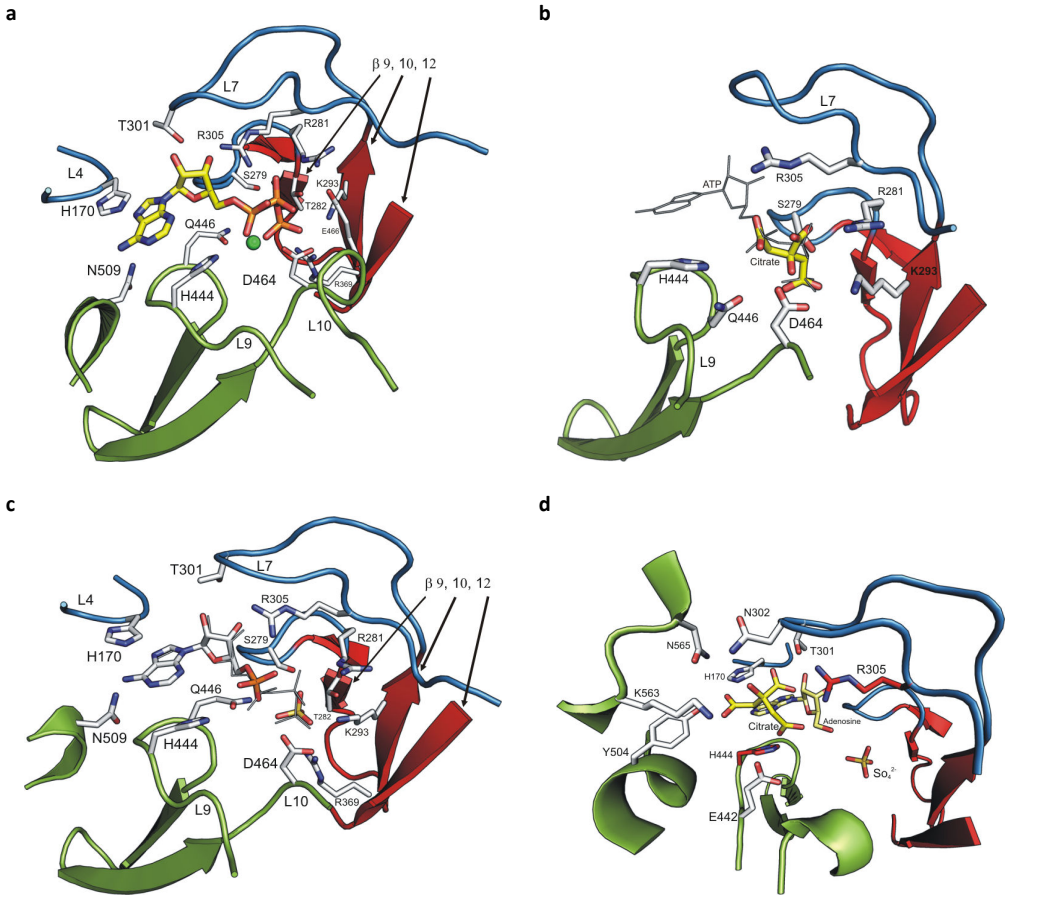


Figure 2.18: AcsD substrate coordination. Important secondary structure elements such as loops (e.g. L4 or L7) or sheets are labeled from the N-terminus. **(a)** ATP binding site with ATP **(b)** citrate artifact, binding position overlaps with ATP phosphates (superposed in wireframe). **(c)** AMP - sulfate co-complex overlaps with ATP (superposed in black wireframe). Carbons of AMP colored in white. **(d)** Citrate adenosine sulfate complex, true binding position.

Q446 adopts a different rotamer in the ATP coordinated structure relative to the apo and citrate (artifact) bound structures to accommodate ATP and Mg^{2+} ion (Figure 2.18a, b). From currently 250 deposited structures that contain ATP and a coordinating Mg^{2+}, only 7 structures show an α- γ phosphate coordination. By far the

most common is the coordination of the β and γ phosphates. Interestingly, the ATP conformer in AcsD, which is shaped akin to a horse shoe, has not been observed in the protein data bank. Binding of AMP and sulfate in the AMP-sulfate co-complex shows a similar coordination as ATP (Figure 2.18c). Only T301 adopts a different rotamer.

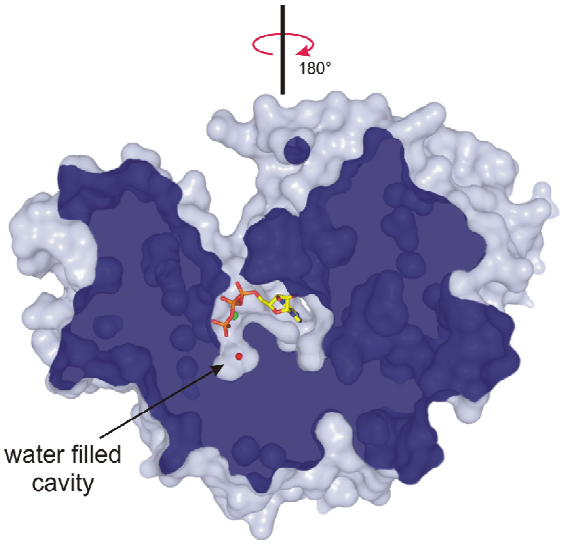


Figure 2.19: Surface representation of the water filled cavity in monomer B adjacent to the ATP binding pocket. AcsD is rotated by 180° compared to the view in Figure 2.17a. The atoms of ATP are colored as above, the water molecule in monomer B is colored in red.

The binding site for ATP is larger than ATP. The bottom end of the ligand binding site extends into a water filled cavity (Figure 2.19). Monomer B binds a single water molecule, but in monomer A three water molecules are located in this pocket. Hence, this cavity is likely to accommodate generated pyrophosphate (PP_i) during catalysis by replacing these water molecules.

2.3.5.2 Citric acid coordination

First attempts to co-crystallize citrate with ATP-γ-S (as a non-cleavable ATP analogue) showed that citric acid was coordinated in a similar position as the triphosphates of ATP (Figure 2.18b). No electron density for ATP-γ-S was observed. The high concentrations of citric acid (300 mM) used in the crystallization condition probably led to this artifact structure. The "correct" citrate structure was obtained by co-crystallization with adenosine and sulfate imitating bound ATP. Citrate was only found bound in monomer B. The electron density in monomer A was not distinct enough to model another molecule of citric acid. Citrate binds to residues located mainly on loops (Figure 2.18d). The electron density shows that the central carboxylate of citrate is recognized by T301 and R305. The hydroxyl group makes a hydrogen bond to Q302. These interactions serve to unambiguously position the citrate with the *pro-R* carboxylate pointing towards the ATP site making hydrogen bonds to H444 and R305. The *pro-S* carboxylate makes hydrogen bonds with H170, K563 and Y504. Real space refinement of citrate in all possible orientation was used to select and determine its correct and final position (Figure 2.18d).

2.3.6 AcsD shares part of the cAPK fold of kinases

A structure similarity search in DALI (Holm and Sander 1993) and SSM (Krissinel and Henrick 2004) reveals that AcsD has an overall novel topology, with no match to any known protein. Superimposing ATP molecules of all 250 in the PDB deposited ATP co-complex structures to the AcsD bound ATP revealed that only 7 structures show an α and γ ATP phosphate coordination with Mg^{2+}. These seven structures are mostly kinases (1csn, 1phk, 1qmz, 1ql6) (Owen et al. 1995; Xu et al. 1995; Brown et al. 1999; Skamnaki et al. 1999) including two cAMP-kinases (1atp, 1rdq) (Zheng et al. 1993; Yang et al. 2004) and one ribonucleotide transformylase which is involved in the purine biosynthesis (1kj9) (Thoden et al. 2002).

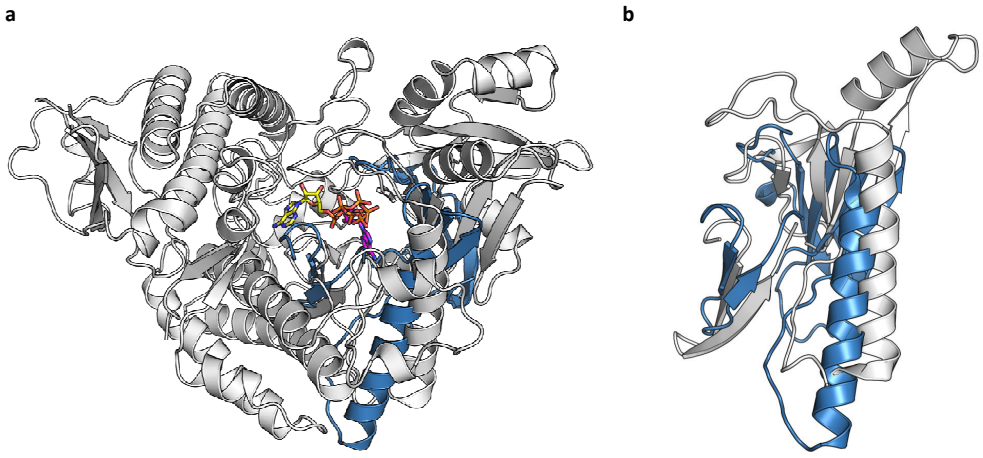


Figure 2.20: Similarity between AcsD and the cAPK fold. **(a)** Partly secondary structure match of AcsD colored in grey and 1e8x (a cAPK kinase) colored in blue. The position of ATP is shown in yellow (AcsD) or magenta (1e8x). **(b)** Close up of the matching region of secondary structure.

Furthermore, a domain similarity search reveals, that the central two domains (palm and fingers) do have partial structural match to several cyclic AMP-dependent protein kinases (cAPK) which bind nucleotide triphosphates (NTPs) including 1e8x (Walker et al. 2000), 2gqr (Ginder et al. 2006) and 1cja (Steinbacher et al. 1999). The

structures superimpose with rmsd between 3.30 and 3.55 Å for around 130 matching C_α atoms (Figure 2.20).

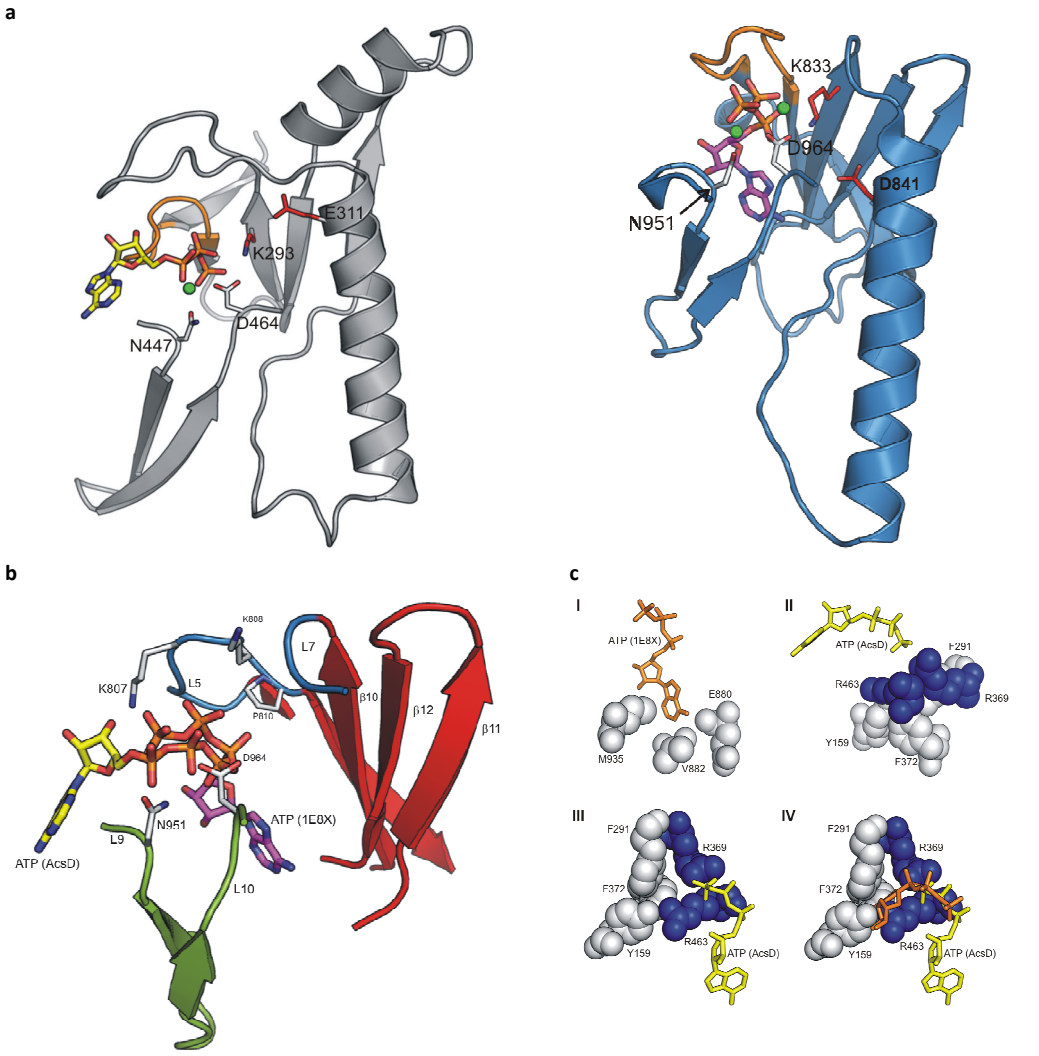


Figure 2.21: Essential features of the cAPK fold required for catalysis. **(a)** The similar structural region of AcsD (grey) compared with 1e8x (blue) showing ATP bound in different orientation with key residues **(b)** ATP binding site of 1e8x as an example of the cAPK fold, the position of residues involved in tethering ATP are conserved in AcsD. ATP binding in 1e8x is shown in magenta. Superimposed in yellow is the position of ATP in AcsD. Secondary structure elements (loops (L5, L7, L9 and L10) and sheets) are labeled from the N-terminus. **(c)** Space fill diagram illustrating the different environment of the conserved ATP binding pocket, orange is ATP from 1e8x and yellow is ATP from AcsD. ***i*** 1e8x binding pocket ***ii*** AcsD binding pocket ***iii*** different orientation than ii; ***iv*** AcsD pocket with 1e8x ATP superimposed shows that it clashes with Y159 and F372.

NTP in the cAPK fold is bound in a conservative manner and the adenosine moiety extends deep into a neutrally charged pocket sitting almost parallel to β-strands 10,

11 and 12 (Figure 2.21). Residues on analogous loops to those in AcsD (L5, L9 and L10) bind NTP through interactions with the exposed phosphate(s) (Figure 2.21b). Other than the adenosine rings, the phosphates occupy the same space as those in AcsD, but the vectors from α to γ phosphates have opposite directions (Figure 2.21a,b). Additional residues in the neutral pocket make interactions with the adenosine moiety. Although these structures share a common fold and phosphate coordination with AcsD, coordination of the adenosine moiety in AcsD is strikingly different (Figure 2.21b).

The cAPK adenosine moiety binding pocket in AcsD has a basic charge and is sterically occluded compared to the neutral pocket in the cAPK fold. In 1E8X residues E880, V882 and M953 are found in a neutral pocket. In the AcsD pocket they are replaced by Y159, F291 and F372 and also occupied by conserved residues R369 and semi conserved R463 (alternative is K) (Figure 2.21c). As a result, the triphosphates of ATP in AcsD sit in a position normal to β-strands 10, 11 and 12 and bind to residues on two insertions, L4 and L7, found in AcsD relative to the classic cAPK fold. L4 and L7 extend across the top of the ATP binding site effectively sealing the site from above. In the cAPK fold L4 is not found and L7 is typically much shorter. As a result only one route (in and out) for large molecules exists and directly exposes the α phosphate of ATP in AcsD for incoming substrates (Figure 2.19).

2.3.7 Mutational studies

Active site and conserved residues in AcsD were mutated as described in section 2.2.4. In Table 2.2 (section 2.2.4.1) all created mutants including progress in purification and testing are summarized. In the sequence alignment (Figure 2.16a) all

mutated sites of AcsD are marked with a star (✱). Five of the 23 planned mutants failed in PCR, expression or later purification steps. This was mainly caused by inefficient primers or mutations which caused more inclusion bodies and hence led to insoluble protein. To ensure correct folding of each purified mutant, circular dichroism (CD) spectroscopy analysis was carried out (see appendix B, Figure 4.7). The previously described AcsD activity assay was used for activity detection with L-serine (Figure 2.22) and/or other substrates (Figure 2.34).

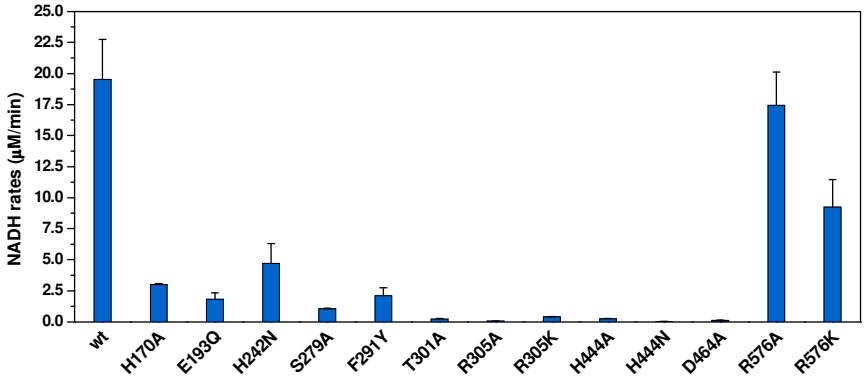


Figure 2.22: AMP activity assay for *wt* AcsD and active site mutants.

Any mutation in T301, R305, H444 and D464 led to total loss in activity (<1 %) and hence are likely to be crucial for AcsD activity. T301 and R305 recognize the central carboxylate of citrate and H444 binds to the *pro*-R carboxylate of citrate (Figure 2.18d). D464 binds to the β and γ phosphates of ATP and is involved in Mg^{2+} coordination. H444A and R305K mutants still have a residual activity of ~1 % and ~1.5 %, respectively. All other tested mutants show activity above 5 % (compared to *wt* activity). Remarkably, mutation of R576 residue had only little (R576A, ~90 %) or a minor (R576K, ~47 %) effect on enzyme activity. This residue is not conserved and located on a flexible loop on top of the active site. It was thought to be involved in

coordination of citric acid (Schmelz et al. 2009). If so, R576 is not crucial or important for catalysis.

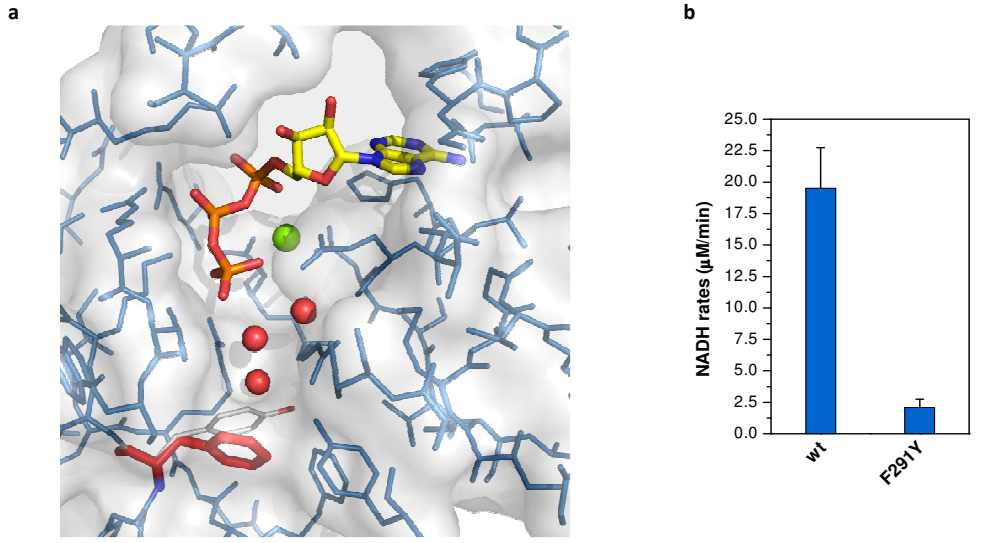


Figure 2.23: Mutation in ATP binding cavity. **(a)** At the bottom of ATP binding cavity three water molecules (red) are located. A model of F291 mutation (colored in red) to tyrosine (carbons in white) shows, that it may lead to displacement of water and therefore decreases space for generated pyrophosphate. **(b)** Though F291 is far away from the α phosphate of ATP, a simple mutation to tyrosine leads to a 10 fold decrease in AcsD activity.

Mutants H170A, E193Q, H242N and S279A show that they are important, but not lethal for catalysis since they still have a remaining activity of ~5-25 % compared to the *wt* protein. E193 and H242 are not located in the active site. Glu193 stabilizes the secondary structure of L4 by forming salt bridges with other conserved residues K177 and R179. His242 is located on the β6-α5 turn and binds weakly to a water molecule (W229 in ATP structure) and the backbone of I280 and V244 on L5 and α5-helix, respectively. The polar H242N mutant is still able to maintain the binding to V244 backbone but not to W229 or I280 on L5, which contains the conserved and ATP coordinating residues S279 and R281 (Figure 2.16). Hence the H242N mutation may lead to a more flexible L5 resulting in a reduced enzyme rate. Mutation of conserved H170 to alanine is more critical and reduces the reaction ~7 fold. This residue facilitates a π-stacking interaction with adenosine ring of ATP and binds to

the *pro-S* carboxylate of citric acid and a mutation therefore is likely to weaken binding of ATP and citric acid. F291, located at the bottom of ATP binding cavity, does not interact with ATP (Figure 2.23a). However, mutation to tyrosine results in a 10 fold loss of activity (Figure 2.23b). The larger tyrosine (-OH) is likely to displace one of the nearby coordinated water molecules and hence reduces available space in the bottom of ATP binding cavity (Figure 2.23a).

2.3.8 Mass spectroscopy of AcsD reaction product

2.3.8.1 Timeline of AcsD reaction

Although AMP production in AcsD reaction was demonstrated in the AMP activity assay (see section 2.3.2.4), mass spectroscopic analysis was used to characterize and verify AMP and final product formation. Different time points of AcsD reaction with ATP, Mg^{2+}, citric acid and L-serine were analyzed with a LCT-ESI MS instrument in negative mode (section 2.2.8).

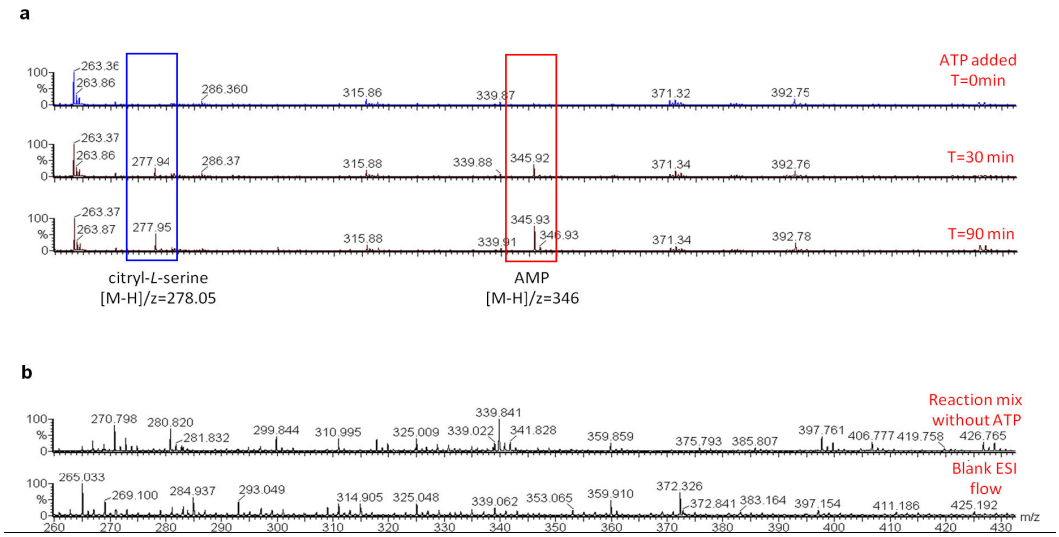


Figure 2.24: Mass spectroscopy of AcsD reaction with L-serine. **(a)** Time line of AcsD with L-serine measured with LCT-ESI MS. Two peaks increase over time corresponding to AMP ([M-H^+] *m/z* =346) and citryl-L-serine ([M-H^+] *m/z* =278). **(b)** Control reactions, where ATP was omitted or only solvent was tested, show not the additional peaks detected in (a).

Two new mass peaks, not detected in control reactions, were observed with *m/z* values of 277.9 and 345.9 after 30 min incubation. These peaks get more prominent as subsequent time points show (Figure 2.24a). These mass peaks were not observed in control solutions or reactions where ATP was omitted (Figure 2.24b). AMP has a theoretical *m/z* value of 346 for [M-H]$^-$ which is close to the detected mass of

m/z=345.9. The lower mass peak (*m/z*=277.9) can be attributed to citryl-L-serine which has a theoretical mass of *m/z*=278.05. These results support the obtained assay results (section 2.3.2.4) and verify that AcsD is able to utilize ATP, Mg^{2+}, citric acid and L-serine producing citryl-L-serine and AMP.

2.3.8.2 Tandem mass spectrometry of citryl-L-serine

Isolated and purified large scale enzymatic derived citryl-L-serine was identified as *N*-citryl-L-serine (carried out in the group of Prof. G. Challis). This was surprising as it was thought that L-serine would catalyze ester bond formation with its hydroxyl, rather than amide formation with its α-amino group and the *pro-R* carboxylate of citric acid. The fragmentation pattern of enzymatic derived citryl-L-serine, isolated *N*-citryl-L-serine and chemically synthesized *O*-citryl-L-serine (by group of G. Challis) were tested and compared using tandem mass spectrometry (section 2.2.8). Directly injected AcsD product (no isolation or purification) of *N*-/*O*-citryl-L-serine led to four major daughter ions with *m/z* = 260.05, 198.08, 173.04 and 111.03 (Figure 2.25a). Each of these ions result from loss of water alone (*m/z* = 260.05), two water plus carbon dioxide (*m/z* = 198.08), loss of serine ($C_3H_5NO_2$, *m/z* = 173.04) and loss of serine plus two waters plus carbon dioxide (*m/z*=111.03) (Figure 2.26, see also Table 2.7), respectively.

Isolated *N*-citryl-L-serine has only three major daughter ions with *m/z* = 260.05, 198.05 and 110.07 (Figure 2.25b). The *m/z* = 260.05 and 198.05 daughter ions are identical to those produced of directly injected citryl-L-serine. The *m/z* = 110.07 daughter ion could result by loss of serine ($C_3H_4O_3$) plus two waters plus carbon dioxide (Figure 2.26). However the *m/z* = 173.04 daughter ion for *N*-citryl-L-serine is

barely detectable. The amide bond is significantly stronger (compared to ester) and hence might not undergo cleavage as readily. Furthermore, the chemically synthesized *O*-citryl-L-serine ester shows the same fragmentation pattern (Figure 2.25c) as directly injected enzymatic product or citric acid (Figure 2.25a, d), identifying it as *O*-citryl-L-serine.

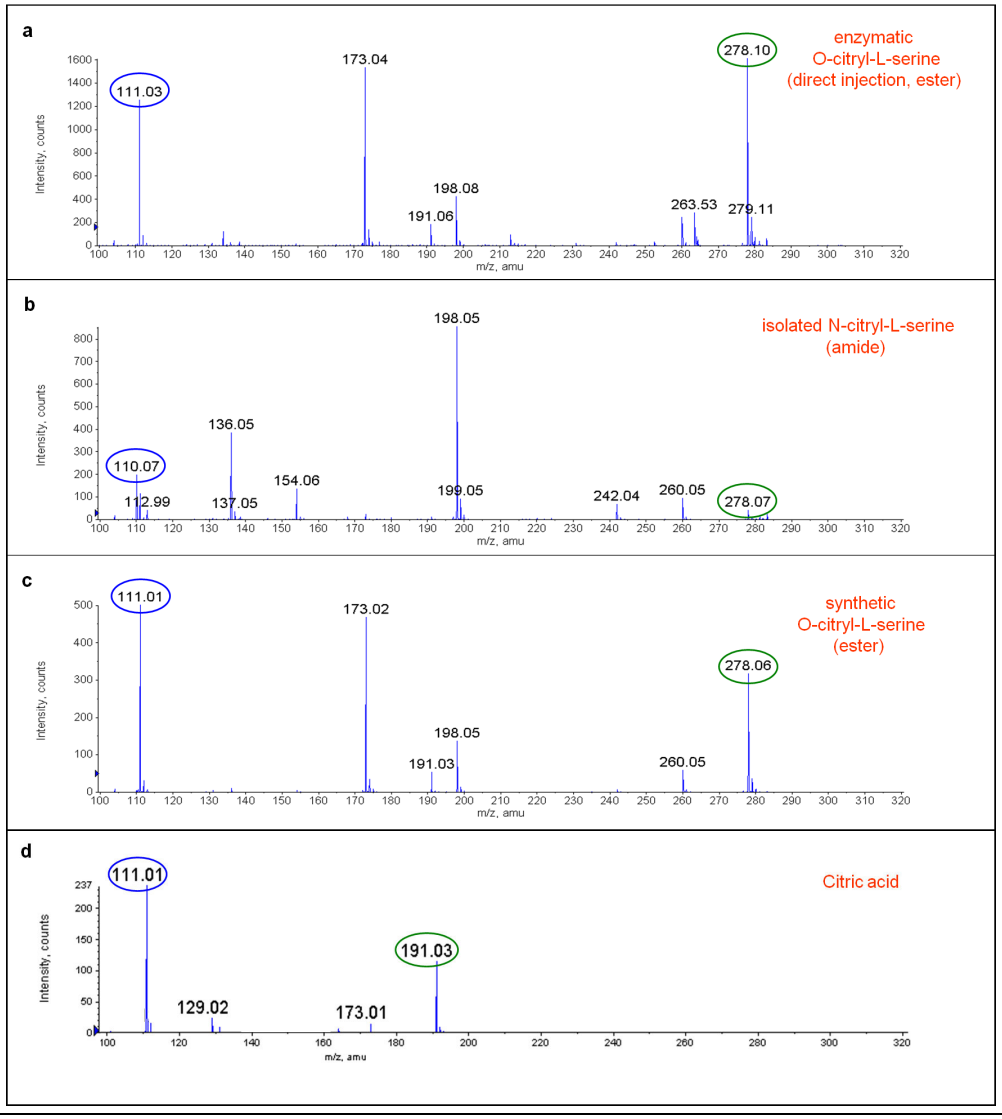


Figure 2.25: MS/MS analysis of AcsD and control reactions. **(a)** Tandem mass spectroscopy of enzymatic derived (but not purified) citryl-L-serine with a strong *m/z* = 111 peak typical for *O*-citryl-ester (see also Figure 2.26). **(b)** The fragmentation pattern for isolated and purified *N*-citryl-L-serine has the *m/z*=110 peak which can be attributed to a citryl-amide linkage (see also Figure 2.26). **(c)** Fragmentation pattern of chemical synthesized *O*-citryl-L-serine ester shows like (a) and (d) the characteristic *m/z*=111 peak. Other daughter ions are identical with (a), identifying the enzymatic product as *O*-citryl-L-serine and not as *N*-citryl-L-serine. **(d)** Daughter fragments of citric acid show a similar pattern compared to *O*-citryl-L-serine (a) and (c).

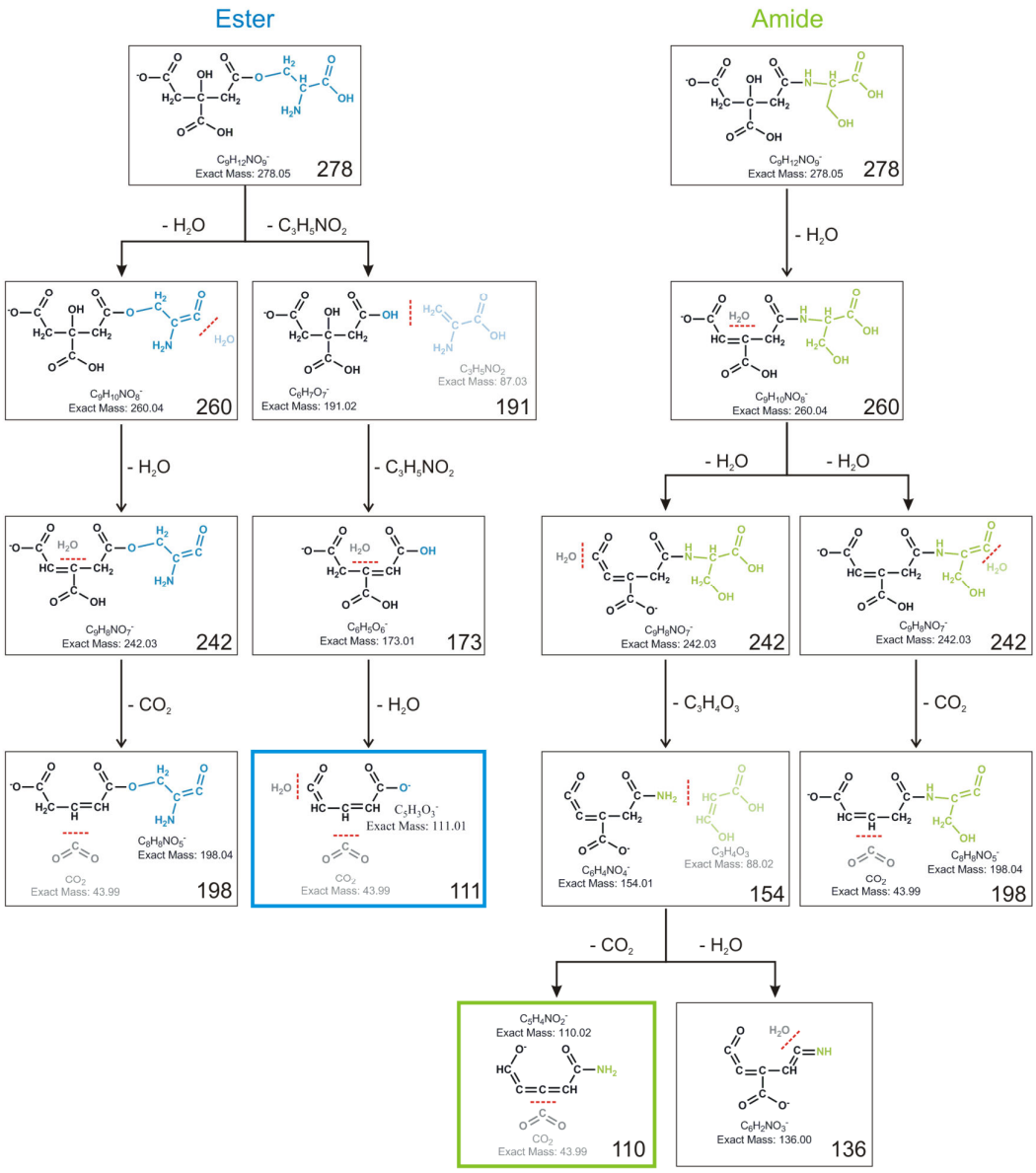


Figure 2.26: Possible *O*-citryl-L-serine (ester) and *N*-citryl-L-serine (amide) MS/MS fragments. The fragmentation scheme shows that a simple fragmentation generated the ester specific *m/z*=111 and the amide specific *m/z*= 110 fragments.

Interestingly, the MS/MS fragmentation produces a specific pattern for authentic ester or amide citrate-L-serine (Figure 2.26). Simple fragmentation of the AcsD ester product with L-serine leads to a daughter ion of *m/z* = 111, as seen for directly injected citrate-L-serine and the chemical synthesized *O*-citrate-L-serine. On the

other hand fragmentation of the amide yields an ion with *m/z* = 110, as seen for isolated *N*-citryl-L-serine (Figure 2.26, Table 2.7). This is strong evidence that AcsD catalyzes ester formation and liberates *O*-citryl-L-serine, consistent with lack of reactivity demonstrated with β-alanine (Figure 2.11). Hence *N*-citryl-L-serine from the preparative scale most likely results from re-arrangement of *O*-citryl-L-serine (the product of the enzymatic reaction) during the purification (for more experimental detail see Schmelz et al. 2009).

Table 2.7: MS/MS fragmentation pattern of *O*- and *N*-citryl-L-serine

Fragments	structure	enzymatic *O*-citryl-*L*-serine (direct injection, ester)	isolated *N*-citryl-*L*-serine (amid)	synthetic *O*-citryl-*L*-serine (ester)	citric acid
O- or *N*-citryl-*L*-serine	$C_9H_{12}NO_9$	278.10	278.07	278.06	
O- or *N*-citryl-*L*-serine -H_2O	$C_9H_{10}NO_8$		260.05	260.05	
O- or *N*-citryl-*L*-serine -$2H_2O$	$C_9H_8NO_7$		242.04		
O- or N-citryl-*L*-serine -$2H_2O$ -CO_2	$C_8H_8NO_5$	198.08	198.05	198.05	
citrate	$C_6H_7O_7$	191.06		191.03	191.03
citrate -H_2O	$C_6H_5O_6$	173.04		173.02	173.02
O- or *N*-citryl-*L*-serine -$2H_2O$ -$2CO_2$	$C_7H_8NO_3$	154.08	154.06		
O- or *N*-citryl-*L*-serine -$3H_2O$ -$2CO_2$	$C_7H_6NO_2$		136.05		
citrate -H_2O -CO_2	$C_5H_5O_4$				129.03
O-citryl-*L*-serine (ester) -$C_3H_5NO_2$ -$2H_2O$ -CO_2	**$C_5H_3O_3$**	**111.03**		**111.01**	**111.02**
***N*-citryl-*L*-serine (amide) -$C_3H_4O_3$ -$2H_2O$ -CO_2**	**$C_5H_4NO_2$**		**110.07**		

2.3.9 Catalysis of ester, amide and thioester bond formation

2.3.9.1 Systematic screening for possible nucleophiles

L-serine and sixteen other related small molecules (Figure 2.27) were tested using the fluorescence based activity assay as described in section 2.2.5. All tested nucleophiles were purchased from chemical suppliers, such as Sigma Aldrich or Fluka. Most nucleophiles are common and essential metabolic products or substrates, such as glycine, L-alanine, L-isoserine or L-lactic acid. In section 2.3.2.3 it

was already demonstrated that water is capable of hydrolyzing citrate adenylate at a detectable rate. This background rate was assigned to zero for nucleophilic activity.

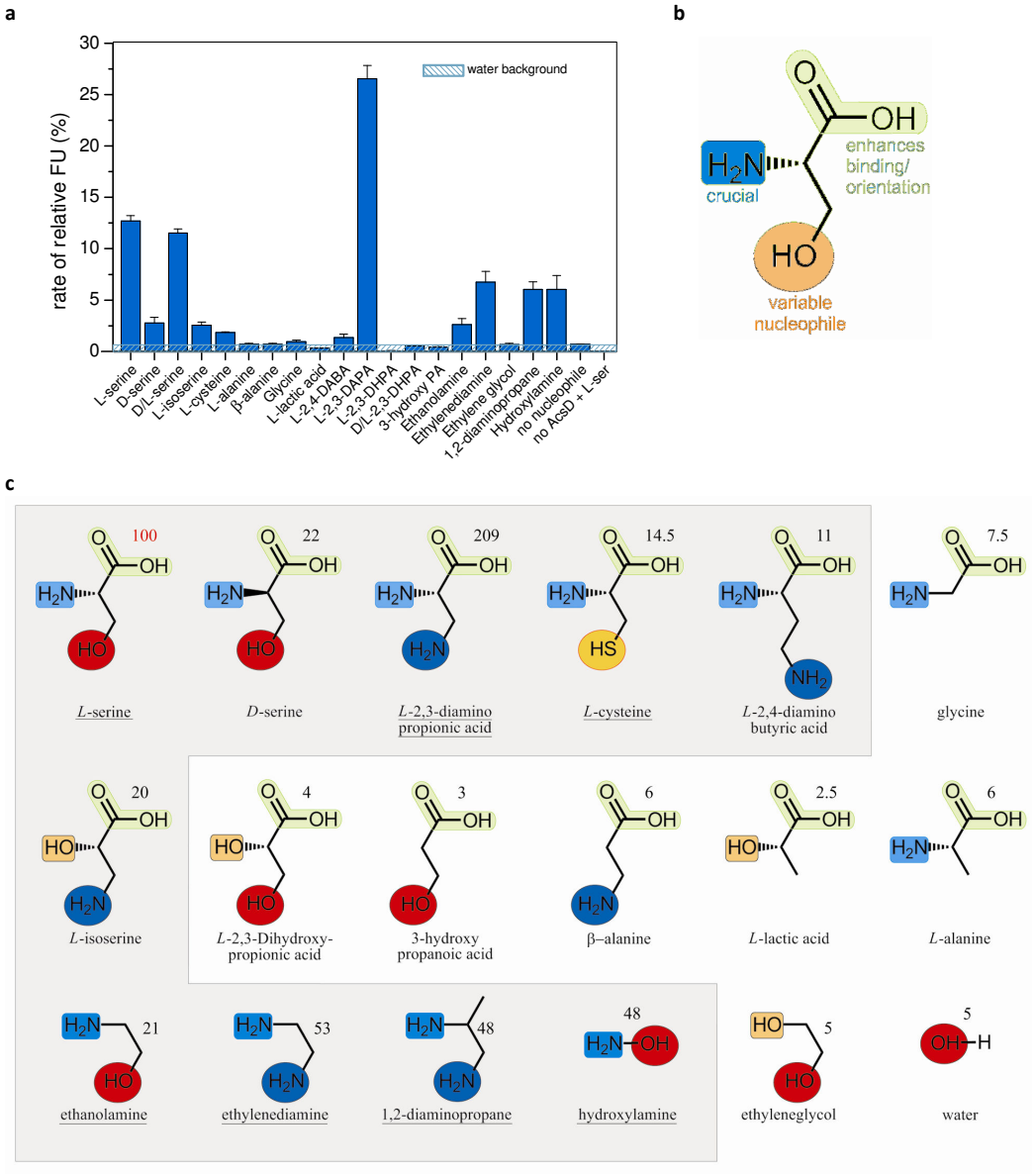


Figure 2.27: Nucleophilic specificity of AcsD: **(a)** AcsD rates with tested nucleophiles in triplicate. **(b)** Schematic representation of L-serine and role of each chemical moiety. **(c)** Chemical structures of tested nucleophiles with numbers indicating their reactivity in reference to L-serine (100). Nucleophiles which were tested twice as fast as water (5) are highlighted in grey and considered to form a citrate derivative. MS and MS/MS analysis were carried out for underlined nucleophiles. Functional groups are highlighted in similar colors for clarity.

Analytic mass spectrometry (MS and MS/MS) was used to determine the chemical composition of the citrate nucleophile linkage where this could be ambiguous (Figure 2.28, Figure 2.29). This approach was already used in section 2.3.8.2 to confirm that L-serine reacts through its hydroxyl, forming an ester linkage rather than an amide bond with its α-amino nucleophile. Ten substrates (including L-serine) showed significant higher activity than that of water (designated active nucleophiles) whilst seven nucleophiles have activity equal to water (within error).

The most striking observation of the active nucleophiles is that all but one have an α-amino group or are simple diatomics (non hydrogen atoms). Replacing this functional group with a hydroxyl or omitting it, leads to loss in activity as seen for L-2,3-dihydroxy-propionic acid, 3-hydroxy propanoic acid or β-alanine. This suggests that the α-amino group of L-serine plays a crucial role in recognizing and orienting the incoming nucleophilic hydroxyl.

The other immediate observation is that compounds which have a nitrogen nucleophile, rather than hydroxyl seen in L-serine, are more reactive. This is exemplified by the compound L-2,3-diaminiopropionic acid which differs from L-serine only in that it has an amine rather than a hydroxyl, however it is over twice as reactive in the assay. The enhanced reactivity is also apparent in ethanolamine compared to ethylenediamine (or 1,2-diaminopropane). In this context it should be pointed out that L-isoserine, which lacks the α-amino group, is still active. This activity can be attributed to the enhanced nucleophilic ability of the amine group which has replaced the hydroxyl and overcomes the substitution of the α-amino group with hydroxyl. The preference for an amine nucleophile is also reflected in the

MS/MS fragmentation pattern of the product between hydroxylamine (which has both nitrogen and oxygen nucleophiles) and citric acid (Figure 2.29). While most of the fragments can be derived from either N- or *O*-citryl-hydroxylamine, the *m/z*=126 fragment corresponding to *N*-citryl-hydroxylamine is far more dominant than the *m/z*=127 fragment belonging to *O*-citryl-hydroxylamine (Figure 2.30). Ethanolamine forms an ester linkage to citric acid, rather than an amide, since the distinct citryl-ester fragment with *m/z*=111 is present (see also section 2.3.8.2).

The results indicate that whilst the carboxylate function of L-serine is important for recognition it is not critical and a number of compounds lacking this group are still active (e.g. ethanolamine or ethylenediamine). The stereochemistry of the nucleophilic group is important, adding an additional carbon ($-CH_2-CH_2-NH_2$) as in L-2,4-diaminobutyric acid (DABA) dramatically decreases activity. However, decarboxylated DABA, 1,3-diamionopropane, was reported to be a suitable L-serine surrogate and found to be incorporated in a functional achromobactin derivative (Berti and Thomas 2009). Shortening the nucleophilic moiety to a methyl group ($-CH_3$), as seen for L-lactic acid or L-alanine, results in loss of function. Replacing the nucleophilic hydroxyl with sulfhydryl (L-cysteine) decreases the rate seven fold (compared to L-serine) but is still significantly above water. Mass spectrometry on the LCT ESI MS instrument indicates that the product is citryl-L-cysteine (Figure 2.31). It was not possible to obtain convincing tandem mass spectrometry of the thioester bond on the Q-Star XL ESI-QTOF instrument. However, no evidence was observed for a peptide bond either. The absence of amide and the known reactivity of thioester linkages imply that reactions with L-cysteine proceeds through the S nucleophile.

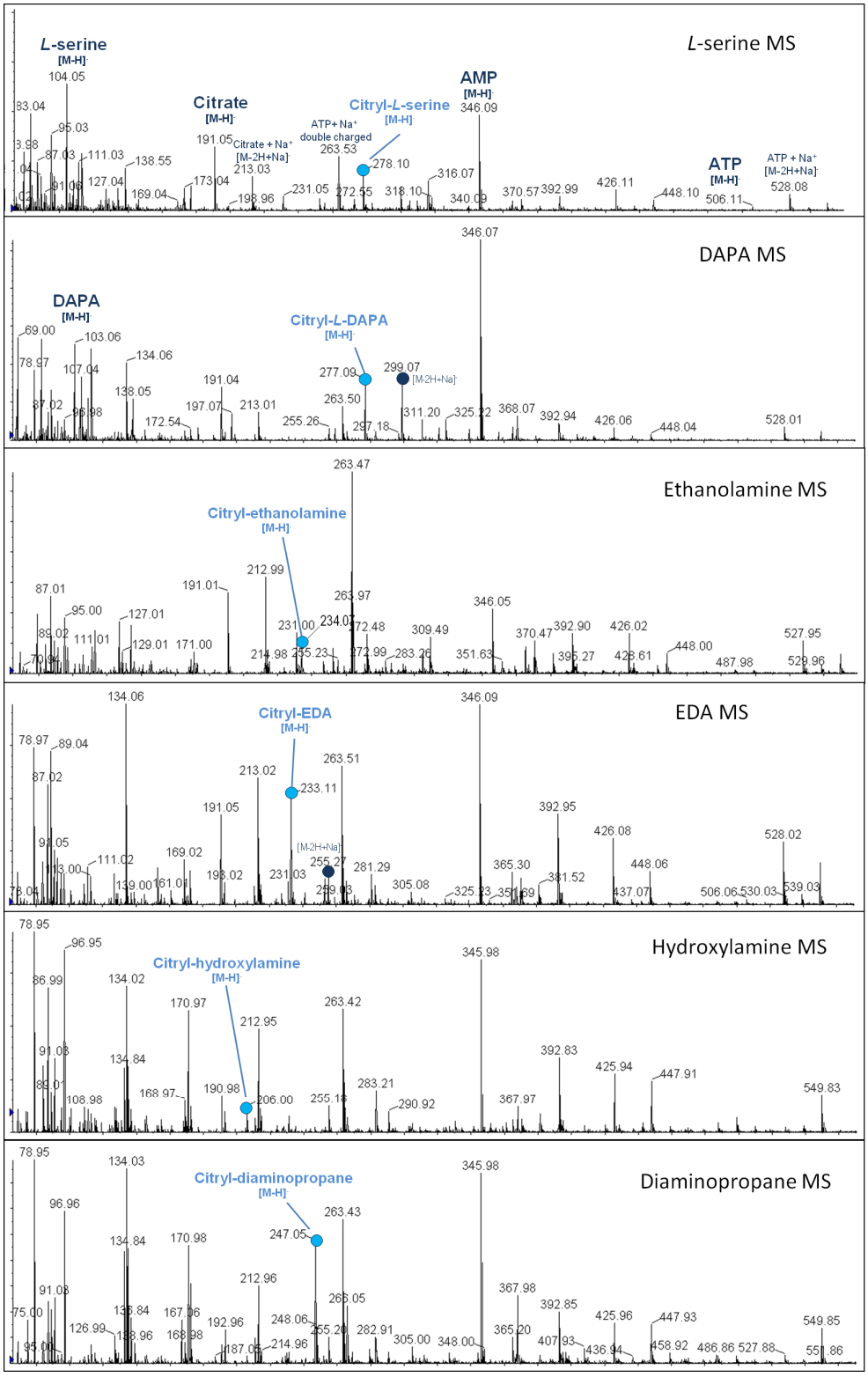


Figure 2.28: MS spectra of AcsD catalyzed citric acid ester and amides. In all tested reactions formation of AMP (*m/z* =346 $[M-H]^-$) was observed, while ATP (*m/z* =506 $[M-H]^-$ and *m/z* =528 $[M-2H+Na]^-$) was consumed. Product peaks were highlighted with light blue circles. For citryl-L-DAPA or citryl-EDA a second peak corresponding to *m/z*= $[M-2H+Na]^-$ (dark blue circle) could be assigned.

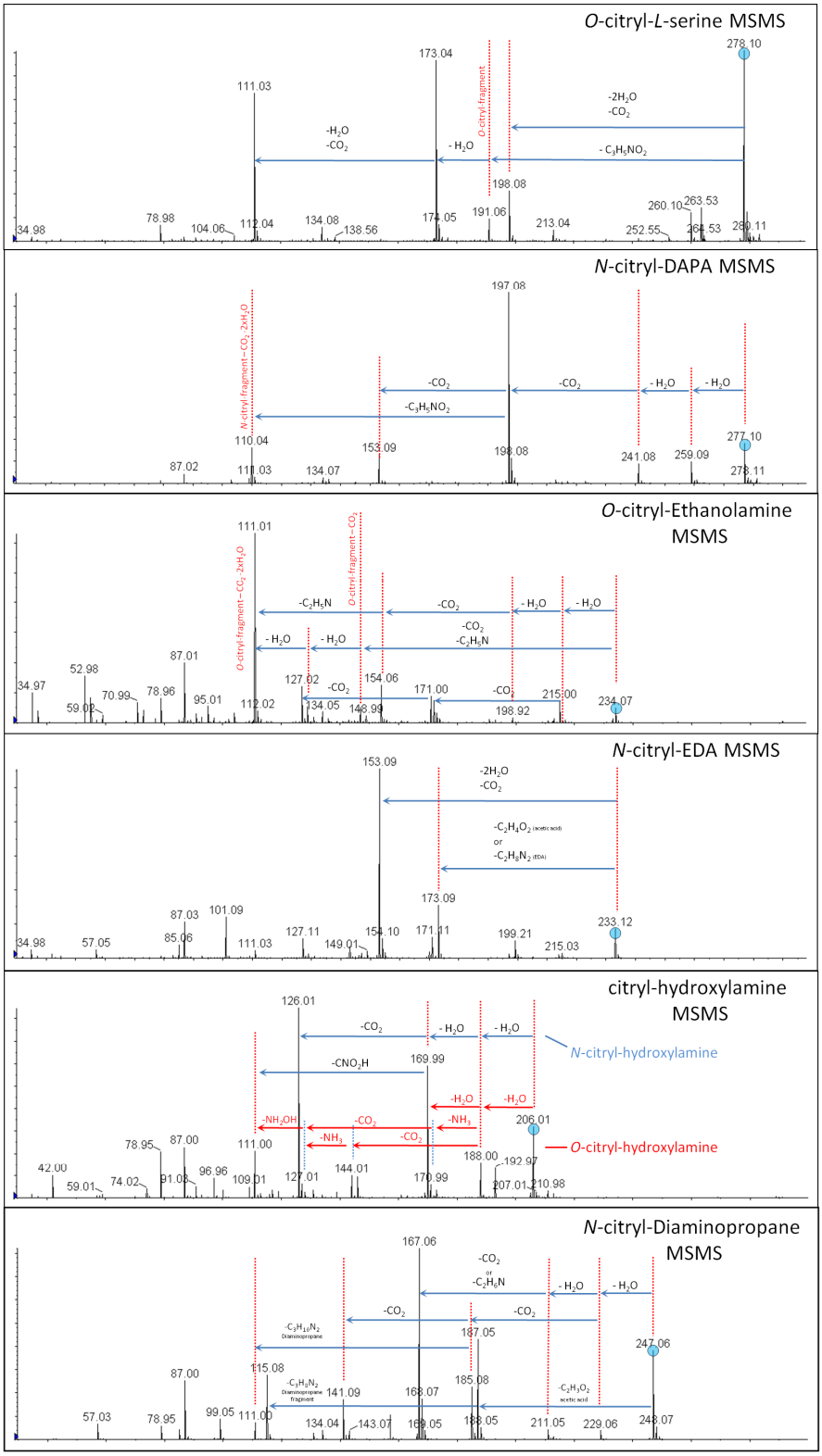
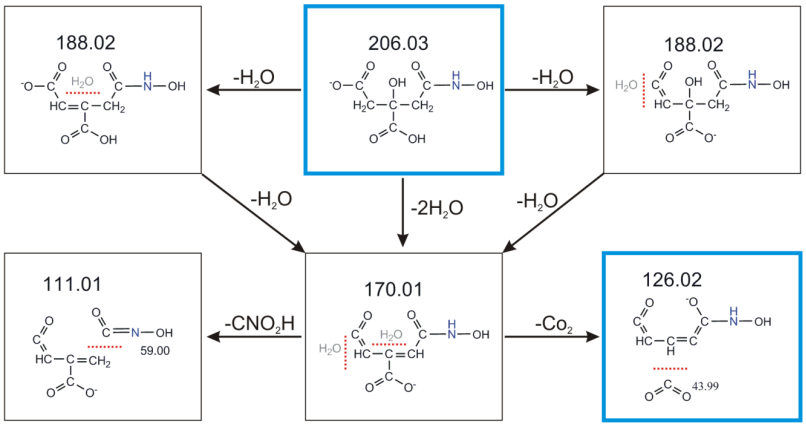


Figure 2.29: MS/MS spectra of AcsD citric acid ester and amide products. Respective product peaks are labeled with light blue circles. Fragmentation pattern including lost fragments and masses were assigned (as far as possible). For hydroxylamine the daughter fragments corresponding to *N*-citryl-hydroxylamine (blue) or *O*-citryl-hydroxylamine (red) were assigned, whereas peaks originated from *N*-citryl-hydroxylamine are dominant.

a



b

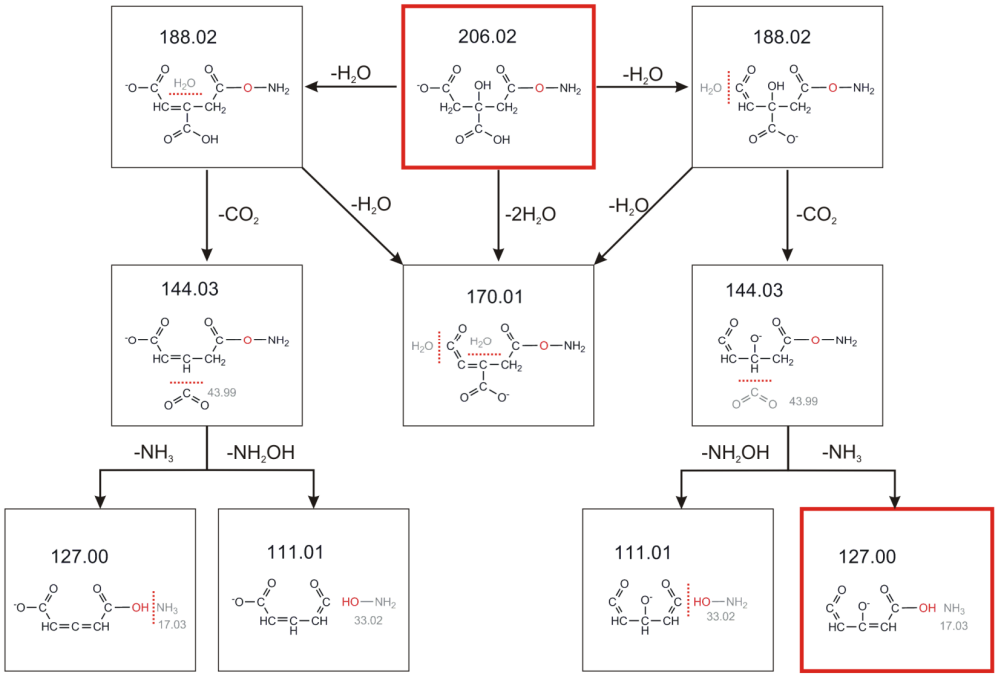
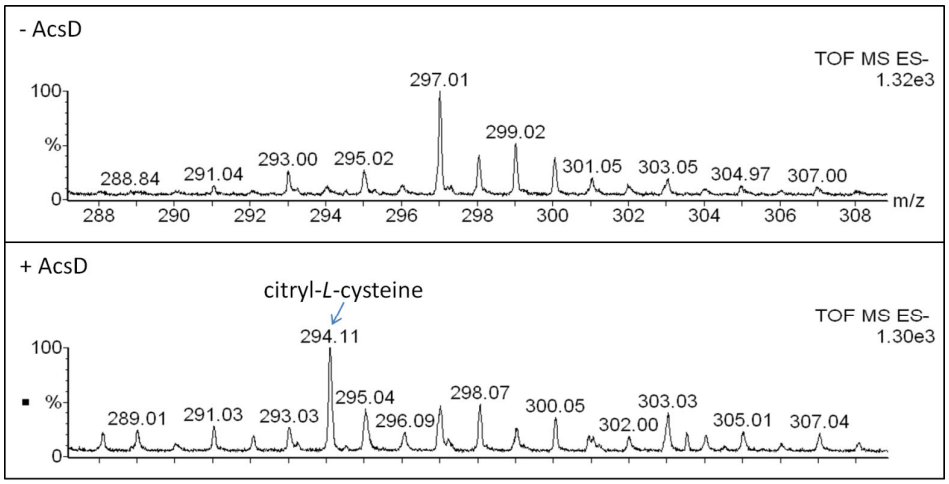


Figure 2.30: Theoretical fragmentation pattern of citryl-hydroxylamine **(a)** Possible fragments for *N*-citryl-hydroxylamine. **(b)** Possible fragments for *O*-citryl-hydroxylamine.

a



b

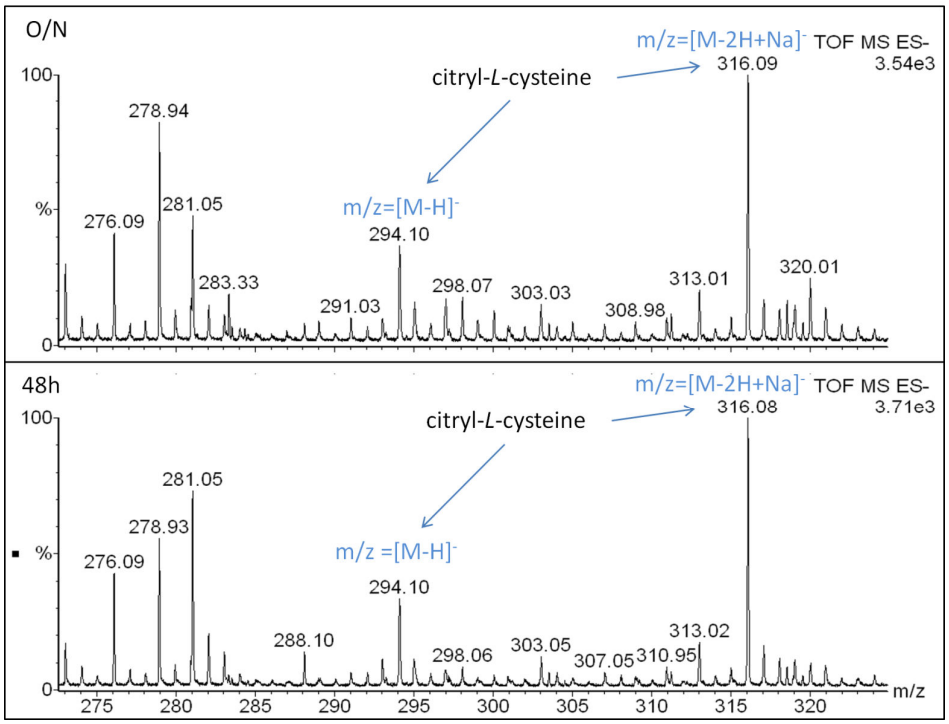


Figure 2.31: MS analysis of AcsD reaction with L-cysteine on the LCT ESI MS instrument (Micromass, Manchester, UK). **(a)** Only in reactions where AcsD was present a new peak with an *m/z* value of 294.11 was detected. This peak corresponds to the theoretical *m/z*= 294.03 $[M-H]^-$ value of citryl-L-cysteine. **(b)** Time points (O/N and 48h) of reaction with L-cysteine. While the *m/z*=294.1 peak is not increasing significantly, a second peak corresponding to citryl-L-cysteine Na^+ adduct (*m/z*=$[M-2H+Na]^-$) increases slowly.

2.3.9.2 Structural studies

Attempts to co-crystallize AcsD with enzymatic derived *O*-citryl-L-serine or *O*-citryl-ethanolamine have not been successful. However, the ability of AcsD producing amides by utilizing L-2,3-DAPA or ethylenediamine (this work and Berti and Thomas 2009) was used to obtain a co-complex structure. AcsD co-crystallization trials were set up with supernatant of incubated AcsD reactions containing L-2,3-DAPA or ethylenediamine, as well as Mg^{2+} and ATP. While both experiments resulted in good diffracting crystals (up to 2Å), only those grown from pre-incubated ethylenediamine reactions showed clear additional electron density for a molecule larger than citric acid and then only in subunit B. This density was interpreted and refined as *N*-citryl-ethylenediamine (Figure 2.32 and Figure 2.4).

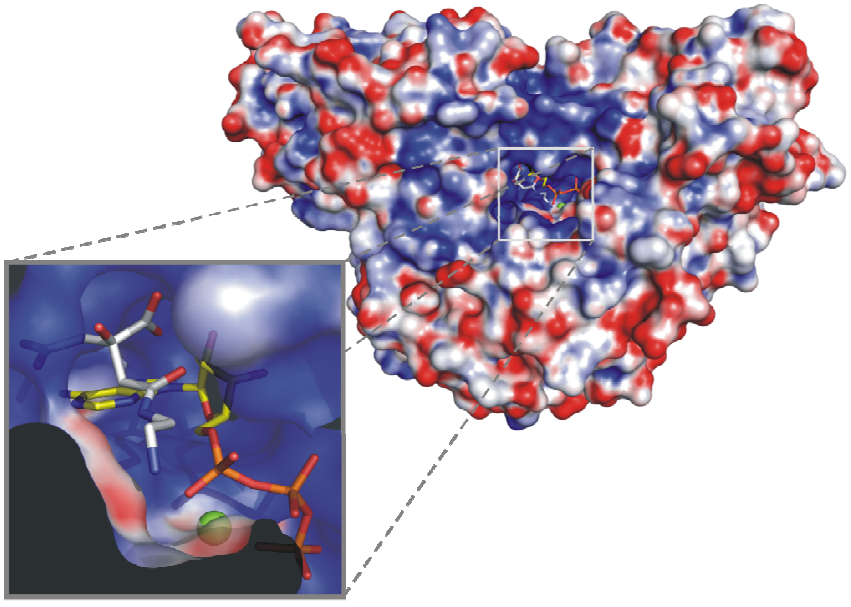


Figure 2.32: AcsD citryl-ethylenediamine (EDA) co-complex. Electrostatic surface representation of AcsD with bound ATP (colored in yellow), citryl-EDA (colored in white) and Mg^{2+} ion (colored in green).

Data collection and refinement statistics of *N*-citryl-ethylenediamine (citryl-EDA) co-complex structure are shown in Table 2.1. Akin to the ATP co-complex (Figure 2.18a) ATP and Mg^{2+} are also bound in the citryl-EDA structure (Figure 2.33a). Most of the citryl moiety of citryl-EDA is coordinated in the same manner as citric acid in the AcsD citrate co-complex (Figure 2.18d, Table 4.2 and Table 4.4). However, there is one crucial difference: the carbon atom involved in forming the amide bond to ethylenediamine has moved away from its position in the citrate co-complex (Figure 2.33b). This change is presumably driven by the presence of the α phosphorous of ATP which would otherwise clash with ethylenediamine. The EDA moiety of citryl-EDA is mainly coordinated by the charged residues E442 and R501 (Figure 2.33a).

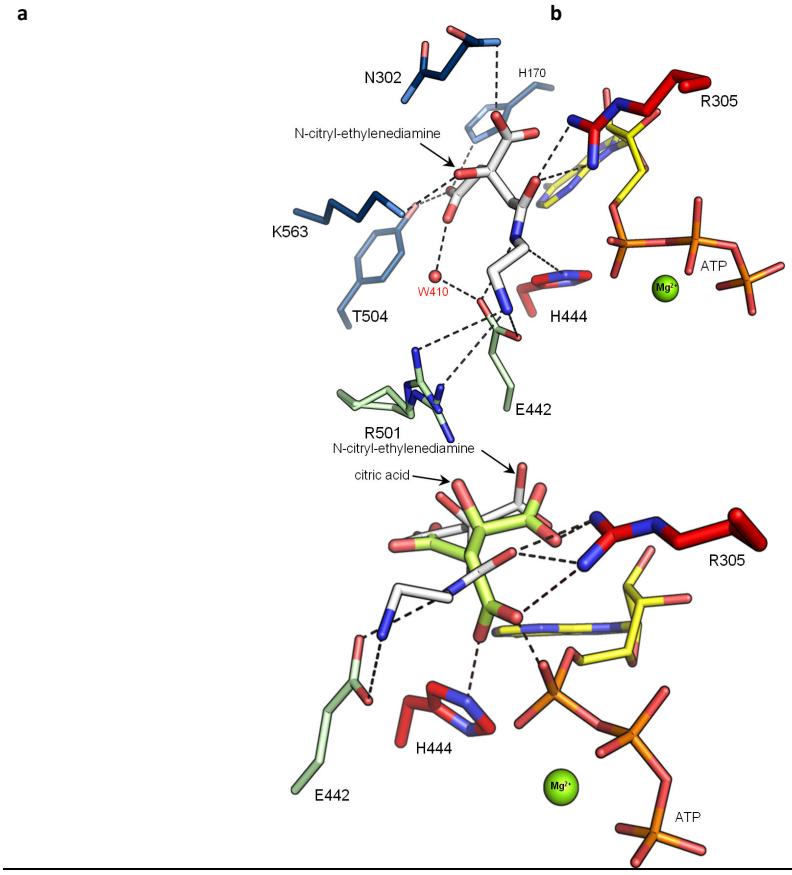


Figure 2.33: AcsD citryl-ethylenediamine coordination. **(a)** Coordination of citryl-EDA; carbons of EDA colored white. **(b)** Superposition of citryl-EDA (colored in white) and citric acid (colored in green) demonstrating the different position of *pro*-R carboxylate in relation to the amide bond of citryl-EDA.

2.3.9.3 Mutational studies

Based on the obtained structural citryl-EDA data, R501 and E442 were mutated and tested for enzyme activity with selected nucleophiles (Figure 2.34a). All tested nucleophiles have the crucial α-amino group and a nucleophilic hydroxyl or amino group, only L-serine comprised the additional carboxylate group. E442D and R501A mutants are inactive for all tested nucleophiles. However, E442Q did show activity (almost halved but above water) with one nucleophile, hydroxylamine. R501K utilizes both ethylenediamine and hydroxylamine at approximately the native rate. Most notably R501K is less active with L-serine, but 1,2-diaminopropane is utilized at a slightly higher rate (~120 %) than with *wt* AcsD.

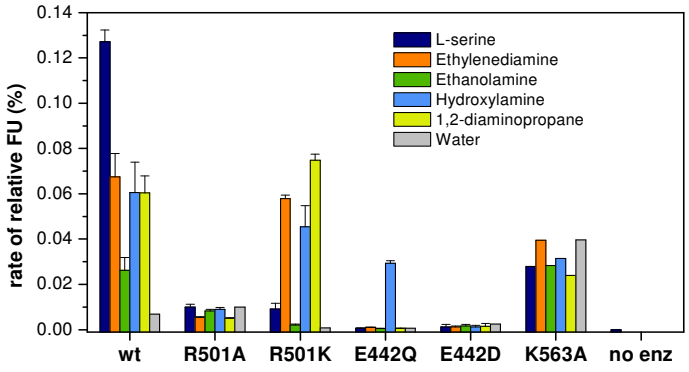


Figure 2.34: Mutational studies of nucleophile binding site. Mutation of residues involved in coordinating distinct functional groups of L-serine.

2.4 Discussion

2.4.1 Structural aspects

AcsD is a new member of adenylate-forming enzymes. A defining characteristic of adenylate-forming enzymes is that they activate the carbon atom of carboxylic groups by making an adenylate ester. The adenylate ester is a highly reactive leaving group which enhances the electrophilic nature of the carbon atom to be attacked by a nucleophilic substrate. Comparison of AcsD active site with the superfamily of adenylate-forming enzymes and the structurally distinct family of aminoacyl-tRNA synthetases, which perform a similar adenylation reaction (Chang et al. 1997), reveals some common chemical themes in catalysis such as the use of Mg^{2+} and positively charged residues to facilitate catalysis. There are however significant differences regarding Mg^{2+} coordination and spatial configuration of active site residues.

In the adenylation domain of GrsA, a NRPS involved in the biosynthesis of gramicidin S (Conti et al. 1997), a Mg^{2+} ion is bound to the α phosphate of AMP. In contrast, no Mg^{2+} ions were found in the active site of the crystal structure of Acyl-CoA synthetase or luc (Nakatsu et al. 2006). In the adenylate-forming enzymes superfamily Arg (R526 in Acetyl-CoA synthetase (Gulick et al. 2003)) or Lys (Lys517 in GrsA) or His (H247 in luc) are in close proximity to the α phosphate of ATP. In contrast AcsD has two conserved residues, H444 and R305, proximate to the α phosphate. Interestingly, the AcsD residues do not superimpose with those in the superfamily. tRNA synthetases can be split in two classes, class I uses one Mg^{2+} and

class II uses three Mg^{2+} ions which bridge α, β (class I, class II) and β, γ (class II) phosphates of ATP. Similar to the adenylating superfamily, a single positively charged residue is proximate to the α phosphate of ATP. These differences in structure and catalytic site organization suggest that AcsD represents a new class of adenylating enzyme superfamily.

2.4.2 AcsD is capable of assembling enantioselective citrate amides and esters

AcsD activates citrate for condensation with the second substrate L-serine. The hydroxyl group of L-serine functions as the nucleophile. Previous work had shown that amines could function in this second step, decomposing the citrate adenylate to generate an amide (Berti and Thomas 2009). Our data suggest that amines can function much more readily (as might be expected from simple chemical considerations) as nucleophiles. In biology this would present a problem as all amino acids have a free primary amine group (except proline). However, the combination of structural and biochemical data suggests that the enzyme solves the problem by strongly recognizing the α-amino group. Only two exceptions are found to this rule, the first is a simple molecule with only two non-hydrogen atoms one of which is an amine (hydroxylamine) and the second one is the compound L-isoserine. Neither of these molecules are expected to exist at significant concentrations in biological systems. The requirement for the α-amino group appears to be so strong that the compound ethanolamine (which has both an amine and hydroxyl moiety) reacts (within error) exclusively with the hydroxyl function rather than the more reactive amine. The carboxyl group is important but not essential and several compounds

(ethanolamine being one example) react without having this functionality. The data also suggest (but do not conclusively show) that the enzyme is capable of catalyzing the formation of the thioester linkage, but at a much slower rate than L-serine. Although sulfur is a stronger nucleophile, the enzyme apparently disfavors its presence. As a result of the much lower activity, it is unlikely to be a significant competitor in biology.

In order to better understand the enzyme requirements for this activity, the complex of *N*-citryl-ethylenediamine was determined (Figure 2.33a). This experimental guide for the recognition of the incoming nucleophile identifies two important residues, E442 and R501. Both mutants E442D and R501A are inactive with all substrates. Interestingly E442Q is weakly active with hydroxylamine but inactive with all the other substrates. R501K can utilize several amine nucleophiles but not L-serine. For E442, both the position and charge of the side chain are crucial, suggesting the residue is important for catalysis (Figure 2.34). The activity of the hydroxylamine can be explained by noting that its pK_a (< 8.0) is significantly below that of the other substrates. E442 might act as a general base deprotonating the incoming nucleophile and activating it. The situation with R501 is more complex, the data suggest that a positive charge is essential (R501A is inactive) but the spatial position is not important for nucleophiles which lack the carboxylate group as they react with essentially the same rate as native. Interestingly 1,2-diaminopropane reacts more readily with R501K than *wt*, an observation which can be attributed to the larger volume at the active site tolerating a secondary amine at the α-amino position. The spatial position is important but is not critical for L-serine which is still utilized by

R501K (but with 14 fold reduced rate). The likely role of R501 is to provide a positive charge which helps to bind and orient L-serine, but has no direct role in catalysis. Remarkably, R501K seems to favor nucleophiles forming amides over those resulting in esters, as seen for the low activity of L-serine or ethanolamine (Figure 2.34a).

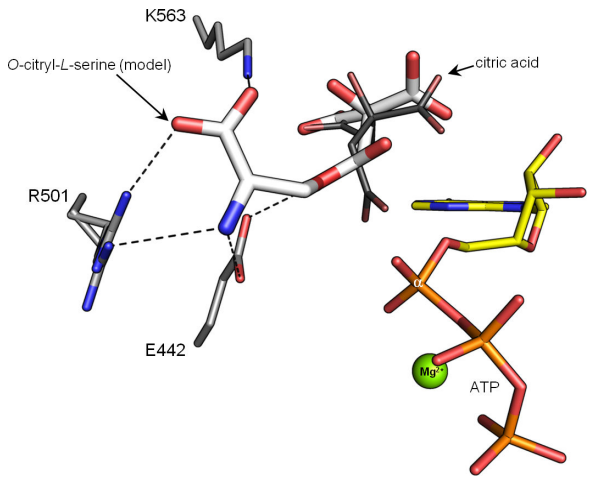


Figure 2.35: Model of coordinated *O*-citryl-L-serine. Coordination is based on the *N*-citryl-ethylenediamine data adding the carboxylate of L-serine, which is coordinated by K563 and R501. Note the different orientation of *pro*-R carboxylate of citric acid (colored in grey, facing the α phosphate) and the ester bond in the final product (moved away to avoid clashing with ATP).

The structural data of *N*-citryl-EDA was used to design a model for a *O*-citryl-L-serine complex (in essence adding a carboxylate group to the observed crystal structure) (Figure 2.35). This model suggests that R501 and K563 might coordinate the carboxylate enhancing the reactivity of carboxylate containing nucleophiles. K563 is involved in the recognition of citrate as well. A K563A mutant was constructed, but expresses only poorly and not all impurities could be removed during purification. The results on this crude material show that the mutant has no activity above water for any nucleophile (Figure 2.34). This suggests that the water rate has increased but this may simply be a feature of impurities. Nevertheless one can see that K563 is an

important residue for activity, its precise role other than in the recognition of carboxylate groups (citrate and L-serine) is unclear and still remains to be explored.

2.4.3 Molecular mechanism

2.4.3.1 AcsD forms an enzyme bound citrate adenylate intermediate

Superposition of the protein atoms of the ATP and adenosine citrate complexes of AcsD assist in constructing a model for the ternary complex (Figure 2.36a-c). This model indicates that the essential Mg^{2+} ion will polarize the α phosphorus oxygen bond, enhancing its electrophilicity. Furthermore, Mg^{2+} will stabilize the generated pyrophosphate as well. Interestingly, the α - γ phosphate coordination of Mg^{2+} in AcsD is rare and only found in seven out of 250 ATP containing structures. The absolutely conserved residues R305 and H444 make contacts with the α phosphate and *pro-R* carboxylate of citric acid. H444 seems a likely candidate to act as a general base to deprotonate the *pro-R* carboxylate of citric acid creating a nucleophile. Both residues (H444 and R305) will further polarize the phosphorus oxygen-bonds (enhancing the electrophilicity of the α phosphate) and also stabilize the negatively charged (deprotonated) form of the citrate carboxylate. Once activated, the citric acid nucleophile is able to attack the exposed α phosphate of ATP in an S_N2 type reaction, creating an enzyme coordinated citrate-AMP intermediate and PP_i (Figure 2.36d).

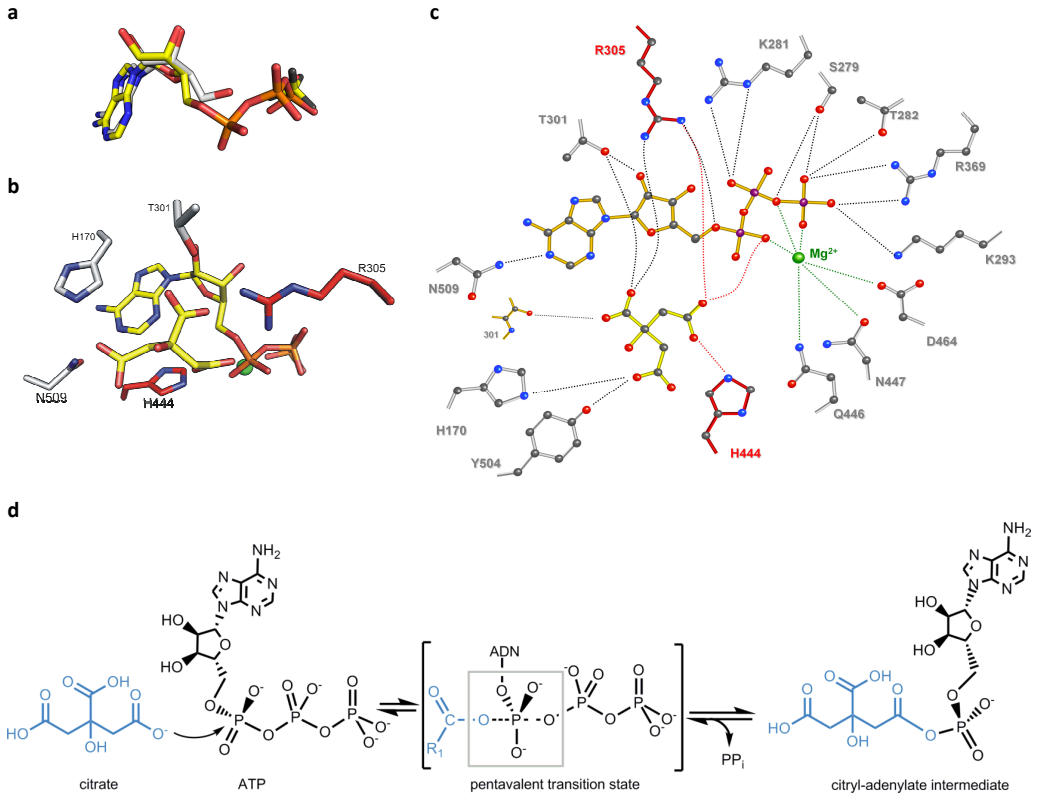


Figure 2.36: First step of AcsD mechanism. **(a)** Adenosine (carbon colored white) and sulfate (yellow/grey) from the citrate adenosine AcsD complex overlap with ATP (carbon colored yellow) from the ATP AcsD complex (based on superposition of protein atoms). **(b)** A model of ternary complex of AcsD, generated by superimposing the ATP and citrate complex. The color scheme is preserved, except H444 and R305 which are highlighted in red. **(c)** Schematic representation of the ternary complex. Hydrogen bonds are shown with doted lines. Catalytic residues are colored in red. **(d)** Reaction of citrate and ATP via a pentavalent transition state forming the enzyme coordinated citryl-adenylate intermediate.

2.4.3.2 AcsD liberates *O*-citryl-L-serine and AMP

The second step in AcsD reaction is the decomposition of the adenylate intermediate by L-serine. The attack of L-serine has to occur from the face of citrate which is exposed to the solvent, as the other face is shielded by the ribose of ATP. On this face E442, R501 and K563 are positioned to recognize L-serine as a model of the likely second ternary complex demonstrates (Figure 2.37a). The nucleophilic hydroxyl of the manually docked serine is located in a similar position as a water molecule

found in the ATP co-complex structure (W140). This location is optimal for the nucleophilic attack at the carbonyl group required for the decomposition of citrate adenylate. Interestingly, the docked model superimposes well with the EDA moiety of the *N*-citryl-EDA co-complex (Figure 2.37b). Mutational studies show that E442 could be the crucial catalytic residue for the second part of AcsD reaction, polarizing and coordinating the respective nucleophile (Figure 2.33). The polarized hydroxyl of L-serine attacks the *pro*-R carbonyl of the citryl-adenylate intermediate liberating *O*-citryl-L-serine and AMP (Figure 2.37c).

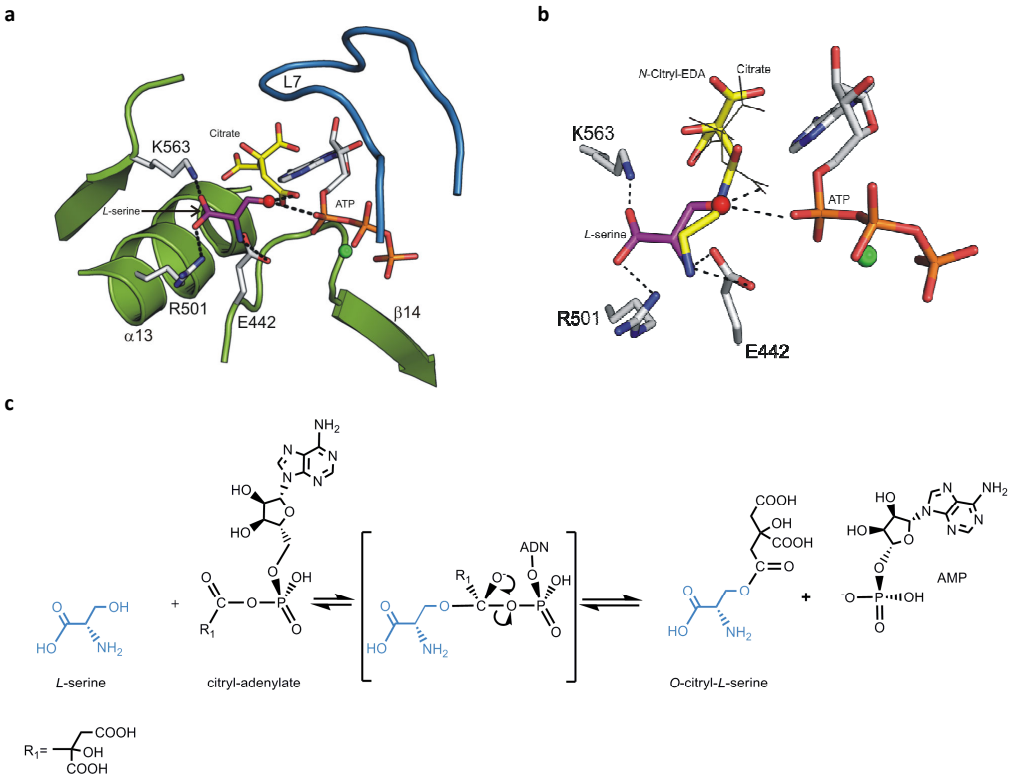


Figure 2.37: Second step in AcsD reaction. **(a)** Model of the second ternary complex. **(b)** Superposition of L-serine model and the *N*-citryl-EDA co-complex. The docked L-serine superposes well with the EDA moiety. **(c)** Chemical representation of second AcsD reaction step. The generated citryl-adenylate intermediate (Figure 2.36d) is attacked by L-serine forming *O*-citryl-L-serine and AMP.

2.4.3.3 The fate of pyrophosphate

Most adenylate-forming enzymes release pyrophosphate immediately after formation of their adenylate such as GrsA or luc, two Acyl CoA synthetases (Conti et al. 1997; Gulick et al. 2003; Nakatsu et al. 2006). However acetyl-CoA synthetases and class II tRNA synthetases bind the PP_i in a pocket using conserved His and Arg (H270 and R480 in Acetyl-CoA synthetase) residues to stabilize it. In AcsD, candidates for this role are the conserved residues R369 and R469. These residues are located in a hydrophilic pocket, which could accommodate PP_i (Figure 2.19). Though F291 in AcsD is far away from the active site a mutation to tyrosine reduces the catalytic activity (Figure 2.23). The larger tyrosine most likely displaces a water molecule at the bottom of ATP cavity and hence confines the possible space to accommodate pyrophosphate. These results suggest that in contrast to other adenylate-forming enzymes AcsD may keep PP_i in this cavity after the first reaction step and liberate it after formation and release of *O*-citryl-L-serine.

2.4.4 Mechanistic link to kinases

Like AcsD, kinases catalyze nucleophilic attack at and cause breakage of phosphate ester bonds. Kinases however catalyze the attack of a nucleophilic substrate at the γ phosphate. But in AcsD the citrate carboxyl group attacks at the α phosphate. Structurally the ATP binding site of AcsD shares elements of the cAPK template, including a metal binding site but crucially lacks the characteristic mobile glycine rich loop (Taylor et al. 1999).

The cAPK fold binds the triphosphate in roughly the same position relative to the secondary structure (Figure 2.20). However, the vectors from α to γ phosphates have opposite directions and the adenosine rings do not superimpose (Figure 2.21b). As a result the key phosphate atoms are located in the same relative position in both AcsD and cAPK (Figure 2.38 and Figure 2.21). Hence both AcsD and kinases may share a similar reaction mechanism.

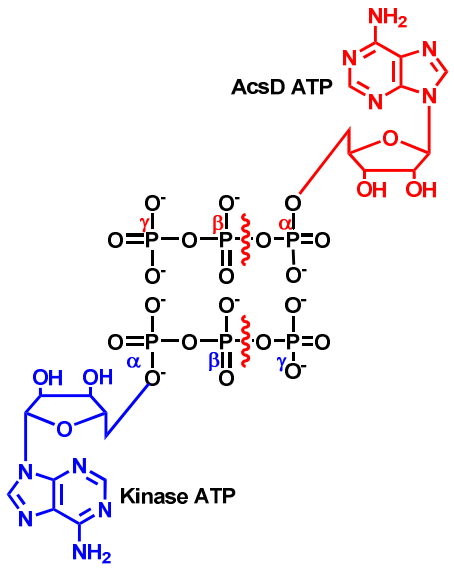


Figure 2.38: Comparison of AcsD and kinase ATP. Schematic illustration of the ATP orientation in AcsD (in red) and in kinases (in blue). The key electrophilic α phosphate in AcsD and γ phosphate in kinases superimpose.

Based on these similarities it is possible that AcsD may have an evolutionary link to kinases. In this context it has to be noted that PurT, a ribonucleotide transformylase involved in purine biosynthesis (Thoden et al. 2002), transfers the γ phosphate to a formyl group creating a reactive intermediate. Thus the enzyme combines kinase like activity (attack at γ phosphate) with AcsD like activity (creation of a reactive intermediate) and may represent a physical link between these families.

2.5 Conclusion

The finding that AcsD catalyzes highly stereoselective desymmetrization of citrate by adenylation of one of its prochiral carboxylate groups indicates that AcsD and other type A NIS synthetases may find use as novel biocatalysts for the preparation of homochiral citric acid derivatives. The obtained structural information for AcsD provides the opportunity for rational mutagenesis to alter or broaden its substrate specificity, further increasing its potential for use in enantioselective biocatalysis.

Mutational and structural studies revealed two catalytic residues H444 and R305. The histidine deprotonates (as a general base) the incoming citric acid, but may also stabilize the pentavalent transition state in addition to R305 and Mg^{2+}.

It was shown that AcsD is able to utilize other nucleophiles than L-serine or ethanolamine. The systematic screening for potential substrates unveiled certain roles for each functional group of the nucleophilic substrate. While the α-amino group is crucial for coordination, the less important carboxyl moiety enhances nucleophilic binding. However, the nucleophilic moiety was shown to be variable. Three nucleophilic groups were tested (-OH, $-NH_2$, -SH) generating citrate esters, amides or thioesters that were verified by MS and MS/MS analysis. However, fragmentation of the thioester product resulting from reaction with cysteine was not successful. Instead, a new mass peak corresponding to citryl-L-cysteine was detected on the LCT ESI MS instrument. The *N*-citryl-ethylenediamine complex identified a possible binding site for the nucleophilic substrate. Mutational studies confirmed that E442 is a key catalytic residue for the second step in AcsD reaction. Along with R501 it also coordinates the crucial α-amino group of the nucleophile. A docked

model of *O*-citryl-L-serine, based on the *N*-citryl-EDA complex, shows that K563 and R501 are likely to coordinate the nucleophilic carboxylate.

Interestingly the R501K mutant shows a surprising preference for amide forming nucleophiles and utilizes 1,2-diaminopropane 20 % faster than *wt* enzyme. Such changes in specificity are often useful for biotransformations. Hence, AcsD and other NIS synthetases have great potential for mutational engineering to broaden their substrate range and specificity. Harnessing the ability of AcsD to desymmetrize organic acids by making amides or esters would have valuable applications in organic chemistry. Such a rational design has always been a challenge. Furthermore, a clear understanding of the substrate recognition sites could assist in the development of AcsD inhibitors. These are likely to affect also homologues of human pathogens. One feasible target is AsbA, which is catalyzing the initial step in petrobactin biosynthesis in *Bacillus anthracis* (Oves-Costales et al. 2007). Thus AcsD inhibitors may be the starting point for designing novel antibiotics used to treat anthrax or other bacterial infections.

2.6 Future work

AcsD generates enantioselective citrate amide and ester derivatives. This could be useful in commercial biotransformation or biosynthetic chemistry. This was exemplified by one mutant, R501K, which is able to utilize 1,2-diaminopropane faster than the *wt* enzyme producing *N*-citryl-diaminopropane. Rational mutagenesis of the nucleophilic binding site will be one of the future targets. Another yet unanswered question is due to the fact that E442Q is inactive for all tested nucleophiles, but not for hydroxylamine. So far only six of approximately twenty nucleophiles have been tested. The remaining ones have to be analyzed to finally reveal if hydroxylamine is the only nucleophile utilized by E442Q.

AcsD is specific for the first substrate, citric acid (Berti and Thomas 2009; Schmelz et al. 2009). Hence, modification of citrate binding site may enlarge AcsD specificity for other carboxylic acid substrates. In this study, T301 (coordinating the central citrate carboxylate) was mutated leading to an inactive enzyme (T301A) when tested with citrate. Possible other candidates for rational mutagenesis are K563, Y504 and N302. Especially, substitution of Y506 and K563 (coordinate the citrate *pro-S* carboxylate) with a smaller amino acid would generate more space for a larger and longer carboxylic acid such as 2-hydroxy-1,2,4-butanetricarboxylic acid or 2-hydroxy-1,2,5-pentane-tricarboxylic acid (Figure 2.39). Mutagenesis of these substrate binding sites is an opportunity to alter or broaden AcsD substrate specificity and further increase its potential for use in enantioselective biocatalysis.

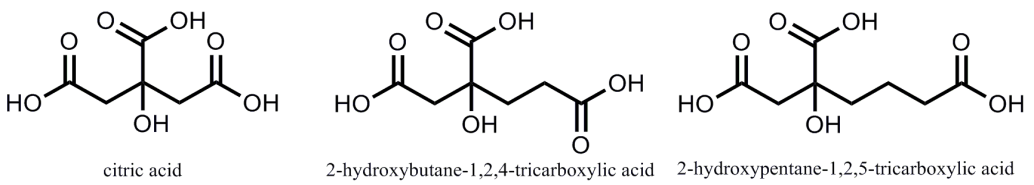


Figure 2.39: Likely citrate analogues for future AcsD mutants.

Finally, it was shown that AcsD favors L-serine over the proposed substrate ethanolamine. However, recently Berti et al. (2009) demonstrated that citryl-ethanolamine or *N*-citryl-1,3-diamionpropane, but not *O*-citryl-L-serine, is incorporated in a functional siderophore, which is able to restore growth of a siderophore-deficient *Pseudomonas syringae* strain (Berti and Thomas 2009). In their experiment combination of all three biosynthesis enzymes (AcsD, AcsC and AcsA) were used with several likely substrates to assemble achromobactin or its derivatives in-vitro. Products were analyzed by mass spectrometry and added to plated cultures of a siderophore-deficient strain. They showed that reactions with L-serine result in formation of α-ketoglutaryl-diaminobutyryl-citryl-serine. This product still lacks a final α-KG moiety to be added by AcsA to yield achromobactin. This reaction seems not to take place probably due to the not yet eliminated carboxyl group introduced by L-serine. AcsE, a postulated PLP-dependent decarboxylase, could catalyze decarboxylation of α-ketoglutaryl-diaminobutyryl-citryl-serine or a prior product such as *O*-citryl-L-serine to allow assembling of the final siderophore. This could be examined by addition of AcsE in the experiment described early (Berti and Thomas 2009). This would finally make it clear whether ethanolamine or L-serine is the natural product of AcsD.

b

Chapter 3 - AlcC

3.1 Introduction

3.1.1 Summary

In this chapter the structural and biochemical work on the type C NIS enzyme AlcC is reported. AlcC is involved in biosynthesis of alcaligin, a hydroxamate siderophore synthesized by iron-starved *Bordetella pertussis* and *B. bronchiseptica*, both being mucosal pathogens of the upper respiratory tract of mammals (Moore et al. 1995). While *B. pertussis* is a human pathogen causing whooping cough, *B. bronchiseptica* causes endemic respiratory diseases in other mammalian hosts such as swine, cats and dogs (Goodnow 1980). AlcC is a processive enzyme performing the final two steps in alcaligin biosynthesis (Challis 2005; Brickman et al. 2007). AlcC is predicted to dimerize *N*-(4-Amino-3-hydroxy-butyl)-*N*-hydroxy-succinamic acid in an ATP driven reaction. This product then is cyclized in a second AlcC reaction liberating alcaligin.

The co-complex of adenosine and ATP were obtained. The ATP structure shows that the ATP phosphate binding site is conserved within NIS enzymes and the adenosine moiety is coordinated by a network of water molecules. The fluorescence based AMP activity assay used for AcsD establishes that AlcC also performs an adenylation reaction. Detailed analysis of the MS/MS fragments of AlcC reaction products with the substrate analogue *N*-hydroxy-*N*-succinylputrescine identified not only substrate dimers, but also the dimer macrocycle. AlcC is also able to generate *N*-hydroxy-*N*-succinylputrescine trimers and a trimer macrocycle. It was not possible to obtain any substrate co-complexes (apart from ATP and adenosine structure). Hence, binding models of *N*-(4-Amino-3-hydroxy-butyl)-*N*-hydroxy-succinamic acid, alcaligin and a *N*-hydroxy-*N*-succinylputrescine trimer macrocycle were developed. The modeled

substrate fits perfectly in a cavity and faces the α phosphate of ATP to perform a nucleophilic attack. Models of alcaligin and the trimer macrocycles demonstrate that there is enough space in the active site to accommodate them.

3.1.2 AlcC apo structure

The apo structure of AlcC was solved, built and partly refined before I started working on this project. A monomeric poly-alanine model of AcsD apo-structure was used as molecular replacement model. The unit cell of AlcC contains one monomer per asymmetric unit. PISA (Krissinel and Henrick 2007) identifies a stable head-to-

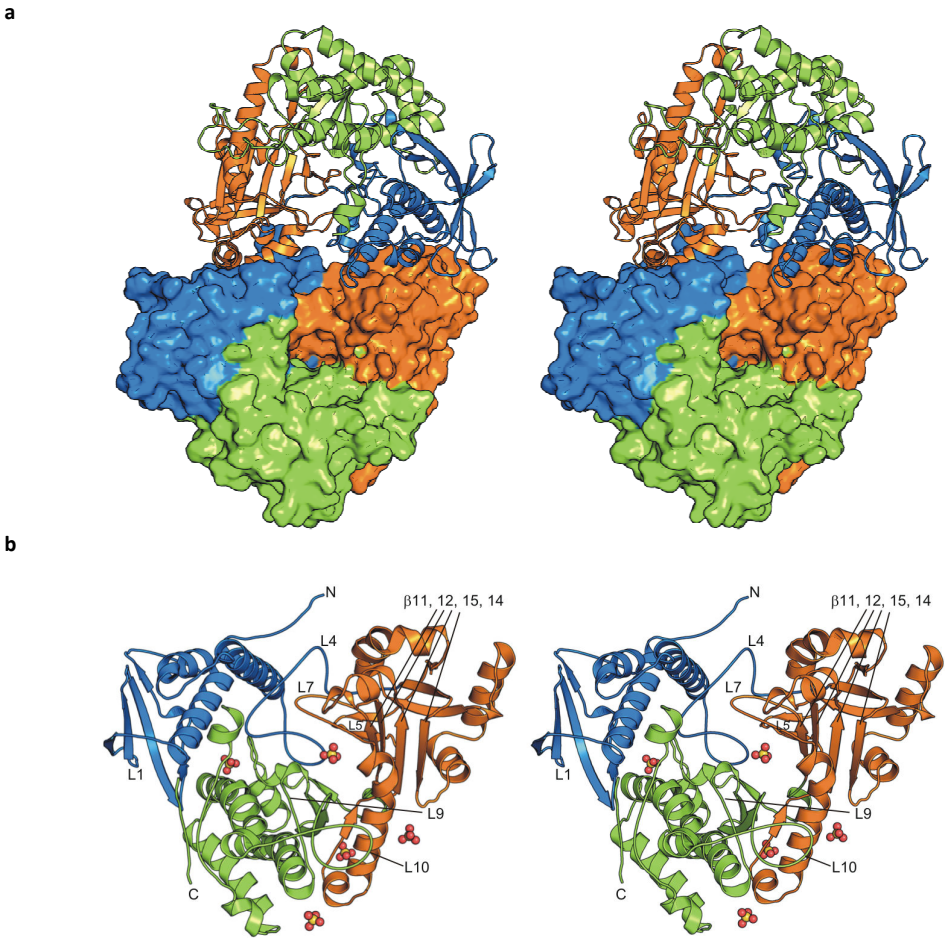


Figure 3.1: AlcC apo-structure. **(a)** AlcC dimer in stereo view shown in cartoon and surface representation. N-terminal domain is colored in blue, middle domain in orange and C-terminal domain in green **(b)** The AlcC apo structure binds five sulfate ions, one located in the active site (shown in ball and stick representation).

head dimer with a complexation significance score (CSS) of 1 (Figure 3.1a).

The active site of each monomer is formed by three domains: N-terminal, middle and C-terminal domain. The N-terminal domain (colored in blue, residues 7 – 138) forms a three helix bundle flanked by five anti-parallel β-sheets (Figure 3.1b). The middle domain (colored in orange, residues 193-389) is linked to the opposite N-terminal domain by loop L4 (residues 156-192). It is folded into a 9 stranded anti-parallel β-sheet interspersed with 5 α-helices and conserved loops L5 and L7. The C-terminal domain (colored in green, residues 390 – 601) is predominantly α-helical (8 helices) and leads back to the N-terminal domain. In the apo-structure five SO_4^{2-} ions are bound, one of them is located in the active site (Figure 3.1b). The AlcC topology shows significant similarities to the previously discussed AcsD structure.

3.1.3 AlcC substrates

In alcaligin biosynthesis AlcB is proposed to carboxylate *N*-hydroxy-putrescine with succinic acid to form *N*-hydroxy-*N*-succinylputrescine (Sub204) (Figure 1.10c). In the subsequent reaction, AlcE adds a hydroxyl and generates *N*-(4-Amino-3-hydroxy-butyl)-*N*-hydroxy-succinamic acid (Sub220), the proposed substrate of AlcC. In a subsequent AlcC reaction an internal amide bond is formed generating the macrocycle alcaligin (Figure 3.11a). AlcC was also tested to utilize Sub204, the product of AlcB, as a substrate analogue. Other than with Sub220 just Sub204 dimers are formed, but not the final macrocycle putrebactin (Figure 3.11b) (Challis, unpublished data). Therefore, alcaligin biosynthesis shares parallels with the recently described putrebactin biosynthesis from *Shewanella* species (Kadi et al. 2008).

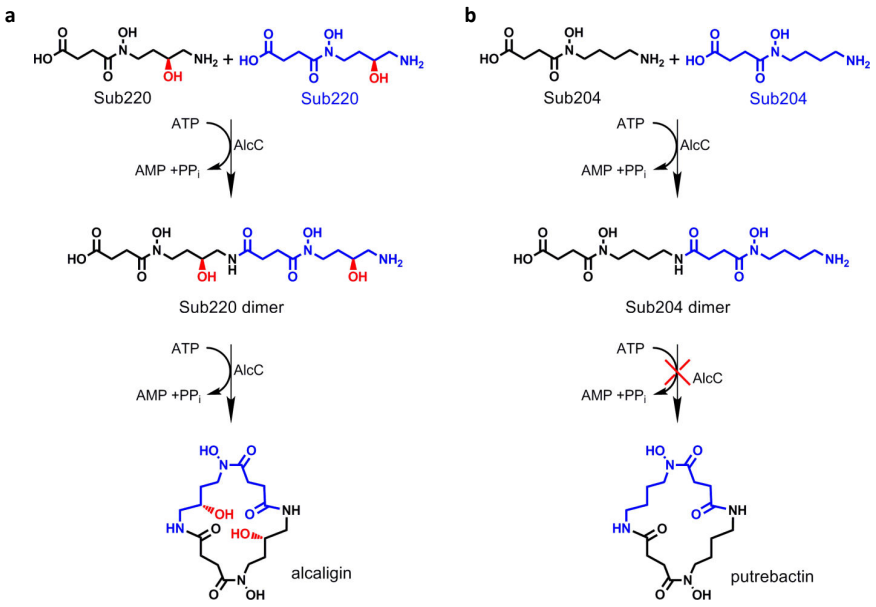


Figure 3.2: AlcC substrate consumption. **(a)** Two molecules of the proposed substrate *N*-(4-Amino-3-hydroxy-butyl)-*N*-hydroxy-succinamic acid (Sub220) are dimerized and then the "open" ends are joined to form alcaligin. **(b)** AlcC was shown to utilize *N*-hydroxy-putrescine (Sub204) generating dimers, but not the macrocycle putrebactin (Challis, unpublished data).

3.2 Materials and methods

3.2.1 Overproduction and purification of AlcC

The pET151/D-TOPO *alcC* construct was provided by Prof. Challis, University of Warwick. It codes for the protein fused with an N-terminal TEV cleavable His_6-tag. For protein over-production the *alcC* plasmid was transformed into BL21(DE3) *E.coli* cells. 1L cultures were launched from 10 mL O/N starter cultures as described for AcsD (section 2.2.1). When cell density reached an optical density A_{600}=0.8 protein overproduction was started by adding 0.5 mM IPTG. The temperature was lowered to 15 °C and cultures further incubated O/N. For cell harvest, subsequent protein purification and tag-removal the evolved AcsD protocol was adopted and employed. For detailed information see section 2.2.1.

After tag removal AlcC has six additional residues at the N-terminus (GIDPFT) with a final weight of 70604 Da. Purified AlcC was concentrated to 5.6 mg/mL in 50 mM Tris pH 7.5, 500 mM NaCl, 10 % (v/v) glycerol using protein concentrating devices (50 kDa Vivaspin). Concentrated protein was divided into 25 µL aliquots that were flash frozen in liquid nitrogen and stored at –80 °C. The purification progress was monitored by SDS-PAGE. The identity and integrity of AlcC was confirmed by mass spectrometry. From one liter of cell cultures ~1.9 mg pure AlcC protein was purified.

3.2.2 Crystallization, data collection and refinement of AlcC

3.2.2.1 AlcC apo structure

Apo crystals were grown by Xiaoxuan Qi during her MRES studies in SPoRT, March 2007. Her optimized crystal condition was 1.4 $MgSO_4$, 0.1 M sodium cacodylate (pH 6.0) and 0.1 M Li_2SO_4. Crystals were cryo-protected by transferring into 2.25 M $MgSO_4$ and 100 mM sodium cacodylate (pH 6.0). Data were indexed and scaled using XDS and XSCALE (Kabsch 1993). The apo structure was solved with a poly-alanine model of AcsD molecule A as a molecular replacement model. This work, the manual fit to various maps and the initial refinement were carried out by Dr Kenneth Johnson, a former member in SPoRT, before I started working on this project. A final refinement was performed to meet the criteria of the structure validation tool MolProbity (Davis et al. 2007). The structure was refined using REFMAC5 (Murshudov et al. 1997) and COOT (Emsley and Cowtan 2004) was used for manual manipulation. The apo structure shows electron density for five sulfate ions, one located in the bottom of the active site. Data collection and refinement statistics are displayed in Table 3.1. Electron densities of ligands (discussed in the following sections) are shown in Figure 3.3.

3.2.2.2 Adenosine-SO_4^{2-} co-complex structure

Adenosine-SO_4^{2-} complex crystals grew in hanging drops from equal mixtures of 0.1 M MES pH 6.5, 1.6 M $MgSO_4$, 0.14 M Li_2S0_4 and protein solution (2.5 mg/mL AlcC, 45 mM MES pH 6.5 and 5 mM adenosine, pre incubated for 30 min on ice) to full size in 2 weeks at 20 °C. Crystals were cryo-protected by increasing the $MgSO_4$

concentration to 2.3 M and cryo-cooled at -173 °C. A data set was collected at a beam wavelength of 0.9835 Å on ID14-4 at European Synchrotron Radiation Facility (ESRF, Grenoble). The detector was set to a distance of 288 mm. Data collection was carried out with an exposure time of 1.0 s (14 % transmission) over 125° starting at 20.0° phi with an oscillation range of 0.5° and no overlap. The data set was indexed and integrated using DENZO and scaled with SCALEPACK (both implemented in the HKL2000 suit (Otwinowski and Minor 1997)) to 2.1 Å. The 2.4 Å apo structure was used to phase the 2.1 Å adenosine SO_4^{2-} complex. The structure was refined using REFMAC5 (Murshudov et al. 1997) and COOT (Emsley and Cowtan 2004) was used for manual manipulation.

3.2.2.3 ATP co-complex structure

Crystals grew in hanging drops from equal mixtures of 0.1M MES pH 6.5, 1.7 M $MgSO_4$, 0.18 M Li_2SO_4 and protein solution (2.5 mg/mL AlcC, 45 mM MES pH 6.5 and 5 mM adenosine, 5 mM *N*-hydroxy-*N*-succinylputrescine (Sub204), pre incubated for 30 min on ice) to full size (approx. 0.2 x 0.4 mm) in 2 weeks at 20 °C. Crystals were soaked in 25 mM ATP, 25 mM substrate 204 and 4 M sodium formate for 30 seconds and flash-frozen in liquid nitrogen. A data set was collected at a wavelength of 0.9796 Å on the IO2 beam line at Diamond Light source Ltd, UK. The detector was set to a distance of 288 mm. Data collection was started at 75° phi for 180° with an oscillation range of 0.5°. Data were indexed and integrated using DENZO and scaled with SCALEPACK (both implemented in HKL2000 (Otwinowski and Minor 1997)) to 1.96 Å. The 2.4 Å native structure was used for phasing the 1.96 Å ATP complex structure, which was refined using REFMAC5 (Murshudov et al. 1997) and COOT

(Emsley and Cowtan 2004). In the active site extra electron density was only observed for ATP, but not for Sub204.

Table 3.1: AlcC data collection and refinement statistics.

	Apo	**Adenosine –SO_4^{2-}**	**ATP**
Data collection			
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Cell dimensions (Å)	*a*= 93.7, *b*= 129.0, *c*= 48.3	*a*= 92.9, *b*= 128.9, *c*= 48.2	*a*= 94.6, *b*= 129.4, *c*= 46.5
Resolution (Å) (high resolution)	40.0-2.4 (2.49-2.4)	45.2 -2.1 (2.18-2.1)	50.0-1.96 (1.96-2.03)
R_{merge}	8.6 (49.4)	11.2 (58.7)	9.3 (55.2)
$I/\sigma(I)$	17.1 (3.3)	12.54 (2.1)	20.8 (3.5)
Completeness (%)	97.9 (99.1)	99.6 (99.9)	99.9 (99.8)
Aver. redundancy	4.1 (3.7)	4.8 (4.3)	7.2 (6.8)
V_m(Å^3/Da)	2.1	2.1	2.0
Solvent (%)	41	41	40
Software	XDS/XSCALE	HKL2000	HKL2000
Refinement	Phenix.refine	Refmac5	Refmac5
Unique reflections	23068	34630	41914
R_{work} / R_{free}	18.2/23.6	19.9 / 24.8	18.2 / 22.1
No. atoms			
Protein	4631	4601	4726
Water	180	285	257
ATP	0	0	31
Adenosine	0	19	0
Mg^{2+}	0	0	2
SO_4^{2-}	25	20	0
B-factors (Å^2)			
Protein	32	28	21
Water	31	32	27
ATP	0	0	17
Mg^{2+}	0	0	20
Adenosine	0	39	0
SO_4^{2-}	57	56	0
Rmsd bonds (Å) / angles (°)	0.003 / 0.725	0.011 / 1.261	0.009 / 1.173
B-factor deviation Bond/angle (Å^2)			
Main chain		0.754 / 1.256	0.554 / 1.070
Side chain		1.808 / 2.760	1.665 / 2.795
MolProbity score (percentile)	96	99	93

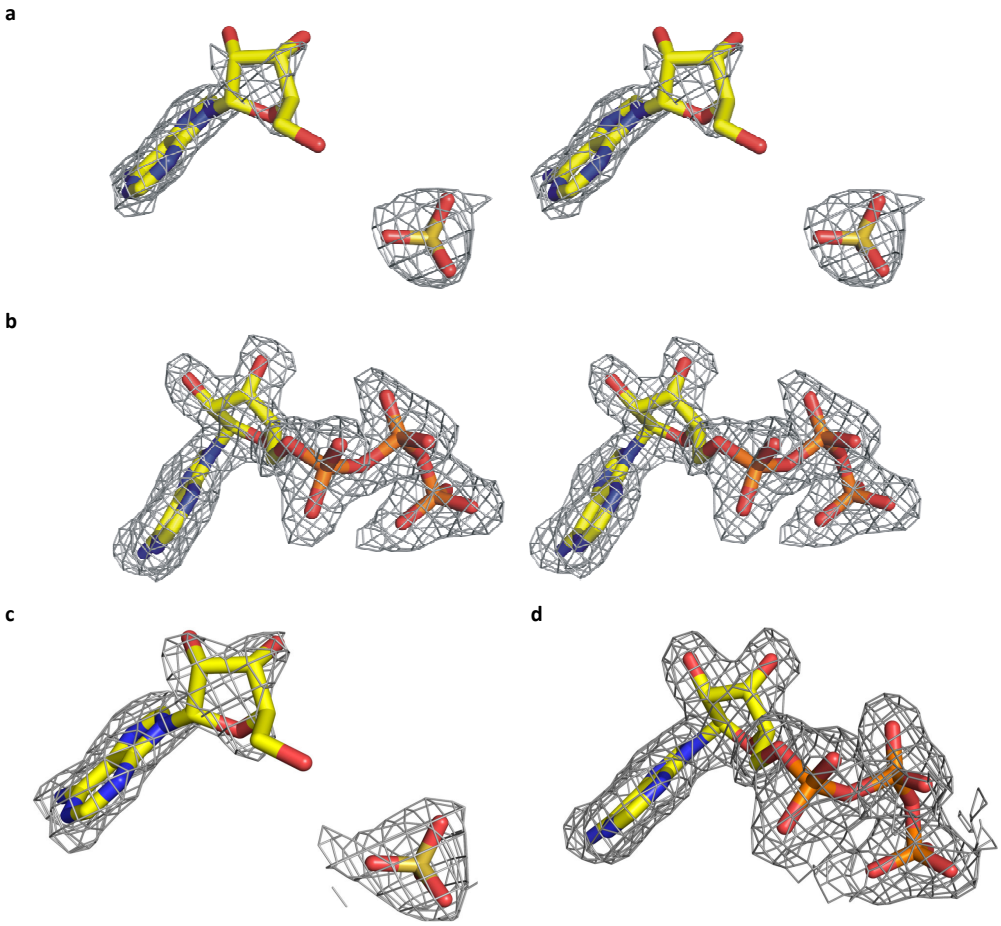


Figure 3.3: Electron density of AlcC co-complex structures. **(a)** Stereo F_o-F_c map of ADN-sulfate co-complex at 2.5 σ. **(b)** Stereo F_o-F_c map of ATP co-complex at 2.5 σ. **(c)** $2F_o$-F_c map of ADN and SO_4^{2-} contoured at 1 σ. The electron density for C5 of the ribose moiety is weak and the C5 hydroxyl less well ordered. **(d)** $2F_o$-F_c map of ATP contoured at 1 σ.

3.2.3 AMP production assay

AlcC activity was tested in the AMP production assay, previously described for AcsD (section 2.2.5). The reaction contained 50 mM Tris-HCl buffer (pH 8.0), 3 mM ATP, 15 mM $MgCl_2$, 1.5 mM phosphoenol pyruvate, 0.25 mM NADH, 2.4 µM AlcC, 12.6 units of lactate dehydrogenase, 8.4 units of pyruvate kinase, 4 units of myokinase and 2.5 mM Sub204 at a total volume of 140 µL. The mixture was incubated at 20 °C for 5 min and the reaction started by adding 2.4 µM AlcC.

3.2.4 Mass spectrometry

Products of the AlcC timeline experiment with Sub204, a substrate analogue of AlcC was characterized by MS and MS/MS analysis. Enzymatic reactions were carried out in 50 mM ammonium acetate buffer (pH 8.0) containing 20-40 µM wt AlcC, 30 mM $MgCl_2$ and 5 mM Sub204. The mixtures were incubated at 20 °C and the reactions were initiated by adding 20 mM ATP. At certain time points 1 µL aliquots were removed, diluted into 50:50 acetonitrile:water (100 µL) and centrifuged at 12,000 rpm for one minute. The samples were analyzed using continuous syringe infusion (Harvard syringe pump, Harvard Apparatus, Kent, UK) into a LCT (Micromass, Manchester, UK) ESI MS instrument or a Q-Star XL (Applied Biosystems, Foster City, USA) ESI-QTOF instrument using the appropriate ionization mode. *M/z*s of interest were subjected to collisionally induced dissociation (CID) on the Q-Star XL instrument, with collision energy adjustment to give some intact signal remaining. The operation of instruments and data acquisition was carried out by Dr Catherine H Botting (BMS Mass spectrometry and Proteomics Facility, University of St Andrews).

3.3 Results

3.3.1 Protein purification

3.3.1.1 Immobilized-metal affinity chromatography

His_6-AlcC plasmid containing *E. coli* cells were grown, induced and lysed as described in section 3.2.1. The first purification step after cell lysis was carried out using Ni-beads (Figure 3.4a). His_6-AlcC was eluted with 500 mM imidazole in three fractions (lane 7-8). Most protein eluted in the first fraction (lane 7) at around 75 kDa and contained a lot of additional proteins that had to be removed in subsequent purification steps.

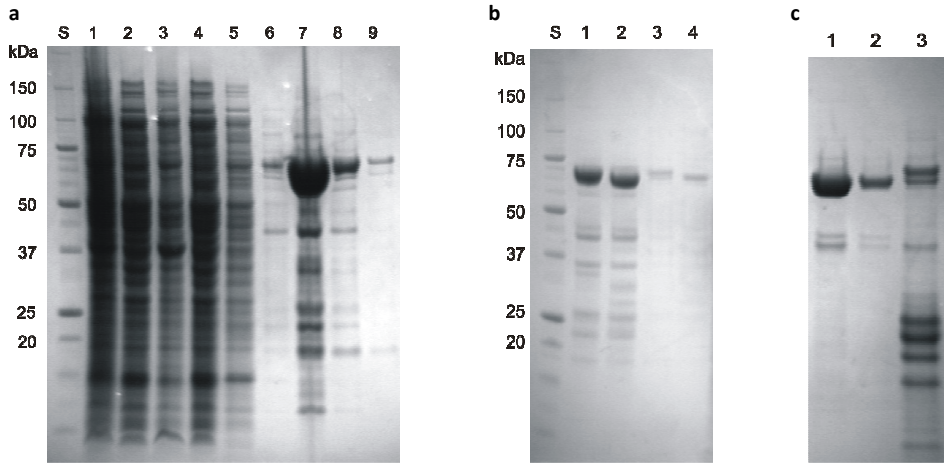


Figure 3.4: His_6-AlcC purification and TEV cleavage. **(a)** First purification of AlcC with immobilized-metal affinity chromatography (IMAC): **S** standard; **1** crude protein solution after cell breakdown; **2** supernatant (S/N); **3** pellet dissolved in 8 M urea; **4** flow through of IMAC column; **5** wash step 1; **6** wash step 2; **7** elution I; **8** elution II; **9** elution III; **(b)** TEV digestion with 1:50 TEV protease. **S** standard; **1** His_6-AlcC from elution I; **2** cleaved AlcC from elution I; **3** His_6-AlcC from elution II; **4** cleaved AlcC from elution II. **(c)** Second IMAC purification to remove the tag and impurities. **1** cleaved AlcC from EI where TEV protease and tag was removed; **2** Cleaved AlcC from EII where TEV protease and tag was removed; **3** Elution from 2nd nickel column. It contains a lot of impurities, which have been removed from cleaved protein in lane 1 and 2.

3.3.1.2 Desalting, tag removal and gel filtration

Eluted protein was directly desalted on a HiPrep 26/10 column to avoid protein degradation caused by imidazole. The elution profile indicates a peak (14 – 30 mL) corresponding to AlcC protein (Figure 3.5a). TEV-cleavage was carried out over night at RT and verified by SDS-PAGE (Figure 3.4b). Cleavage was successful with a protease to protein ratio of 1:50. TEV protease and cleaved tag were removed by a second IMAC purification step using a 5 mL pre-packed HisTrap column (GE Healthcare). Protein and wash buffer contained 50 mM imidazole to avoid TEV-cleaved AlcC from binding to the column. It was collected directly in the flow-through. Several impurities (*E. coli* proteins and non cleaved AlcC) have been removed through this purification step (Figure 3.4c).

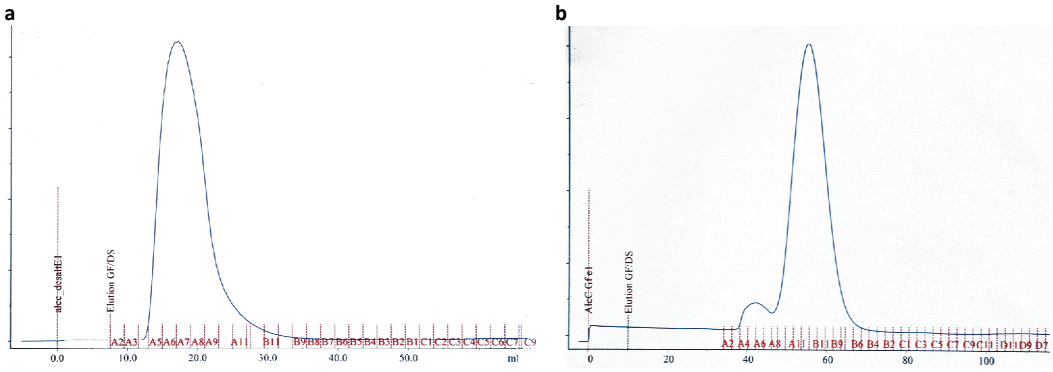


Figure 3.5: Desalting and gel filtration elution profiles for AlcC. **(a)** Elution profile of the desalting step on a HiPrep 26/10 desalting column to remove imidazole used for elution in the IMAC purification step. **(b)** Elution profile of the gel filtration step on a Superdex S200 column. Fractions A4 – A8 is aggregated protein found in the void volume. Protein in A9-B5 was identified as AlcC by mass spectrometry (see also Figure 3.6).

After tag removal, AlcC protein was concentrated and finally purified on a Superdex S200 (GE Healthcare) gel filtration column. The elution profile shows a Gaussian shape with its peak at a retention volume of 55 mL (Figure 3.5b). A second smaller peak that eluted in the void volume (~40 mL) is aggregated AlcC (Figure 3.6). The

purest fractions (A9-B12) were kept separate from less pure fractions (B11-B7). Both set of fractions were combined and concentrated for crystallization trails (see also section 3.2.1)

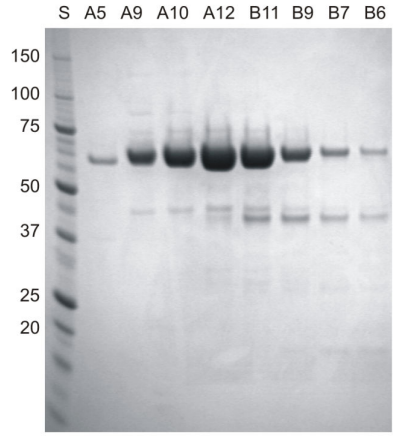


Figure 3.6: SDS-PAGE of AlcC gel elution profile. Fractions analyzed from gel filtration (Figure 3.5b) show that A5 (fraction of the void volume) contains degraded AlcC protein. Purest AlcC protein is found in fraction A9-A12, while fractions B11-B6 still contain impurities.

3.3.2 Crystallization of AlcC

3.3.2.1 Co-crystallization of various substrates

AlcC was found to form single, large orthorhombic crystals by using high concentrations of $MgSO_4$ as a precipitant. Five SO_4^{2-} binding sites were identified in the apo structure, whereas one is located in the active site. This high sulfate concentration made it rather difficult to co-crystallize any AlcC substrates in the active site. While most co-crystals still grew in perfect shape and reasonable size (some smaller and some larger than 0.2 x 0.4 mm), it was not possible to coordinate ATP, AMP, ADP or the substrate analogue Sub204 alone or in combination (e.g. ATP + Sub204, or AMP + Sub204). However, co-crystallization with 5 mM adenosine was successful (Figure 3.7a, Figure 3.3a,c), but not in combination with Sub204 (Figure 3.7b). Attempts to find a sulfate free crystal condition with commercially available crystal screens were not successful, even though parameters like protein concentration were changed or different buffers were tested. Screening in combination with micro-seeding of AlcC apo-crystal seeds failed as well.

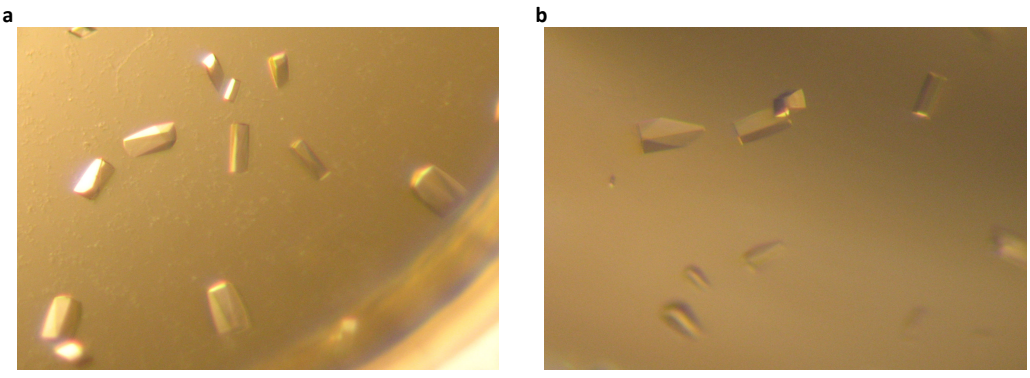


Figure 3.7: Co-complex crystals of AlcC. **(a)** AlcC adenosine – co-crystals grown with 5 mM adenosine. **(b)** Crystals grown with 5 mM Sub204 and 5 mM adenosine sill have a perfect shape, but don't coordinate any substrates, but sulfate.

3.3.2.2 Soaking of AlcC crystals

Since co-crystallization was shown to be not successful, other methods had to be tested. To enhance success of a soaking experiment, the possible binding site which was blocked by sulfate had to be “emptied” prior binding of substrate. This was obtained by transferring crystals into 4 M sodium formate. Thereafter, AlcC crystals were transferred into the soaking solution containing 10-200 mM substrate and 4 M sodium formate for a few seconds and flash-frozen in liquid nitrogen. This approach was successful when crystals were soaked with 25 mM ATP and 25 mM Sub204, binding ATP in the active site (Figure 3.3b, d), but not Sub204. Attempts to increase the soaking time (up to 22 h) or the concentration of Sub204 (with and without ADN or ATP) resulted only in electron density for ADN or ATP, but not for Sub204. Soaking with 15 mM ATP and 15 mM of desferrioxamine B (Figure 3.8), a substrate analogue for a final ring formation, failed. Here again only density for ATP was detected.

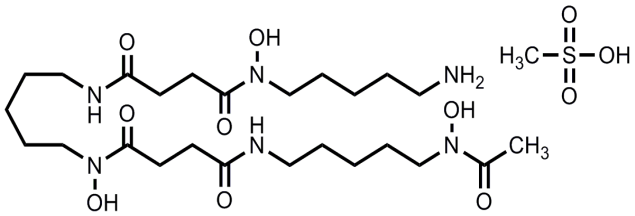


Figure 3.8: Chemical structure of desferrioxamine B mesylate.

3.3.3 Co-complex structures

The native protein and all co-complex structures are essentially identical with a root mean square deviation (rmsd) for 577 C_α atoms of 0.32 Å. No large conformational changes have been observed upon ligand binding. AlcC forms a deep cavity alike the AcsD active site. In this cavity ATP is coordinated between middle, C-terminal domain and loops 4 and 7 and is cradled in a position perpendicular to β-strands 12, 15 and 14 (Figure 3.9a, b).

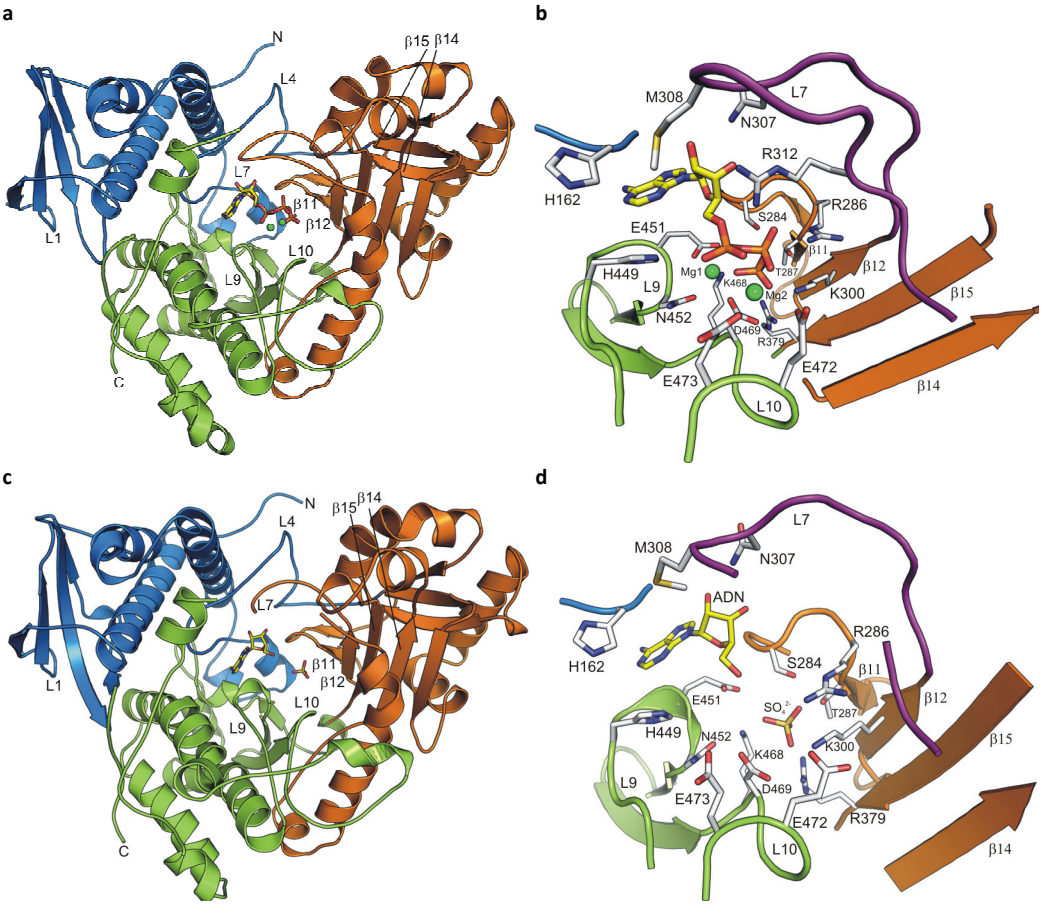


Figure 3.9: Coordination of AlcC ligands. **(a)** Soaked ATP complex. **(b)** ATP binding site. ATP carbons are colored in yellow. **(c)** Adenosine and SO_4^{2-} co-complex. **(d)** Adenosine and SO_4^{2-} coordination. L7 is less well coordinated and no electron density has been observed for residues 310-312.

Adenosine and sulfate of the ADN-SO_4^{2-} co-complex show a similar coordination (Figure 3.9c, d). Adenosine adopts position of the ATP adenine ring and the ribose while sulfate is coordinated in the γ phosphate position of ATP.

Upon ATP binding, loop L7 seems to be more ordered as in the apo structure or ADN-SO_4^{2-}co-complex which is lacking residues 310 – 312 (Figure 3.9b, d). This primarily affects the highly conserved R312. Upon ATP binding R312 moves in from a remote position to coordinate the α-phosphorous of ATP (Figure 3.10). Interestingly, ATP in AlcC is coordinated by two Mg ions rather than one as seen in AcsD (Figure 2.18a). The second Mg is coordinated by residues D469, E472 and E473 and bridges the α and β phosphates in AlcC. E451 adopts a different rotamer in both co-complex structures to allow adenosine or ATP binding. In the ATP co-complex E451 is involved in coordination of the first Mg^{2+} ion (Figure 3.10).

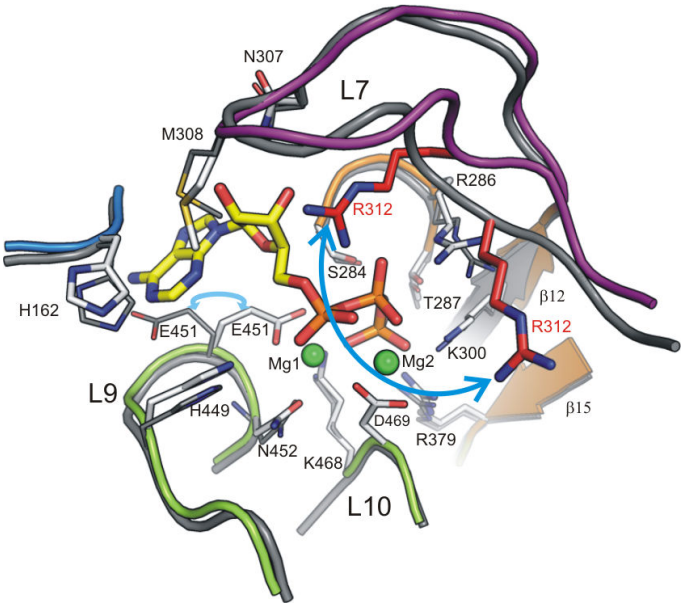
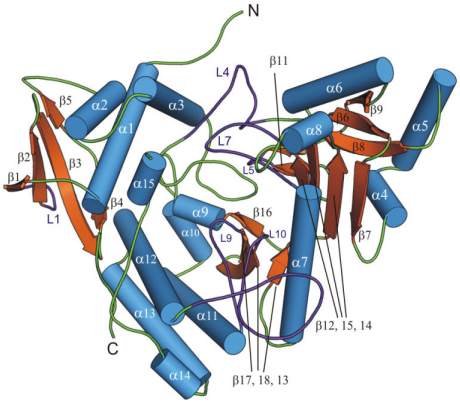


Figure 3.10: AlcC active site is more ordered upon ATP coordination. ATP binding helps to stabilized loop L7. R312 is moved in (indicated by a blue arrow) to bind the α phosphate of ATP. E451 adopts a different rotamer to allow adenosine or ATP to bind. AlcC apo structure is colored in grey, AlcC ATP structure as in Figure 3.9.

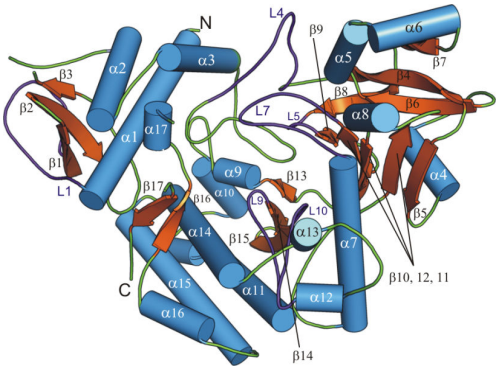
3.3.4 AlcC structure comparison with AcsD

Although AlcC has only 24 % sequence identity to AcsD (39 % similarity, for more detail see Appendix B, Table 4.5) it has a three domain topology that is similar to AcsD (Figure 3.1a). In contrast to AcsD in AlcC a 10 residue extended N-terminal loop is found (Figure 3.11a). But L1 in AlcC comprises only 7 residues compared to 28 residues in AcsD and connects the first two β-sheets, but is not involved in AlcC dimerization.

a



b



c

NIS family	sequence identity to AlcC	sequence similarity to AlcC
A	20-26 %	35-43 %
B	21-25 %	36-42 %
C	24-52 %	43-70 %

d

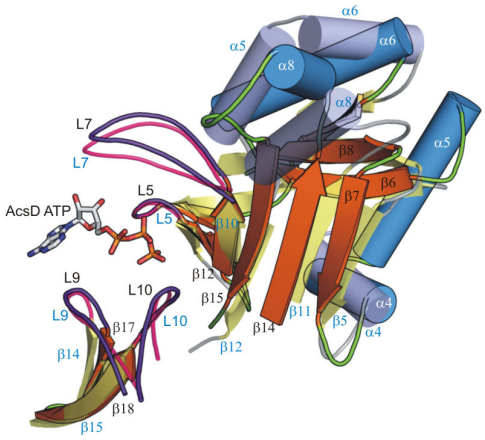


Figure 3.11: AlcC and AcsD have a similar topology. **(a)** AlcC monomer in cartoon representation. Striking differences to AcsD (b) are the extended N-terminal loop, the shorter L1 and the inserted 15 residue long helix α5. **(b)** Secondary structure elements of AcsD. **(c)** Sequence identity and similarity of NIS family members in relation to AlcC (detailed analysis in appendix B, Table 4.5). **(d)** AlcC active site shows structural similarities with AcsD, which is shown in lighter, transparent colors. AcsD secondary structure elements are labeled in black/white colors and AcsD elements in light blue colors.

The middle domain of AlcC contains 198 residues and is 15 residues longer than that one in AcsD. These additional residues form helix α5, which is flanked by α4 and β8. Helices α5 and α6 in AcsD are merged to helix α6 in AlcC. The C-terminal domain leads back to the N-terminal domain as seen for AcsD. The main difference in the C-terminal domain is found in L10. It extends a further 14 residues and replaces the two helices α12 and α13 found in AcsD (Figure 3.11a, b). Though AlcC has only ~40 % sequence similarity to type A NIS synthetases (Figure 3.11c) the ATP binding motif, the five stranded anti-parallel beta sheets and loops L5, L7, L9 and L10 (Figure 3.11d) are conserved.

3.3.5 Comparison of AlcC and AcsD ATP coordination

Superposition of AlcC and AcsD co-complex structures reveals that ATP in both structures is preliminary coordinated by matching residues (Figure 3.12). These are strongly conserved throughout NIS family A and C (Table 3.2) and are basically involved in the coordination of ATP tri-phosphates. The adenosine moiety of ATP in AlcC however adopts a slightly different position and is coordinated differently.

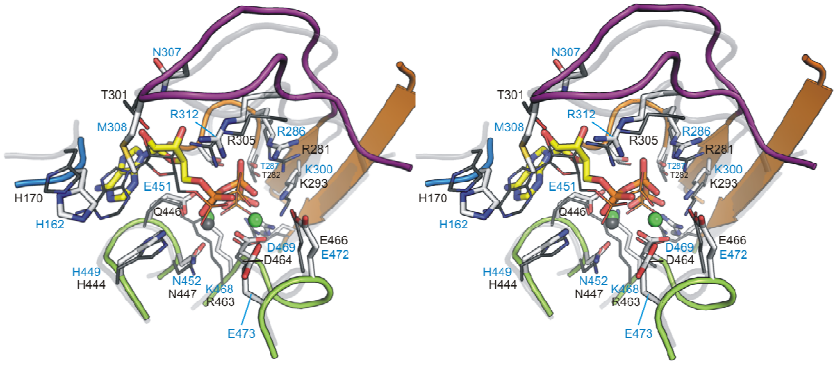


Figure 3.12: Comparison of ATP coordination in AlcC and AcsD. Cartoon representation in stereo view. ATP in both enzymes is preliminary coordinated by conserved residues. AlcC ATP co-complex is colored as in Figure 3.9. AcsD residues and Mg^{2+} are colored in dark grey and the backbone in transparent grey.

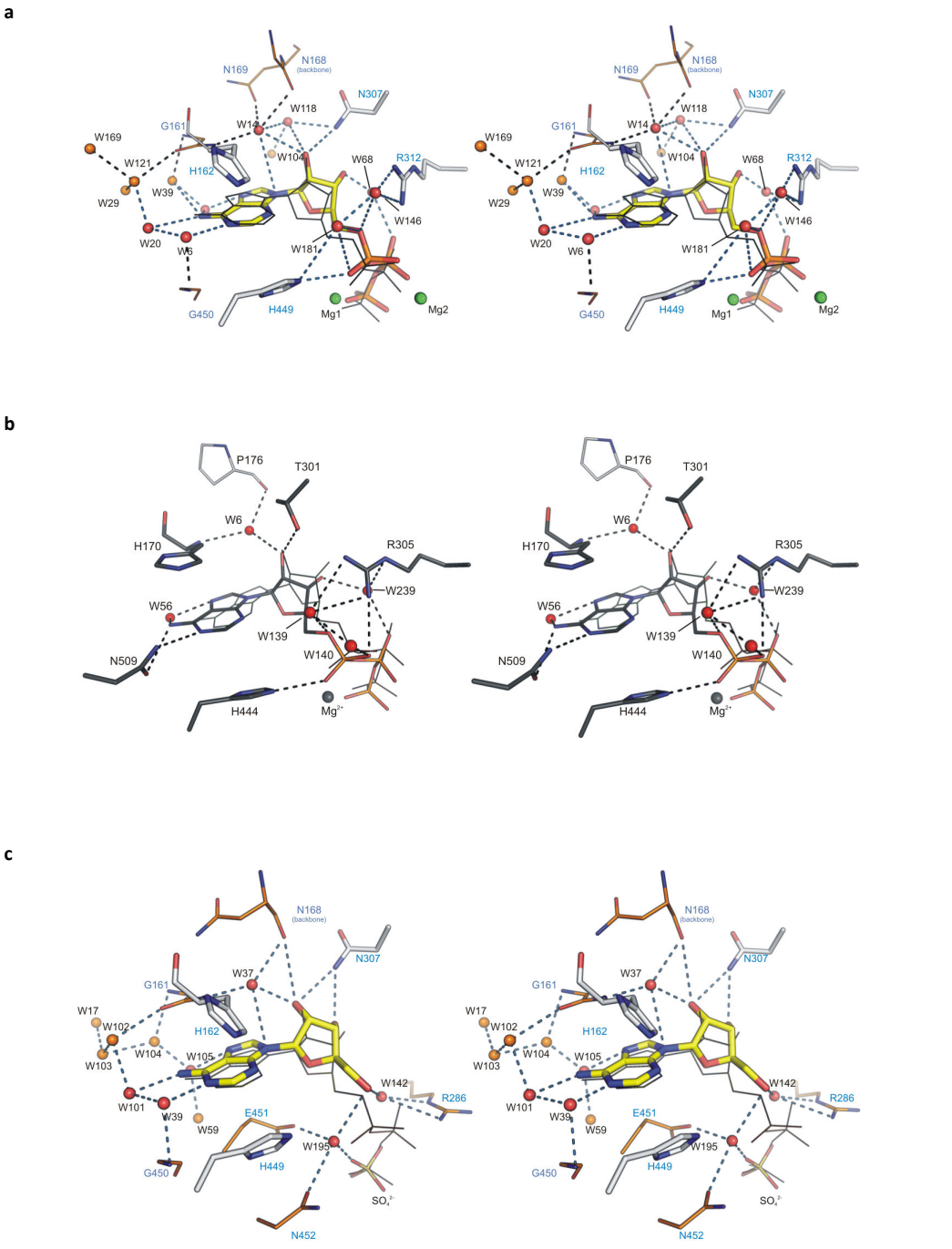


Figure 3.13: Coordination of ATP adenosine moiety in AlcC and AcsD. **(a)** The adenosine moiety in AlcC is coordinated by interaction with several water molecules. The different ATP position in AcsD (superposed based on active site residues) is shown in black wireframe. **(b)** The AcsD adenosine moiety is coordinated by less water molecules. Here the family A conserved N509 is involved in coordination of the adenine ring. AlcC ATP is superposed in black wireframe. **(c)** Adenosine in AlcC is coordinated in a similar water network as seen for ATP in (a). AlcC ATP is superposed based on active site residues and shown in black wireframe. Hydrogen bonds are indicated by dotted lines.

The most striking observation is that adenosine moiety in AlcC is essentially coordinated by a network of 13 water molecules (eight in direct contact with adenosine) (Figure 3.13a). In AcsD four adenosine binding waters are found (Figure 3.13b).

Table 3.2: ATP coordination in AlcC and AcsD. The table compares matching residues of AlcC and AcsD involved in ATP coordination. Each set of residues was defined as "conserved" when totally conserved in NIS family A, C. Number of family members carrying the same amino acid of partly conserved residues are shown in brackets (e.g (4/17)), whereas the last number is the total number of known family members.

AlcC	AcsD	conserved	conserved in family	partly conserved in family	not conserved
H162	**H170**	x	A,C	B(4/17)	
S284	**S279**	x	A,C	B(9/17)	
R286	**R281**	x	A,C	B(8/17)	
T287	**T282**		C	A (most T, but some S)	B
K300	**K293**	x	A,C	B (9/17)	
M308	**T301**		A		B,C
N307			C		A,B
R312	**R305**	x	A,C	B(9/17)	
R379	**R369**		C, most A (R/Y/V)	B (G/L/S/Y)	
H449	**H444**		A,C	B (6/17)	
E451	**Q446**		A, C (Q/E)	B (9/17)	
N452	**N447**	x	A,C	B (9/17)	
K468	**R463**		A (R conserved), C (K conserved)		B
D469	**D464**	x	A,C	B (D/E)	
E472	**E466**			A (E/Q), C (E/D/G)	B
E473	**G467**		A (G conserved),	C (E,D,G)	B
	N509		A		B,C

Since the AlcC ATP complex was derived from a soak, these water molecules could be an artifact. However the ADN co-complex derived from co-crystallization shows a similar water coordination network and demonstrates that this is not the case (Figure 3.13c). Another difference in ATP coordination is that in AcsD the adenine ring is stacked between two histidines (H444 and H170). H449 in AlcC adopts a similar position to H444 in AcsD, but H162 is found in a less ideal position for

stacking. N509 in AcsD (only conserved in family A members) coordinates the N1 and N6 atoms of the adenine moiety (Figure 3.13b). The matching G450 in AlcC is not close and binding of N1 and N6 is maintained by two water molecules (W20 and W6) instead (Figure 3.13a). Interestingly G450 is strongly conserved in type C enzymes (see also appendix D).

ATP in AlcC is polarized by two Mg ions. The first one is coordinated by AlcC residues E452, N452 and D469, the analogue residues of AcsD (Q446, N447, D464), bridging the α and γ phosphate of ATP. The second Mg in AlcC is positioned by residues D469, E473 and E473 binding to the α and β phosphorous (Figure 3.12). Only the first two residues (D469, E472) are found in AcsD (D464 and E466). The third residue E473, which is weakly conserved in NIS family C (other possibilities are D or G), is replaced by G467 in AcsD. Interestingly G467 is conserved within family A. Why AlcC coordinates a second Mg ion is still unclear. R312 and H449 in AlcC (conserved residues in NIS enzymes) face the α phosphate of ATP akin the catalytic important residues H444 and R305 in AcsD.

3.3.6 Activity assay for AlcC

Since AlcC is proposed to perform a similar reaction as AcsD, it presumably liberates AMP as well. Therefore its activity could be examined by the AMP activity assay, which was extensively tested for AcsD (see also chapter 2). Preliminary experiments with Sub204 as substrate analogue showed that this is the case and AlcC leads to NADH conversion in the coupled assay, while reactions without AlcC showed no decrease in NADH concentrations (Figure 3.14).

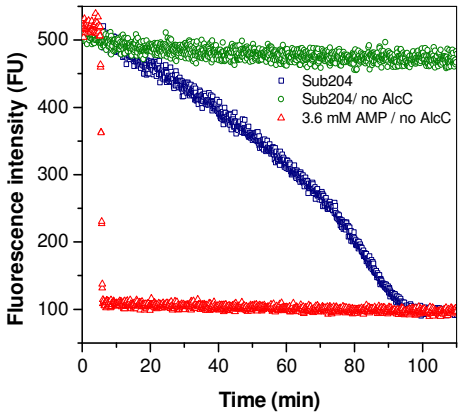


Figure 3.14: AMP production by AlcC. NADH consumption is observed using AlcC and substrate analogue Sub204. Reactions with Sub204 were AlcC was omitted showed no decrease in NADH levels. The positive control with AMP, but no AlcC enzyme, shows that the coupled enzymes are not rate limiting, but AlcC.

These first results indicate that the fluorescence based AMP activity assay is transferable to other AMP forming enzymes and could be used to obtain kinetic data for AlcC. However, due to the relatively large quantities of substrate analogue (Sub204) (synthesized by group of Prof. Challis, Warwick) required for this assay, only a few experiments were carried out to test the proof of principle. Tests with Sub220, the proposed substrate, were not possible. Unfortunately, Sub220 is not commercially available and synthesis optimization to yield adequate amounts is still in progress.

3.3.7 Mass spectroscopic analysis

3.3.7.1 AlcC forms a Sub204 dimer

The processive enzyme AlcC is proposed to perform dimerization and macrocyclization of Sub220. The substrate analogue Sub204 was tested to be active and to synthesize more easily in adequate amounts. As mentioned earlier, preliminary tests indicated that AlcC utilizes Sub204 catalyzing its dimerization, but not its macrocyclization. Mass spectroscopic analysis of AlcC product formation with Sub204 revealed two new mass peaks with an *m/z* value of 391.2 and 413.2 that increase over time (Figure 3.15). The theoretical mass of Sub204 dimer is 390.2, which correlates well with the observed *m/z*=391.2 peak in positive TOF mode ($M+H^+$). The second mass *m/z*=413.2 could be a dimer sodium adduct ($M+Na^+$).

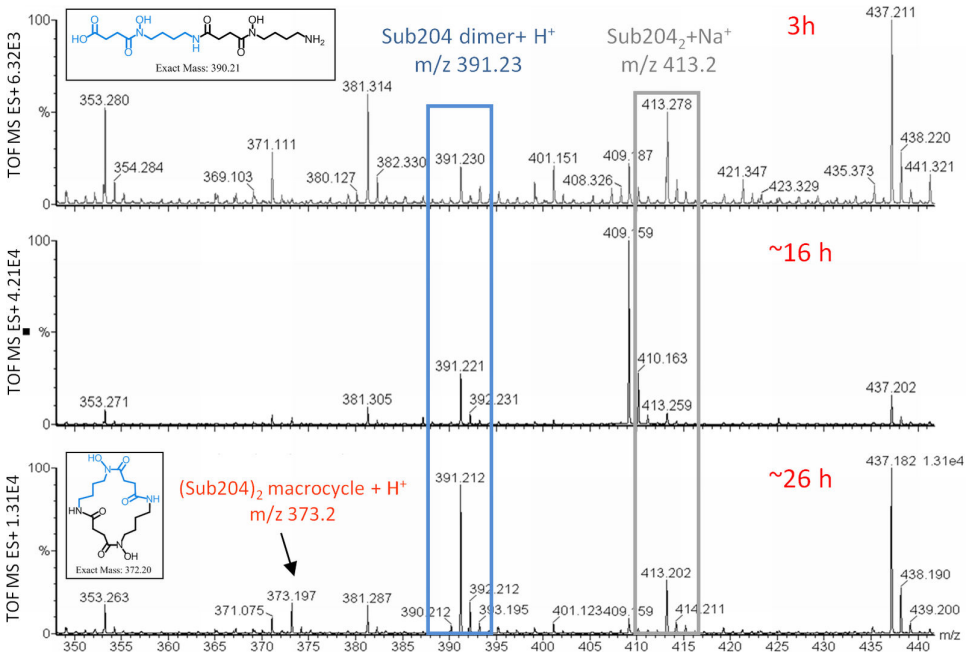


Figure 3.15: AlcC mass spectrometry analysis of Sub204 dimer formation. Two new peaks have been observed after 3 h incubation. The observed *m/z* values 391.2 and 413.2 correspond to Sub204 dimer (inset) and a Sub204 Na^+ adduct, respectively in positive TOF mode. A third mass peak at *m/z*=373.2 is observed after 26 h incubation that would match with the theoretical mass of a Sub204 dimer macrocycle ($M+H^+$) (inset).

3.3.7.2 AlcC forms a Sub204 dimer macrocycle

The mass spectrometry spectra of Sub204 product formation shows a third mass peak at *m/z*=373.2 after 26 h of reaction incubation (Figure 3.15). On closer examination, one can detect the 373.2 peak already after 3h of reaction initiation. The peak increases steadily over time if compared to the 371.11 peak (Figure 3.16). The 373.2 peak is 18 Daltons lighter than the Sub204 dimer peak (*m/z*=391.2) and hence could be the mass of the Sub204 dimer macrocycle, formed in a second AlcC reaction by losing water and forming an ester linkage. The observed mass ($M+H^+$ *m/z*=373.2) matches the theoretical mass of a $Sub204_2$ macrocycle (M=372.2) in positive TOF mode. The Sub204 derived macrocycle is also known as putrebactin in the literature (Kadi et al. 2008). This was surprising as it was thought that AlcC could only form Sub204 dimers.

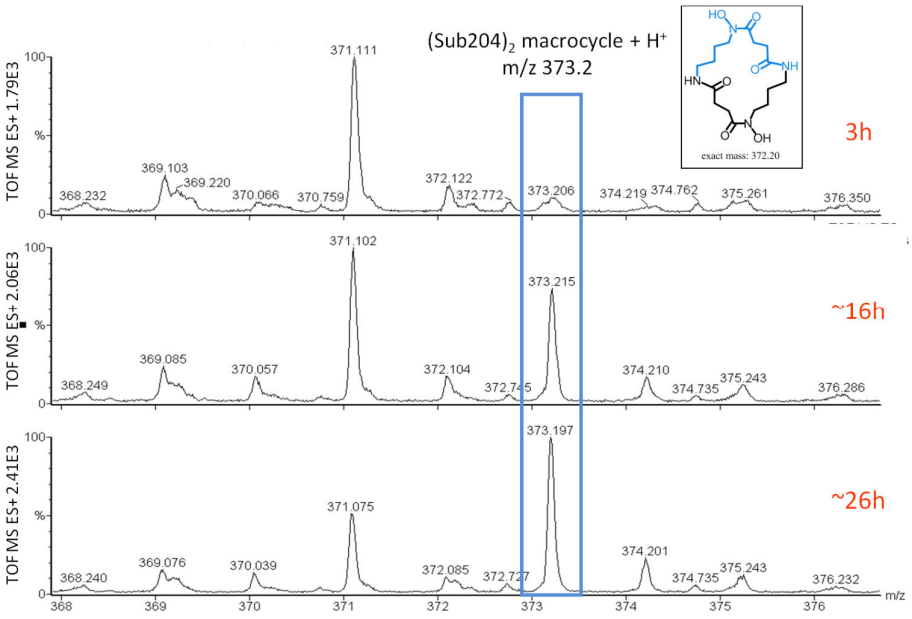


Figure 3.16: AlcC mass spectrometry analysis of Sub204 macrocycle formation. The 373 peak is detectable after 3 h of incubation and increases to dominate the spectra after 26 h in this *m/z* range. The *m/z* value of 373.19 ($M+H^+$) correlates well with the theoretical mass of a Sub204 dimer macrocycle (M=372.2) shown in the inset.

3.3.7.3 MS/MS analysis of Sub204 dimer and Sub204$_2$ macrocycle

Fragmentation of the three new peaks with an *m/z* value of 373.2, 391.2 and 413.2 in tandem mass spectrometry experiments led to several daughter ions for each species (Figure 3.17a, Figure 3.18a, Figure 3.19a). Interestingly daughter ions from Sub204 dimer (*m/z*= 291.2, 287.1, 269.1 and 187.1) and the dimer sodium adduct (*m/z*= 327.1, 313.2, 227.1 and 127.1) do not have the same masses. This was unexpected as one would think that the fragmentation of the same or related compounds would lead to at least some identical daughter ions, as seen for the fragments of Sub204 dimer and the Sub204 macrocycle which share the *m/z*=269.1 and 187.1 daughter ions (Figure 3.17a and Figure 3.19a).

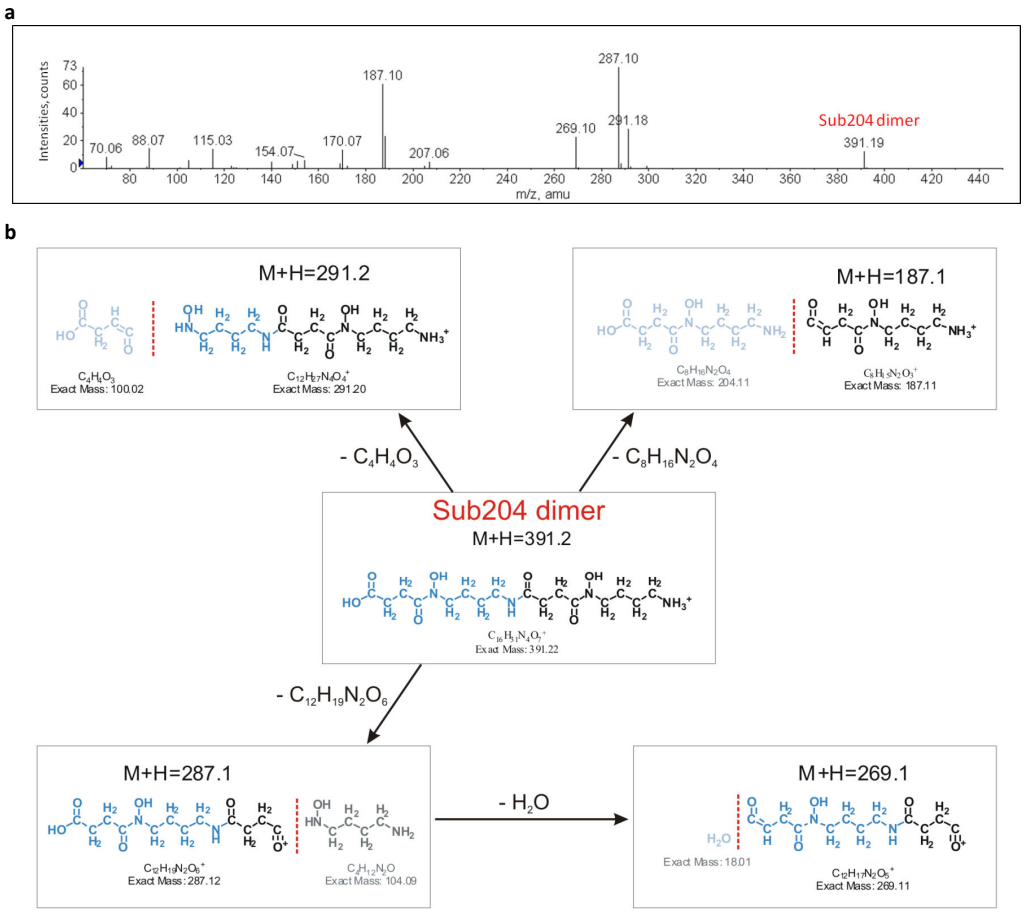


Figure 3.17: MS/MS fragmentation pattern of Sub204 dimer product. **(a)** Fragmentation of Sub204 dimer results in four main daughter ions. **(b)** Possible fragmentation of Sub204 generating the four daughter ions.

To verify that the Na^+ adduct fragmentation pattern is related and derives from Sub204, a theoretical fragmentation was performed. The software ChemDraw Ultra (CambridgeSoft) was used for drawing and calculation of fragmentation masses. In all three cases simple fragmentation of amide bonds led to the observed fragments (Figure 3.17b, Figure 3.18b and Figure 3.19b).

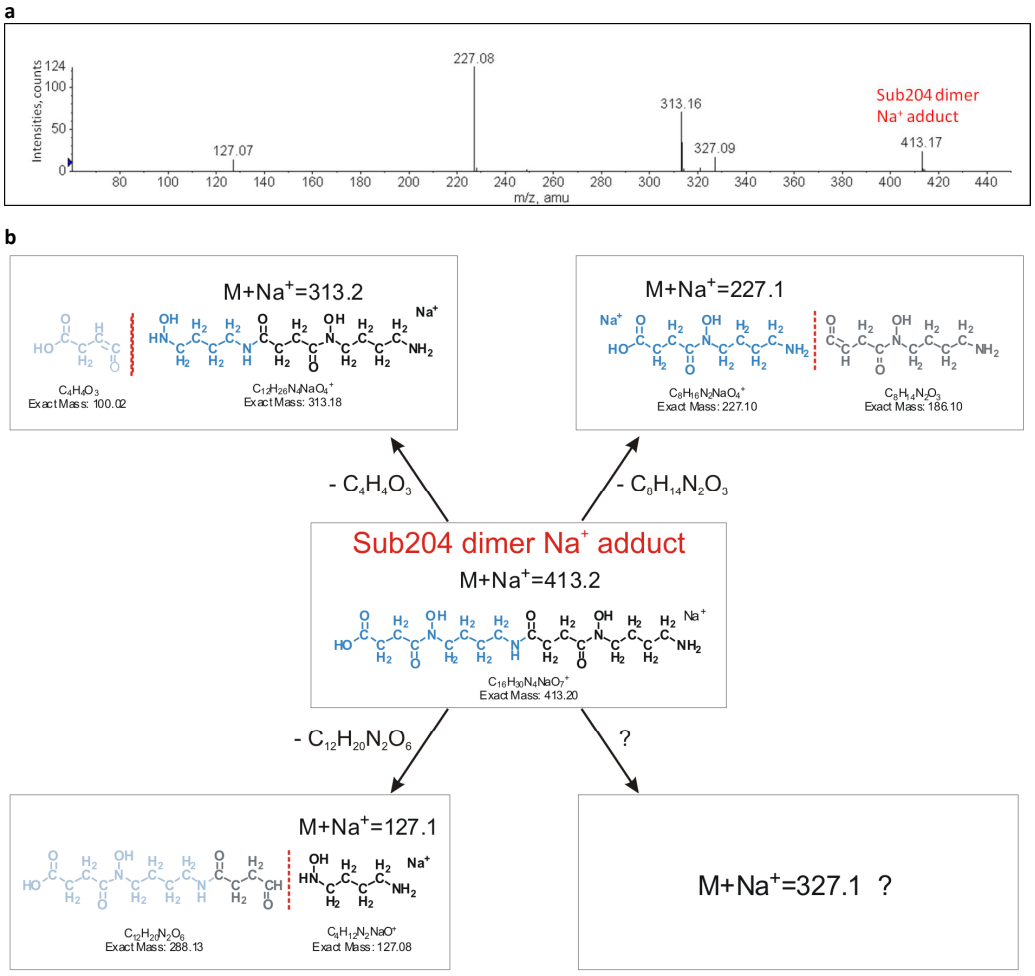


Figure 3.18: MS/MS fragmentation pattern of Sub204 dimer sodium adduct. **(a)** MS/MS spectra of dimeric Sub204 sodium adduct shows four daughter ions. **(b)** Possible fragmentation of Sub204 dimer adduct to generate the daughter ions in (a). No simple fragmentation of the sodium adduct yield a fragment with *m/z*=327.1.

For the Sub204 dimer, the three major fragments (*m/z*= 291.2, 287.1 and 187.1) can be generated by cleaving one of the three amide bonds. The fourth fragment (*m/z*=269.1) is possibly created in a second event from fragment *m/z*=287.1 by eliminating a water molecule (Figure 3.17b). Theoretical fragmentation of the sodium adduct results in three daughter ions (*m/z*=313.2, 227.1 and 127.1), but not in the *m/z*=327.1 fragment (Figure 3.18b). All daughter ions are also sodium adducts and therefore 23 Dalton heavier than the actual fragment. Hence, the *m/z* values of 313.2, 227.1 and 127.1 correlate to fragments with a mass of 290.2, 204.1 and 104.1 Daltons, respectively. Interestingly, these masses/fragments correlate with the non-charged fragments of the dimer in the theoretical fragmentation scheme and are not detectable (Figure 3.17b), apart from 290.2 which is observed as the protonated daughter ion 291.2 in the MS/MS spectra (Figure 3.17a).

Fragmentation of the Sub204 macrocycle results in two major daughter ions with *m/z* =269.1 and 187.1 and three minor ions with *m/z* =169.1, 151.1 and 140.06 (Figure 3.19a). An initial fragmentation is mandatory to break and linearize the macrocycle. This event presumably takes place by cleaving one of the four amide bonds. However the linearized macrocycle has no mass difference and is therefore not distinguishable from the closed ring in an MS/MS analysis (Figure 3.19b). In a second event, the linearized macrocycle is cleaved at another amide position resulting in fragments *m/z* =269.1 or 187.1 (Figure 3.19b). Two of the minor observed fragments (*m/z*=169.1 and 151.1) could be derived by two successive eliminations of water from fragment 187.1. Fragment *m/z*=140 is likely to be generated by losing *N*-methyl-hydroxylamine (OH-NH-CH_3) from fragment 187.1

(Figure 3.19b). Both major fragments (*m/z* =269.1 or 187.1) are also observed in the fragmentation pattern of dimeric Sub204 (Figure 3.17b). These results indicate that all three peaks (*m/z*= 373, 391 and 413) are related to each other and most likely originate from a Sub204 product generated by AlcC. This also shows that AlcC is capable of synthesizing the macrocycle putrebactin, which lacks the two additional OH-groups in alcaligin (Figure 3.2).

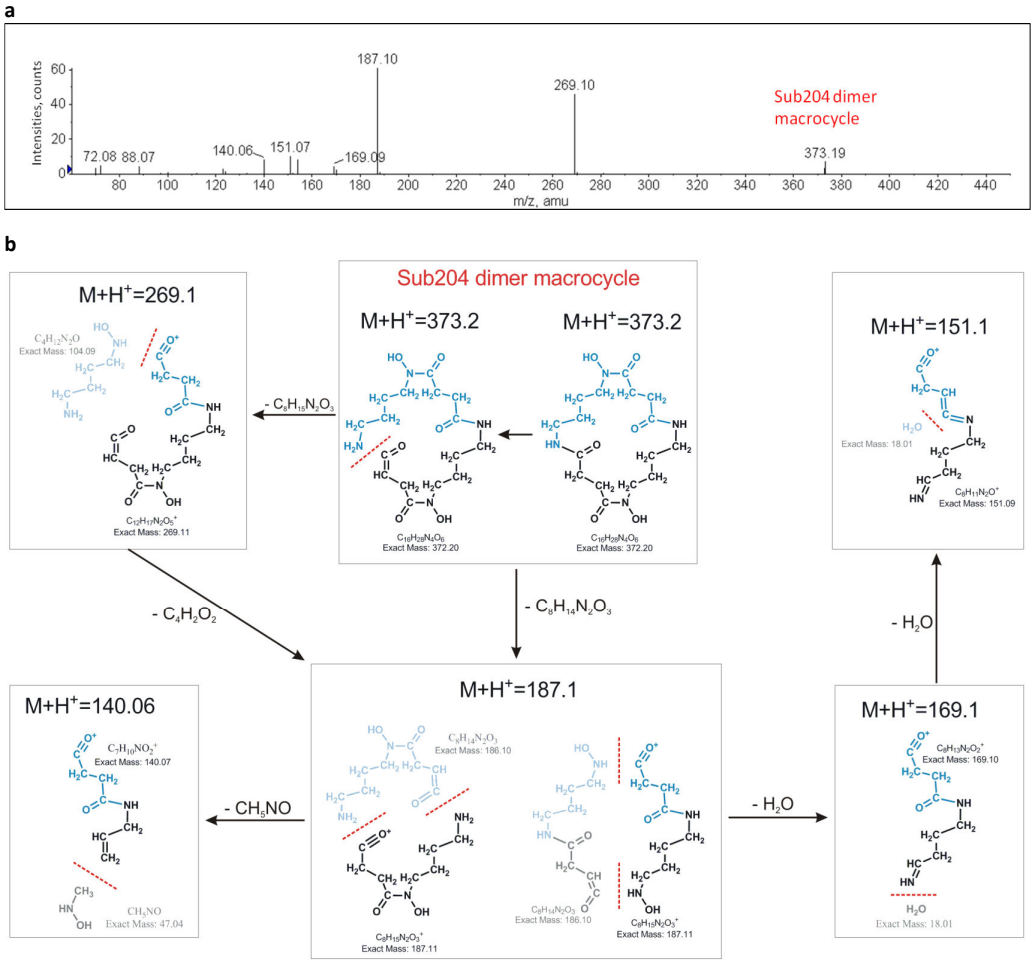


Figure 3.19: MS/MS fragmentation pattern of Sub204 dimer macrocycle. **(a)** Fragmentation results in six main daughter ions. **(b)** Possible fragmentation scheme of $Sub204_2$ macrocycle to generate the observed daughter ions.

3.3.7.4 AlcC forms a Sub204 trimer and trimer macrocycle

Long incubations of AlcC reaction (~3 days) showed that AlcC incubated with ATP, Mg^{2+} and Sub204 results in three additional mass ions (along the previous discussed dimeric peaks) with an *m/z* value of 559.3, 577.3 and 599.3 (Figure 3.20a). These masses correlate with the theoretical masses of a Sub204 trimeric macrocycle (558.3 Da), Sub204 trimer (576.3 Da) and Sub204 trimeric Na^+ adduct (599.3 Da) in positive acquisition mode (Figure 3.20b).

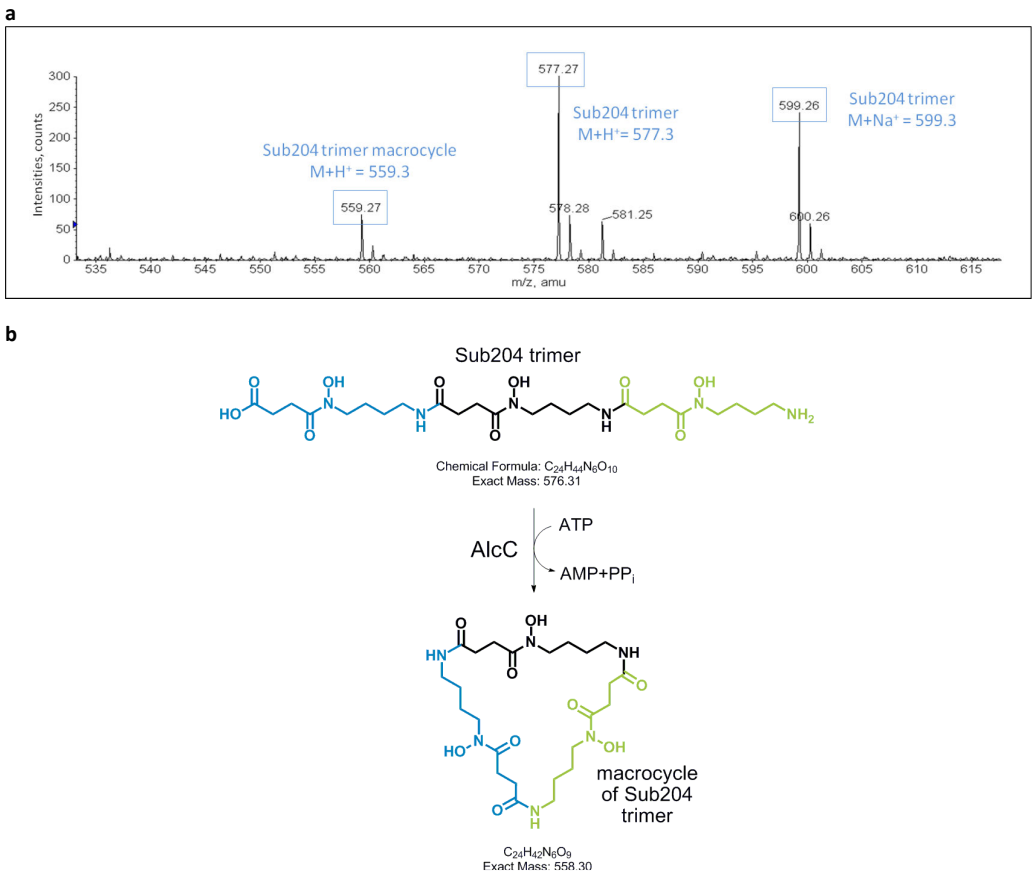


Figure 3.20: AlcC is capable of producing Sub204 trimers. **(a)** MS/MS analysis of AlcC reactions incubated for three days with Sub204 show three additional peaks that correlate with the masses of trimeric Sub204 in open (576.3 and 599.3 (Na^+ adduct)) and in ring formation (558.3). **(b)** Chemical structure of a Sub204 trimer and its conversion catalyzed by AlcC into a trimer macrocycle. Each Sub204 moiety is highlighted in a different color.

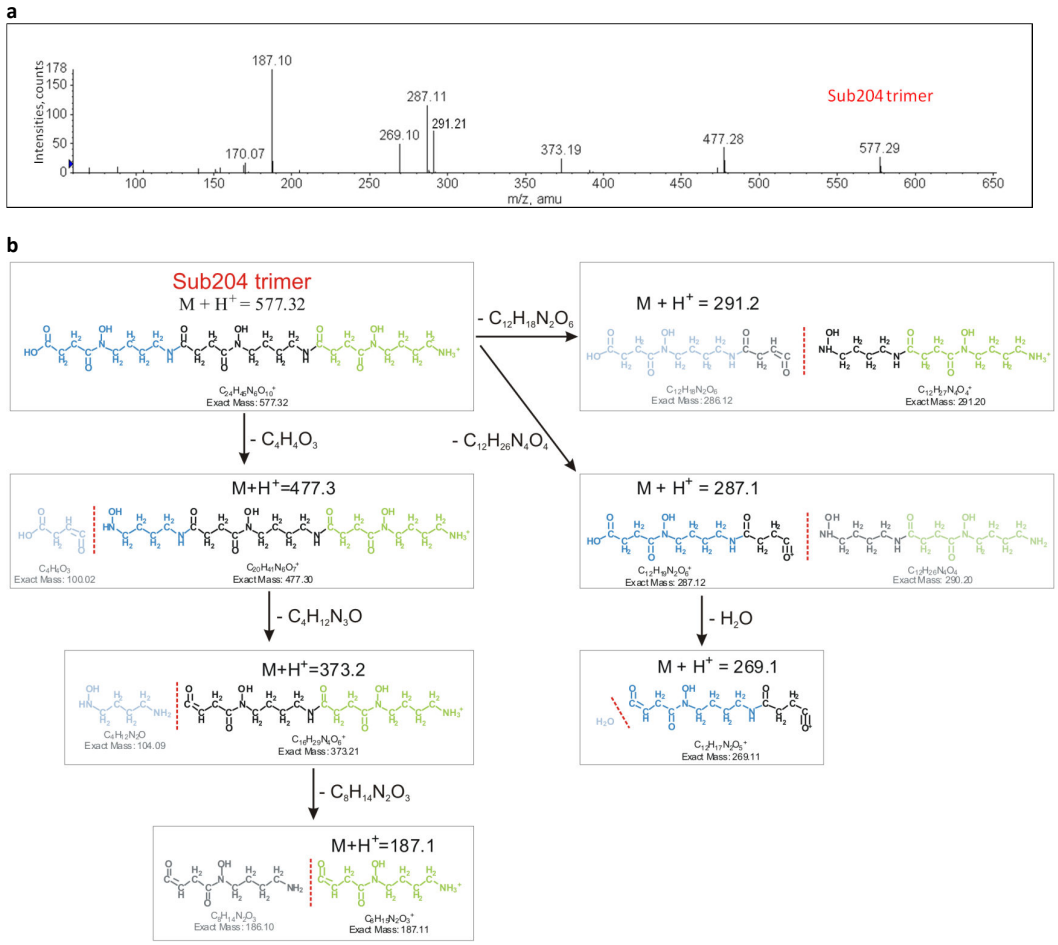


Figure 3.21: MS/MS analysis of Sub204 trimer peak. **(a)** MS/MS fragmentation pattern of Sub204 trimer peak 577.3. **(b)** Theoretical fragmentation to reproduce the daughter ions observed in (a). Each Sub204 moiety is highlighted in a separate color. The fragmentation sites are indicated by red dotted lines.

The detected MS/MS daughter ions of these three peaks indicate that they are related to the fragmentation patterns observed for dimeric Sub204 products (Figure 3.21a, Figure 3.23a, Figure 3.22a). All daughter ions of dimeric macrocycle, Sub204 dimer or the Na^+ adduct are found in the fragmentation pattern of the three peaks *m/z*=559.3, 577.3 and 599.3, respectively. Only fragment *m/z*=140 of the dimeric macrocycle has not been observed in the pattern of peaks 577.3 and 559.3 (Figure 3.19, Figure 3.21, Figure 3.22). As for the dimeric species, theoretical fragmentation

of each trimer species was performed to see if the resulting fragments correlate with the respective MS/MS spectrum.

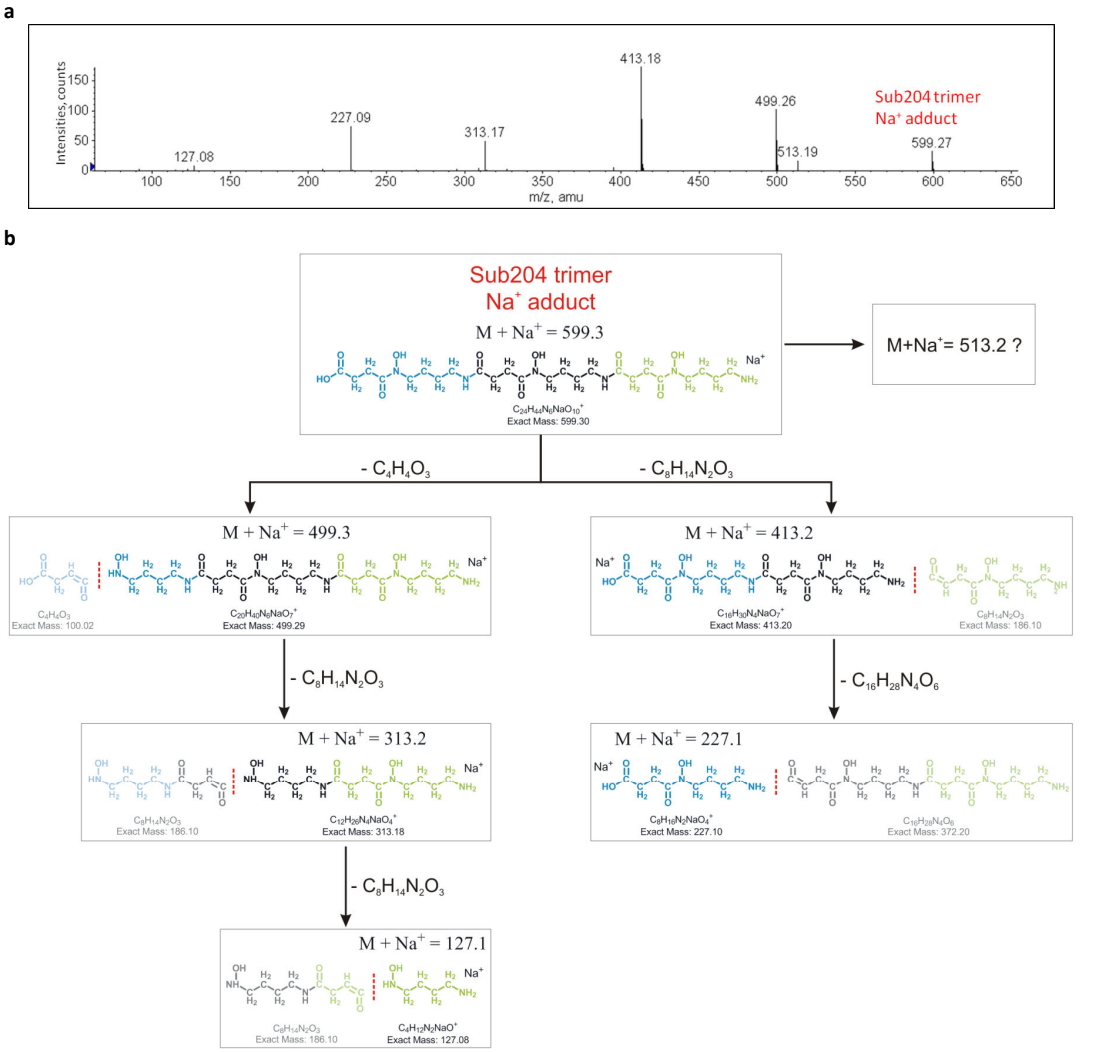


Figure 3.22: MS/MS analysis of Sub204 trimer sodium adduct peak. **(a)** MS/MS fragmentation pattern of Sub204 trimer peak 599.3. **(b)** Theoretical fragmentation of the trimer sodium adduct to reproduce the observed daughter ions in (a). Coloring as in Figure 3.21. No simple fragmentation of the trimer Na^+ adduct led to the observed fragment 513.2.

As for dimeric Sub204, the observed fragments could be reproduced in all three cases by simple fragmentation of amide bonds or by elimination of water molecules (Figure 3.21b, Figure 3.23b, Figure 3.22b). These results strongly suggest that AlcC is capable of extending a Sub204 dimer with another Sub204 molecule forming a

trimeric molecule. The data also suggest that AlcC catalyzes cyclization of this compound, liberating a trimeric macrocycle (Figure 3.17b).

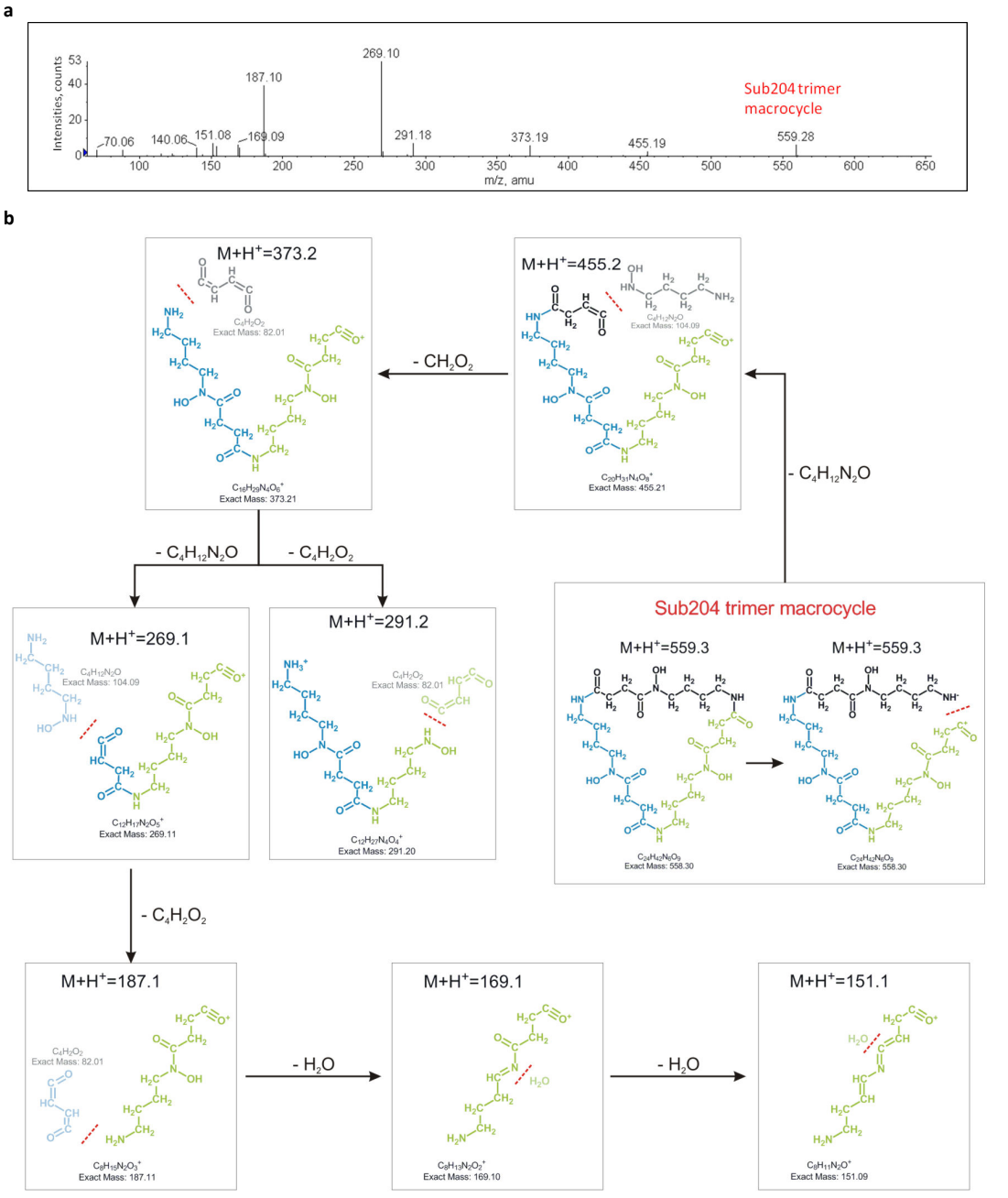


Figure 3.23: MS/MS analysis of Sub204 trimer macrocycle peak. **(a)** MS/MS fragmentation pattern of Sub204 trimer peak 559.3. **(b)** Theoretical fragmentation of the trimer macrocycle to reproduce the observed daughter ions in (a). Coloring as in Figure 3.21.

3.4 Discussion

3.4.1 AlcC dimerization interface

AlcC and AcsD form both dimers, but involve different dimerization domains. The AcsD interface is formed by L1, which comprises a four residue helix located in the N-terminal domain (Figure 2.1c, Figure 3.24a). The four residues involved in hydrogen bonds in the AcsD interface are not conserved in AlcC, and L1 in AlcC is much smaller and cannot establish the AcsD dimerization interface (Figure 3.24b). Instead residues mainly located on helices (α1-α3, α6 and α8), L4 and a longer N-terminal loop are involved in forming the AlcC dimer interface (Figure 3.25). Twelve hydrogen bonds and two salt bridges (K255-D123 in each monomer) are found. The interface buries 2130 $Å^2$ per monomer and is more than 2.5 times larger than that of AcsD, which buries 821 $Å^2$ per monomer as analysis with PISA shows (Krissinel and Henrick 2007).

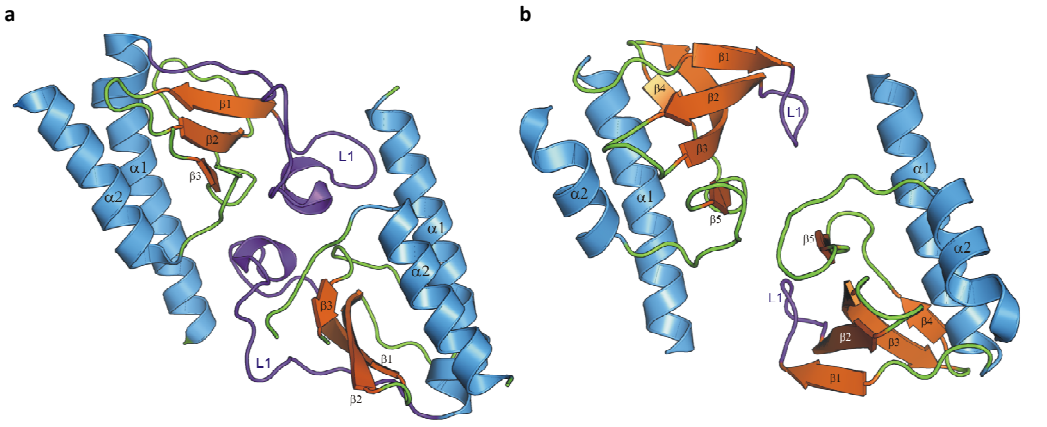


Figure 3.24: AcsD and AlcC dimer interface comparison. **(a)** Dimerization domain of AcsD **(b)** Model of an AcsD like dimerisation domain for AlcC. The model was generated by superposing two AlcC monomers independently with monomer A and B of AcsD. It demonstrates that AlcC cannot form a "side-by-side" dimer of AcsD due to its shorter L1 loop.

The AlcC dimer interface contains three family C conserved motifs (Figure 3.26a) comprising half of all hydrogen bond forming residues. They are located on α3, α6 and α8. D123 and K255 form two intermolecular salt bridges and are situated on α3 and α6, respectively. Both residues are weakly conserved and D123 is substituted by E or Q or in the case of K255 by several other amino acids (Figure 3.26a).

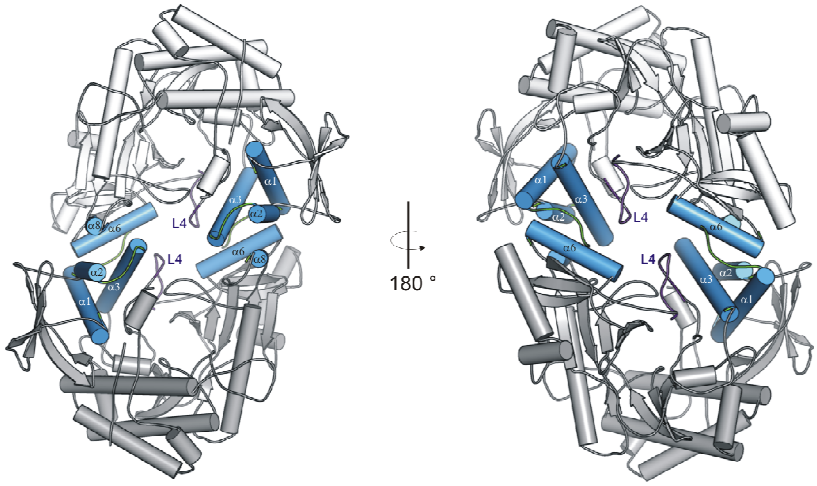
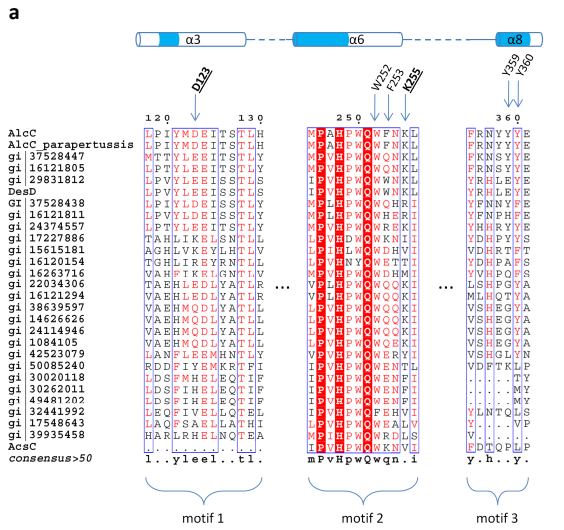


Figure 3.25: AlcC dimer interface. The "real" AlcC dimer is a head-to-head dimer. The interface is formed by residues located on helices (colored in blue) in the N-terminal and C-terminal domain. L4 and the long N-terminal loop are also involved in forming the interface.

PISA analysis (Krissinel and Henrick 2007) of AlcC dimerization shows that the interface area decreases by ~500 $Å^2$ per monomer if one deletes residues 7-17 of the N-terminal loop and that the dimer would be formed by six hydrogen bonds rather than by twelve (Figure 3.26b). As a result the free energy ($\Delta^i G$) released upon dimerization would be 10 kcal/mol less than for full length AlcC. This indicates that the additional N-terminal residues in AlcC (10 more than in AcsD) may help to interlock the two monomers and stabilize the dimer as demonstrated in Figure 3.26c. These dimerization motifs suggest that type C enzymes presumably form "head-to-head" rather than "site-by-site" dimers as observed in AcsD. Whether these

differences in dimer formation have specific function in catalysis or any other role is still unclear and needs further investigation.



AlcC PISA analysis	Apo structure	N-term. loop removed
complexation significance score (CSS)	1	1
interface area (Å^2)	2129.9	1606.3
hydrogen bonds	12	6
salt bridges	2	2
Δ^iG (kcal/mol)	-34.4	-24.5

Figure 3.26: Conserved motifs of the AlcC dimer interface. **(a)** AlcC contains three partly conserved motifs located on α3, α6 and α8 which comprise six hydrogen bond forming residues in the dimer interface. D123 and K255 form also two salt bridges. **(b)** PISA analysis of AlcC apo structure and of a model, lacking the N-terminal loop (residues 6-17). The interface area per monomer decreases by 500 Å^2 and less energy (~10 kcal/mol) upon dimerization is released. **(c)** The extended N-terminal loop (colored in red) interlocks the two AlcC monomers. AlcC is shown from its front (facing the active site) and its back in surface representation. N-terminal domain is colored in blue, middle domain in green and C-terminal domain in orange.

3.4.2 Conserved ATP binding motifs in NIS family C

Coordination of ATP in AlcC and AcsD involves several conserved residues. It was shown that most of these residues are close to the phosphates of ATP. However the adenosine moiety is bound differently. In AcsD the adenine ring is stacked between two conserved histidines (H444 and H170). H162 in AlcC is placed in a suboptimal position for stacking. Instead, a network of water molecules organizes the binding of the adenosine moiety as observed in the adenosine co-complex or the ATP complex structure (Figure 3.13). A sequence alignment of 27 NIS family C members shows that they have five conserved ATP binding motifs located on L5, L7, L9, L10 and on β12 and β15 (Figure 3.27).

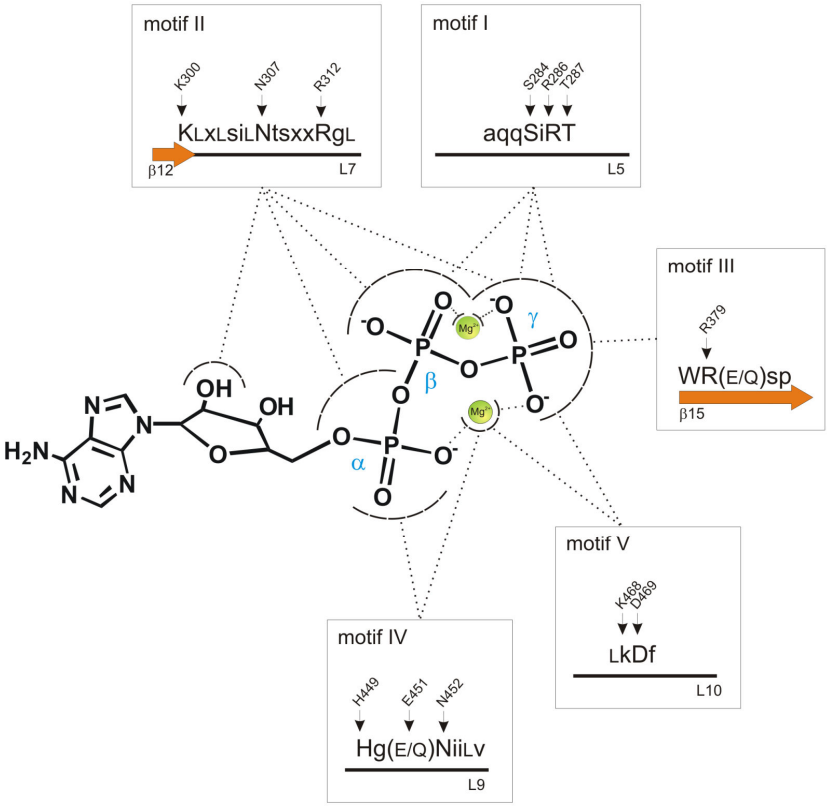


Figure 3.27: Conserved ATP binding motifs of NIS synthetase family C. Schematic representation of motifs and their binding association to ATP moieties. Labeled residues are conserved residues found in AlcC. Motifs and structural elements are numbered from the N-terminus. Upper cases are family conserved, lower cases are partly conserved and positions labeled with "x" are not conserved residues.

Remarkably, all five motifs are involved in phosphate coordination and contain the likely catalytic key residues R312 and H449. The most conserved motif is motif I (aqqSiRT) which contains three highly conserved residues (S284, R286 and T287 in AlcC) located on L5 that are involved in β and γ phosphate coordination (Figure 3.27). These three residues are also found in NIS family A members. Motif II is fifteen residues long and contains three highly, nine weakly and three non conserved residues on β12 and L7. This motif is involved in the binding of all ATP phosphates and the 2'-OH of the ribose and contains the presumed catalytic arginine (R312 in AlcC).

Motif III (WR(E/Q)sp) contains two highly conserved residues (W378 and R379 in AlcC), whereas only arginine is involved in the coordination of the γ phosphate. Motif IV contains three family conserved residues (H449, E451 and N452 in AlcC). These residues coordinate the α phosphate and the first Mg ion. A NIS superfamily conserved histidine (H449 in AlcC), which was shown to be one of the catalytic residues in AcsD (H444 in AcsD), is located in motif IV. Motif V is less well conserved and is involved in binding the γ phosphate and the first Mg ion. The binding site of the first Mg is conserved between AlcC and AcsD, in contrast to the second Mg coordination which is not conserved. Only a few other type C enzymes have matching residues (see also appendix D). This suggests that not all family C members bind a second Mg ion. These binding motifs indicate that the phosphates of ATP in family C but also in family A may be coordinated in a similar manner.

3.4.3 AlcC is a processive enzyme and forms macrocycles

Although AcsD and AlcC have a conserved ATP binding pocket (described in detail in the previous section), their substrate binding site must be different. AcsD, a type A enzyme, uses citric acid and L-serine, but AlcC, a type C enzyme, consumes and performs dimerization of the succinic acid derivatives Sub204 or Sub220. For AcsD, the co-complex of citric acid and the final product citryl-EDA assisted in elucidation of the AcsD reaction mechanism. Unfortunately no such co-complex structures were obtained for AlcC. However, one still can infer substrate binding from the AcsD co-complex structures for a possible ligand location in AlcC. The O1 atom of the citric acid is facing the α-phosphorous in an angle of 151° which is a consistent position for a nucleophile attack to form an enzyme bound citryl-adenylate. Furthermore, two conserved catalytic residues (R305 and H444) are in close proximity to the carboxylate of citric acid. Bearing all these facts in mind the first ternary complex with Sub220 can be modeled in the AlcC ATP co-complex. Sub220 perfectly fits in a small active site pocket and would be coordinated by R312, H162, M447, H449 and D506 (Figure 3.28a). The Sub220 carboxylate and not its amino group would face the α-phosphorous of ATP allowing a nucleophilic attack of the carboxylate. A conserved histidine (H162 in AlcC, H170 in AcsD) is close to coordinate Sub220. In AcsD, H170 is involved in binding of citric acid as well as stacking the adenosine ring of ATP with H444. However in AlcC, H162 is in a less favored position for stacking. The remaining Sub220 binding residues M447 and D506 are only weakly conserved (found in nine other type C enzymes, see also appendix D, Figure 4.5).

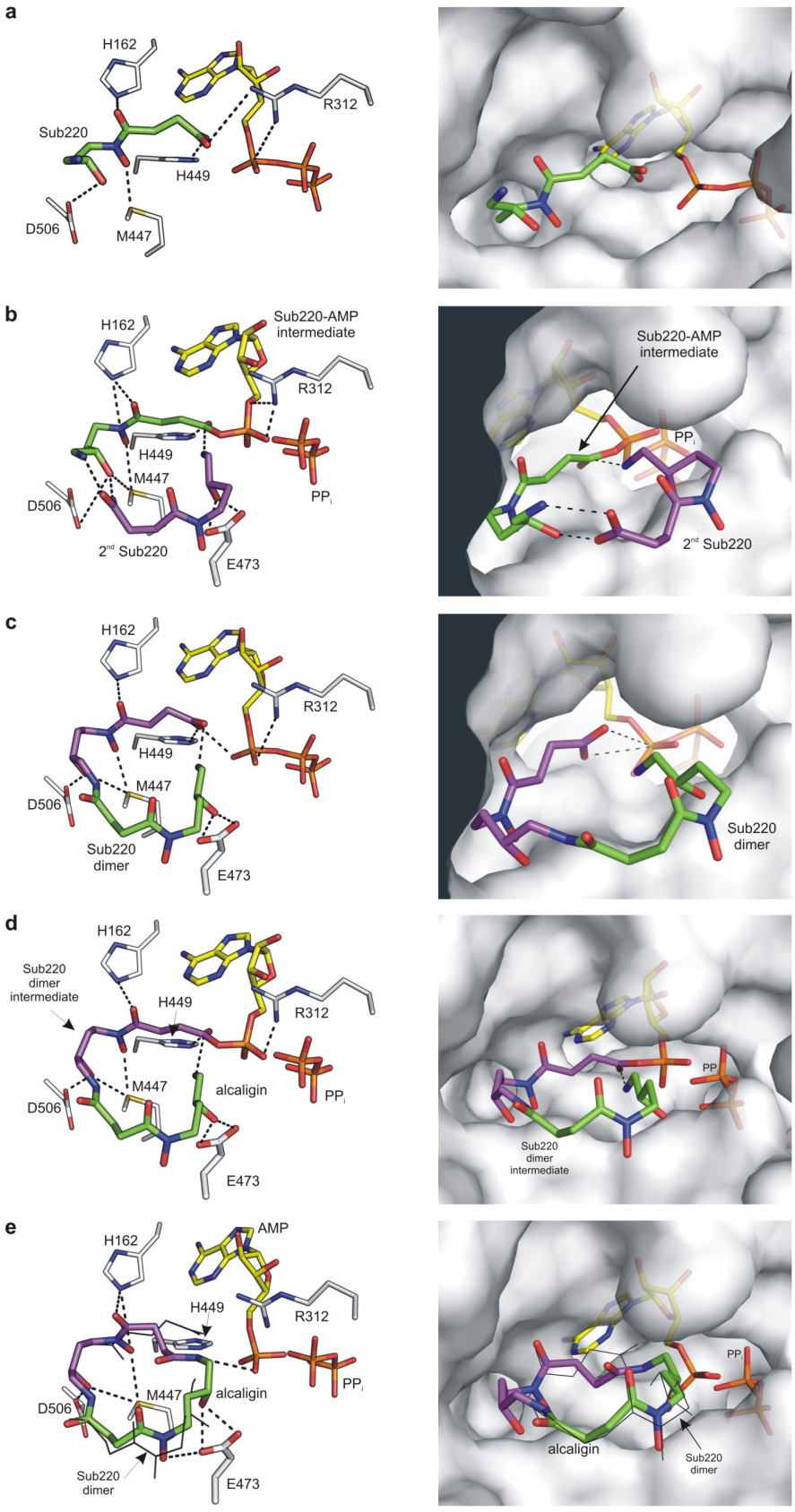

a
H162
R312
Sub220
H449
D506
M447
b
H162
Sub220-AMP
intermediate
R312
H449
M447
PPi
D506
2nd Sub220
E473
Sub220-AMP
intermediate
PPi
2nd Sub220
c
H162
R312
H449
M447
D506
Sub220
dimer
E473
Sub220
dimer
d
H162
Sub220
dimer
intermediate
H449
R312
M447
alcaligin
PPi
D506
E473
PPi
Sub220
dimer
intermediate
e
H162
AMP
H449
R312
PPi
M447
alcaligin
D506
Sub220
dimer
E473
PPi
alcaligin
Sub220
dimer

Figure 3.28: Docked models of Sub220 and final products in AlcC active site. **(a)** The docked Sub220 (colored in green) faces with its carboxylate the α phosphate of ATP. Sub220 is coordinated by three conserved residues (H162, R312 and H449) and two weakly conserved residues (M447 and D506). **(b)** A second Sub220 molecule (colored in purple) faces the Sub220-AMP intermediate (Sub220 moiety colored in green, AMP in yellow) from the solvent accessible face and binds to E473. **(c)** The Sub220 dimer has to bind in a permutated orientation to allow macrocyclization. **(d)** The Sub220 dimer is activated by adenylation and a Sub220 dimer intermediate and PP_i is generated. **(e)** A docked alcaligin molecule superposes well with a docked Sub220 dimer (shown in black wireframe).

Similar to AcsD where the L-serine attack must occur on the solvent exposed face of citric acid, in AlcC only one position for the second substrate molecule is conceivable. In the second ternary complex another Sub220 molecule presumably is coordinated by E473 facing the ester bond of the Sub220-adenylate with its nucleophilic amino group (Figure 3.28b). The Sub220 carboxylate binds to the free amino group of the Sub220-adenylate intermediate. After dimer formation the product has to be released. This would be consistent with mass spectrometry results where a mass peak corresponding to Sub204 dimer is detected although the enzyme was removed by filtration prior analysis. The $Sub204_2$ macrocyclization can happen in two different ways: enzyme catalyzed or autocatalytically. Reactions with Sub204, where AlcC was removed after O/N incubation (to yield sufficient Sub204 dimer product), showed that the ratio of Sub204 dimer (+ Na^+ adduct) to dimer macrocycle peak levels do not change significantly over a period of 13 days (see appendix E, Figure 4.8). This indicates that AlcC catalyzes the macrocyclization and that the macrocycle is not formed autocatalytically. Therefore a generated dimer would have to bind in a permutated orientation in the active site facing the α-phosphorous of ATP with its free carboxylate (Figure 3.28c). As for the first reaction the dimer has to be activated by adenylation generating an enzyme bound Sub220 dimer intermediate and pyrophosphate (Figure 3.28d). A model of the resulting alcaligin superimposes well

with both substrate position and the macrocycle does not clash with active site residues (Figure 3.28e).

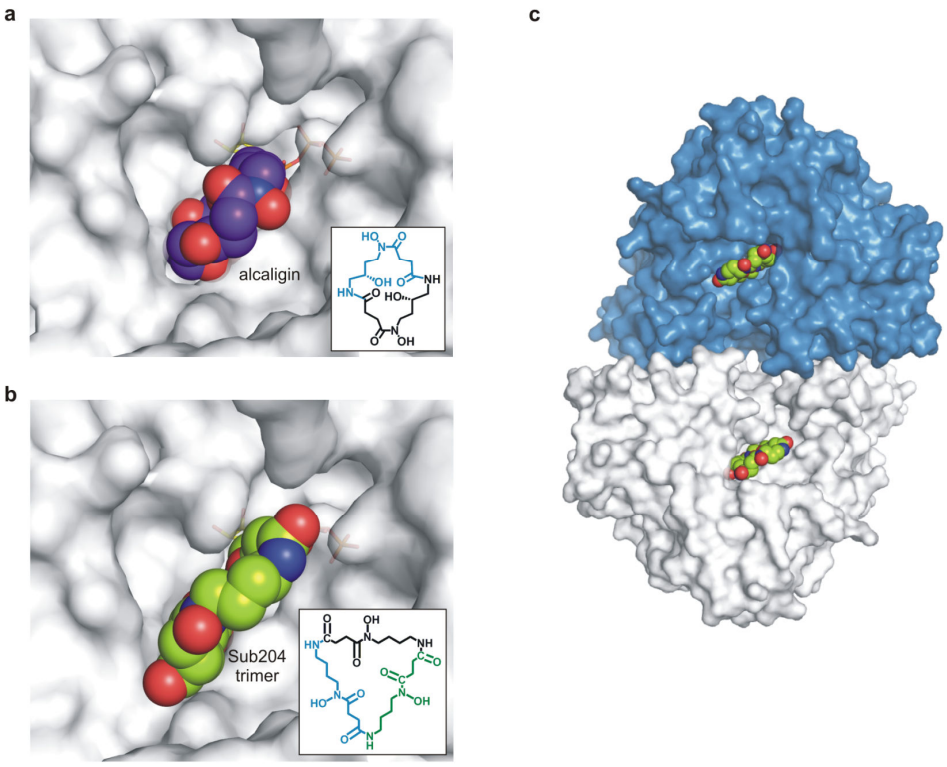


Figure 3.29: Docked models of alcaligin and the Sub204 trimer in AlcC active site. **(a)** Space filling model of alcaligin. **(b)** Space filling model of Sub204 trimer demonstrates that there is still enough room to bind the large trimeric macrocycle. **(c)** Docked model of two Sub204 trimer macrocycles in both AlcC monomers.

AlcC was shown to form a Sub204 trimer in an open and a closed ring formation. This was quite surprising since the actual biosynthesis model proposes AlcC to form alcaligin, a dimeric macrocycle (Challis 2005; Brickman et al. 2007). A docked model of alcaligin and Sub204 trimer macrocycle shows that the larger macrocycle is still able to fit in the AlcC active site. This is demonstrated by the space filled models of each macrocycle (Figure 3.29a, b). A model of AlcC dimer with two trimeric macrocycles bound reveals that each active site has to work independently since

they are too far apart to interact with each other. These models suggest that both reactions (dimerization and macrocyclization) are likely to happen in the same place.

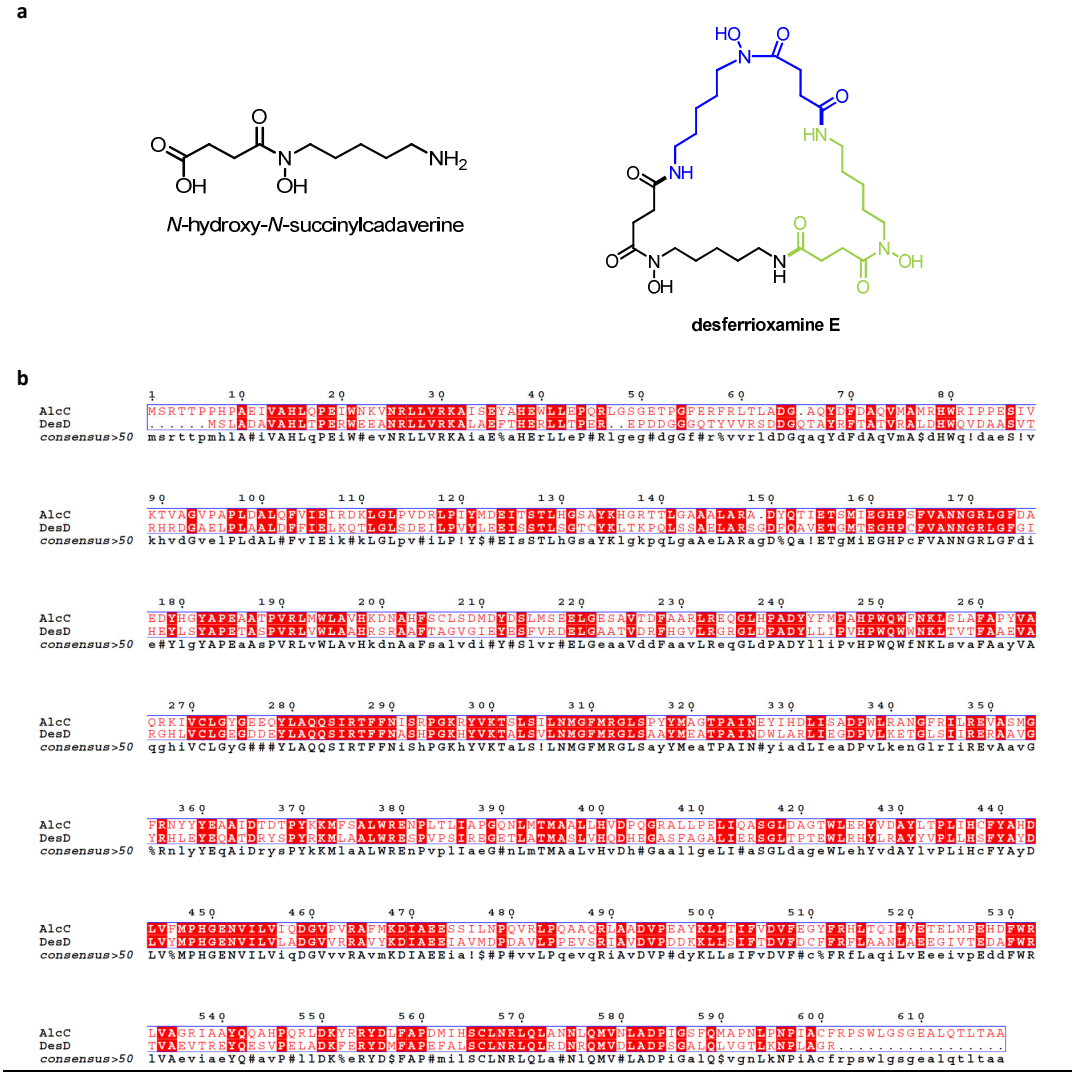


Figure 3.30: Similarities between AlcC and DesD. **(a)** Chemical structure of *N*-hydroxy-*N*-succinylcadaverine and desferrioxamine E, substrate and product of DesD. **(b)** A sequence analysis of AlcC and DesD shows that key active site residues (H162, R312, M447, H449, E472, E473, D506 etc.) are conserved in both enzymes.

DesD, another type C enzyme, is involved in biosynthesis of desferrioxamine in *Streptomyces coelicolo*r (Kadi et al. 2007). It has 51 % sequence identity (69 % sequence similarity) with AlcC (see also appendix C, Table 4.5). DesD is proposed to perform trimerization and macrocyclization of *N*-hydroxy-*N*-succinylcadaverine to

generate desferrioxamine E (Figure 3.30a). The DesD substrate is a 10-aminocarboxylic compound and differs only by an additional carbon compared to Sub204 (Figure 3.2b), which is a 9-aminocarboxylic substrate. A sequence alignment of both enzymes shows that residues involved in ATP coordination as well as the possible catalytic residues R312 and H449 are found in both enzymes (Figure 3.30b). Residues H162, M447, D506 and E473 which are likely to coordinate Sub220 in AlcC are also present in DesD. Hence the active site must be organized in a similar way. Therefore it is likely that AlcC is able to utilize DesD substrate catalyzing not only dimerization, but also trimerization and macrocyclization.

3.5 Conclusion

The apo and co-complex structure of AlcC are the first structural insights into a NIS family C member. While the adenosine structure was acquired by co-crystallization, the ATP structure was only obtained after replacing the active site bound sulfate. This was accomplished by soaking apo crystals in high concentrations of sodium formate, followed by an ATP/sodium formate solution. The most prominent observation is that ATP is coordinated in a "similar" manner to ATP in AcsD, and that two conserved residues, R312 (that moves in from a remote position) and H449, are found in close proximity to the α-phosphorous. This indicates that these residues, as it was shown for AcsD, are the catalytic residues. Sequence analysis of family C members showed that coordination of the tri-phosphates and the first Mg ion is partly conserved in five binding motifs. The binding site of the second Mg is not conserved and is presumably not found in many other type C enzymes.

AlcC and AcsD share a three domain topology. There are only minor differences in secondary elements such as an added helix (α5) or the extended N-terminal loop in AlcC. Even though they share a similar topology, AlcC has a "head-to-head" dimer and is not able to form the "side-by-side" dimer seen for AcsD due to its shorter L1 loop. The extended N-terminal loop (missing in AcsD) was shown to interlock the two AlcC monomers and stabilize them. It still remains unclear whether dimerization is functionally significant.

In biochemical analysis with the substrate analogue *N*-hydroxy-*N*-succinylputrescine (Sub204) it was demonstrated that AlcC is capable of forming dimers, trimers and

their macrocycles. This was unexpected, since preliminary tests in the group of Prof. Challis indicated that with this substrate analogue only a dimer but not a macrocycle was formed. A detailed fragmentation analysis of the MS/MS data illustrated that simple cleavage at amide bonds or elimination of water reproduces the observed fragments. Since no substrate or product co-complex was obtained substrate binding models were created. The docked model of Sub220 substrate perfectly fits in a small cavity facing the α phosphate of ATP with its carboxylate similar to citric acid in AcsD allowing a S_N2 like attack to generate a Sub220 adenylate. The model of alcaligin and the trimer macrocycle demonstrate that the active site of AlcC is large enough to accommodate both without any domain movements. This would be consistent with the AcsD structure where no structural movement upon nucleophile binding was observed. A model of AlcC dimer with two coordinated $(Sub204)_3$ macrocycles demonstrates that the two active sites of AlcC are too far apart to interact with each other and are not likely to form one “large” active site. Interestingly DesD, another type C enzyme, performs a similar trimerization and macrocyclization and shows a high sequence similarity to AlcC. Key active site residues including the likely substrate coordinating residues are found in both enzymes. DesD is involved in desferrioxamine biosynthesis, a substance also used as a FDA approved drug in medical applications such as iron chelating therapies after repetitive blood transfusions or treatment of chronic iron overload or acute iron intoxications (Hua et al. 2008; Suzuki 2008). But so far no structural data of DesD are published and the acquired structural results on AlcC might help in developing new or modified iron chelating compounds used in medical or commercial applications.

3.6 Future work

AlcC and AcsD have a similar three domain topology, but don't share the same dimer interface. The dimer interface in AlcC seems at least partly conserved in family C. Site directed mutagenesis of hydrogen bond forming residues involved in stabilizing the dimer and their effect on enzyme activity is required to probe their role. Additional biochemical data such as kinetic parameters or mutational studies on H449 and R312 are desirable.

Structural models of AlcC substrate and product coordination were developed based on structural information of AlcC and AcsD co-complex structures. Hence future work will concentrate on obtaining substrate or product co-complexes. Since soaking has not been successful, a different method or a revised protocol is needed. One possible approach is mutation of likely coordinating residues to enhance substrate binding and/or test substrate analogues that act as active inhibitors. However the main problem seems to be the sulfate containing crystallization condition, which does not allow binding of ATP until sulfate is removed or diluted. So far intensive screening for a sulfate-free condition has failed. Surface reduction entropy (SER) approach has been shown to enhance crystallization of several proteins (Derewenda 2004; Goldschmidt et al. 2007). The surface modified AlcC protein could be re-screened with commercial or customized stochastic screens, where sulfate or phosphate is omitted.

AlcC was shown to be capable of catalyzing trimerization of *N*-hydroxy-*N*-succinylputrescine (Sub204). DesD also catalyzes a trimerization reaction. It utilizes

three molecules of *N*-hydroxy-*N*-succinylcadaverine to generate desferrioxamine E. Since both substrates are quite similar (just a carbon longer), AlcC might also be able to utilize the DesD substrate. The docked model of the Sub204 trimer shows that there is enough space for larger substrates and products. Hence, finding and testing of other natural or chemically synthesized substrates could increase its use in biotransformation.

Chapter 4 - Adenylation

4.1 Introduction

4.1.1 Summary

Thioesters, amides and esters are common chemical building blocks in a wide array of natural products. The formation of these bonds can be catalyzed in a variety of ways. For chemists, the use of an activating group is a common strategy and adenylate-forming enzymes are examples of this approach in biology. Here the hydroxyl leaving group of the otherwise unreactive carboxylic acid is activated by adenosine monophosphate liberating pyrophosphate (the driving force). The resulting carboxylate adenylate is very reactive and would decompose in water. Necessarily the enzymes that forms this reactive intermediate also catalyze a second step, in which a nucleophile (either amine, alcohol or thiol) reacts with the intermediate to generate the desired product, liberating AMP (Figure 1.2a).

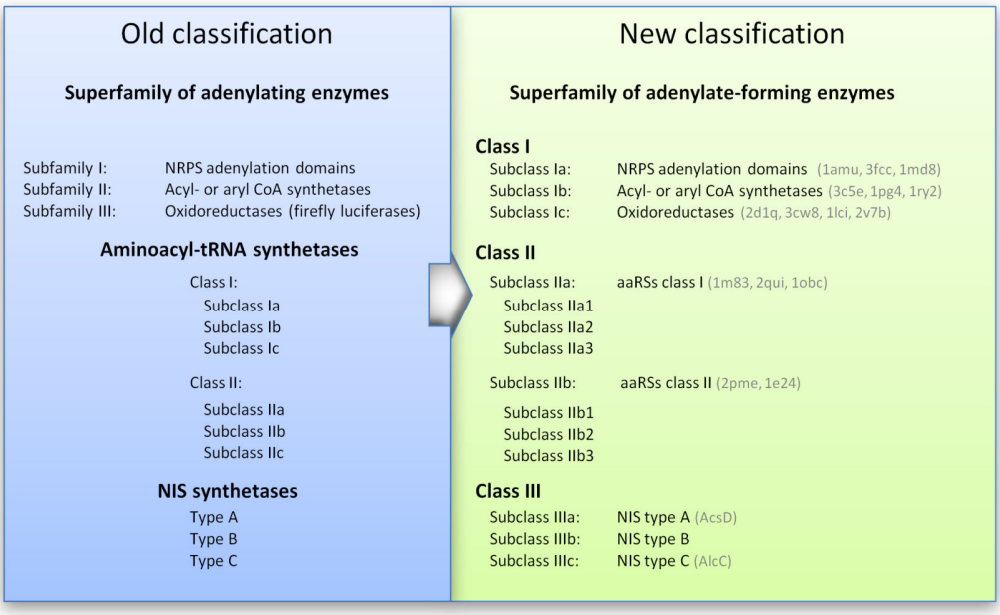


Figure 4.1: Old and new classification system of adenylate forming enzymes.

In this chapter the structural relation of AcsD and AlcC to other adenylate-forming enzymes is investigated. Coordination of ATP and metal ions were compared and the different catalytic residues examined. Since the existing superfamily (Fulda et al. 1994) did not include tRNA-synthetases or the new identified family of NIS synthetases (comprising AscD and AlcC), a new classification system for adenylate forming enzymes was desirable. A new superfamily is proposed comprising three basic classes (Figure 4.1). The first class constitutes the former superfamily with subclasses Ia, Ib and Ic. The second class comprises aminoacyl-tRNA synthetases with subclasses IIa and IIb, while class III contains the NIS enzymes with subclasses IIIa, IIIb and IIIc.

4.2 Results and discussion

4.2.1 Same reaction but very different domain architectures

All adenylate-forming enzymes catalyze the difficult condensation between a nucleophilic carboxylic acid and the electrophilic phosphate of ATP. This mechanism proceeds through a negatively charged phosphorus group which must be stabilized. Two different chemical mechanisms can perform such a reaction: a dissociative or an associative substitution.

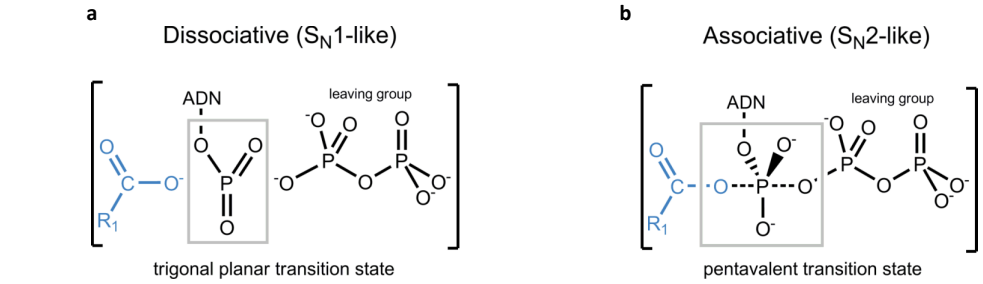


Figure 4.2: Comparison of the transition state of a dissociative and associative reaction.

In the first step of a dissociative mechanism the leaving group is released, forming a trigonal planar transition state (Figure 4.2a). In the second step the carboxyl-nucleophile attacks the phosphorous. The other alternative is an associative substitution. Here the nucleophilic carboxylate attacks the α-phosphorous on the opposite site (180°) of the pyrophosphate leaving group in a one step reaction. This mechanism proceeds through a trigonal-bipyramidal transition state and covalently binds the carboxyl and the leaving group to the pentacoordinated α-phosphorous (Figure 4.2b). The pyrophosphate leaving group is pushed out. The latter reaction

mechanism is favored since pyrophosphate, a relatively big and bulky leaving group, is more suitable for an associative “back face” attack. The activated carboxyl-adenylate intermediate is then attacked by a nucleophilic substrate liberating ester, amide or thioester products (depending on the utilized nucleophile) and adenosine monophosphate (Challis 2005; Abe et al. 2008).

Class I adenylate-forming enzymes all have a large N-terminal domain with a small C-terminal domain connected via a flexible hinge (see section 1.2.1). The relative orientation of the two domains varies presumably as a consequence of the hinge. The C- and N-terminal domain were described as a “hammer-and-anvil” structure. The N-terminal domain (anvil) is subdivided into three sub-domains with an α/β topology, topped by a αβαβα sandwich and a β-barrel (Conti et al. 1996). The C-terminal domain (hammer) sits on top like a “lid” and is made up of five β-strands and three α-helices. The two-domain architecture is consistent for all class I enzymes. Class II enzymes do not show such conserved overall domain arrangement. Subclass IIa aminoacyl-tRNA synthetases (aaRSs) have four domains: a zinc-binding, central catalytic, anticodon-binding and an amino acid specific domain. Some aaRSs (as LeuRS, IleRS or ValRS) have an additional proofreading/editing domain, called CP1 (connective polypeptide 1), which was reported to be important for enzyme fidelity (Figure 1.3c) (Mascarenhas and Martinis 2008). In contrast to class IIa enzymes, subclass IIb consists of an anticodon-binding domain and the central catalytic domain, although additional insertions are known (Figure 1.3) (Xie et al. 2007). While subclass IIa aaRSs are generally monomeric, subclass IIb aaRSs are dimeric or tetrameric (Cusack 1997). Class III enzymes have three domains (Figure 2.1 and

Figure 3.11), an N-terminal domain, a central domain and a C-terminal domain (Schmelz et al. 2009). Until more structures are solved in this class, consistency of their overall architecture remains to be proven.

4.2.2 ATP binding

The catalytic site in adenylate-forming enzymes varies from class to class. In class I enzymes the active site is formed by the interface between the N-terminal and C-terminal domain, which are connected by a flexible hinge (Wu et al. 2009). The domains significantly alter their relative orientations to adjust the catalytic site depending whether the protein is catalyzing adenylate formation or the second reaction. For example in human acyl-CoA synthetase (Kochan et al. 2009) a conserved catalytic lysine important for adenylation is moved out after adenylate formation to facilitate thioester formation (Figure 4.3a) (Kochan et al. 2009; Osman et al. 2009).

Class II enzymes locate the active site in the central catalytic domain. Other than in class I enzymes no large conformational changes are observed in the catalytic domain upon substrate binding. However significant domain movements of the editing and the leucine-specific domain in leucyl-aaRS upon leucyl-tRNA coordination were identified (Weimer et al. 2009) (Figure 4.3b). In ProRS and HisRS from *T. thermophilus*, subclass IIb aaRSs, conformational changes in the active site (such as ordering adjacent loops) coupled to global inter-domain and inter-subunit movements have been reported (Yaremchuk et al. 2001).

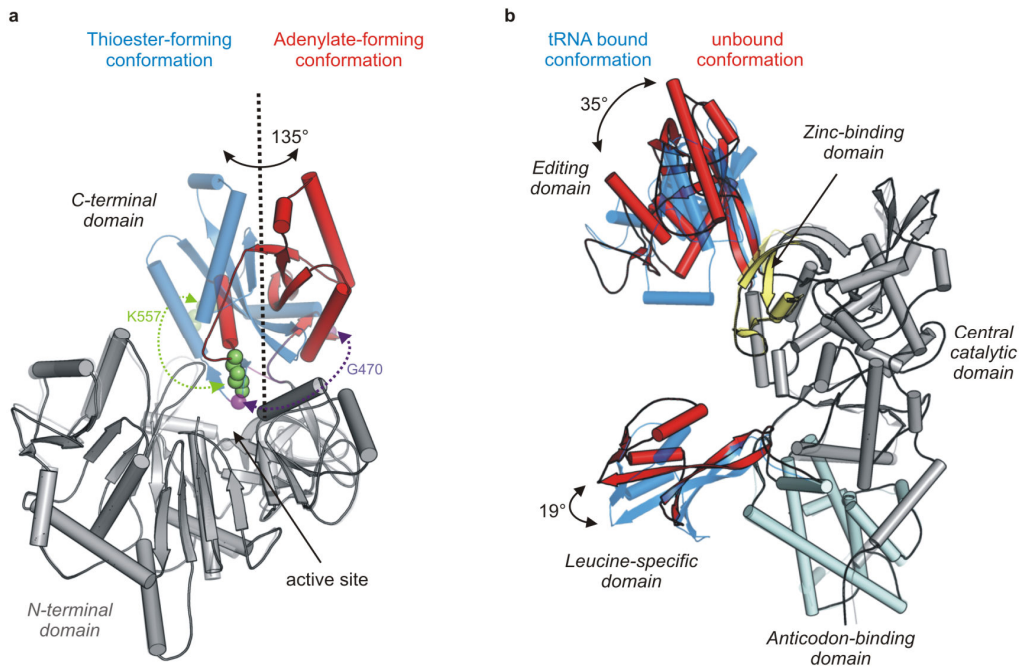


Figure 4.3: Domain movement in class I and II adenylating enzymes. **(a)** Domain movement in class I enzymes. Changing from adenylate-forming conformation (red) to thioester-forming conformation (blue) results in a domain rotation of up to 135°. A conserved catalytic lysine (K557 in human acyl-CoA synthetase, colored in green) is moved out while a glycine residue (G470, colored in purple) is moved in the active site when shifting from adenylate to thioester-forming state. The rotation axis is indicated as a dotted line. Used PDB codes are 3c5e (adenylate-forming conformation) and 2wd9 (thioester-forming conformation). Residues 289-303 of the N-terminal domain are hidden to allow view on active site. **(b)** Domain movement in class IIa enzymes upon tRNA coordination. While the central catalytic domain does not change conformation the editing and the leucine-specific domain shift 35° and 19° in Leucyl-tRNA synthetase upon tRNA coordination, respectively. Used PDB codes are 1h3n (unbound conformation (colored in red), coordinating a leucyl-adenylate analogue) and 2byt (tRNA bound conformation, colored in blue).

Class III enzymes employ all three domains to make a deep cavity in the middle of the protein (Schmelz et al. 2009), which, apart from limited ordering of surrounding loops, do not alter their structure upon substrate binding. Class I and IIa enzymes bind ATP in a Rossmann-fold, but class IIb and class III do not. In subclass IIb aaRSs ATP is bound by three conserved motifs within a seven folded anti-parallel β-sheet (Cusack 1995; Crepin et al. 2006), but in class III enzymes ATP is bound by a novel fold, which shares parts of the cAPK nucleotide triphosphates binding fold (see also section 2.4.4).

4.2.3 Role of metals

All adenylate-forming enzymes are reported to be Mg^{2+} dependent (Airas 2007), but in at least one case Mn^{2+} can substitute for Mg^{2+} (Belrhali et al. 1995). Interestingly, biochemical studies show that human TyrRS requires K^{+} for catalysis. It is reported that it adopts the function of substrate activation, which is normally carried out by a missing but otherwise highly conserved lysine residue in subclass IIa aaRS enzymes (Austin and First 2002).

In each structure with ATP bound, a Mg^{2+} is reported to coordinate the phosphate. The divalent metal ion will neutralize the charge of ATP, stabilize the transition state and neutralize the leaving pyrophosphate. With such an acknowledged central role, it is surprising that the coordination geometry and even more the number of Mg^{2+} ions varies across the superfamily. The NRPS synthetase DltA (class Ia) (3fcc) shows a β – γ coordination (Yonus et al. 2008). Mg^{2+} in tryptophanyl-tRNA synthetase from *G. stearothermophilus* (1m83) (class IIa) is coordinated by α-β-γ phosphates unlike the human homologue (2qui) (Shen et al. 2008), where an α-β coordination is found. In AcsD, a class IIIa enzyme, a single Mg^{2+} coordinates the α–γ phosphate, a geometry only found in seven other structures consisting of kinases and the ribonucleotide transformylase PurT (1kj9). In AlcC a second Mg is bound to the α and β phosphate of the tri-phosphate. In contrast lysyl-tRNA synthetase, a subclass IIb enzyme, binds 3 metal ions, coordinating the α and β phosphate and the β and γ phosphates (Desogus et al. 2000).

4.2.4 Catalytic residues

At the α phosphate of ATP it is possible to identify chemical similarities of surrounding amino acids despite the discussed differences in protein structure. In every structure an Arg, Lys or His (Figure 4.4 and Figure 4.5) residue contacts the α phosphate but not always in a conserved position. In many but not all enzymes an additional positively charged residue surrounds the phosphate.

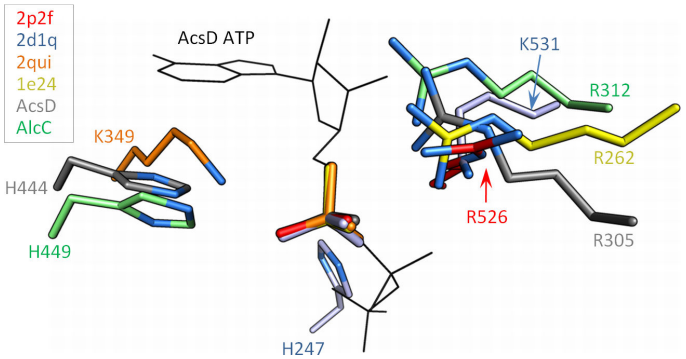


Figure 4.4: Coordination of α phosphate in different classes of adenylate-forming enzymes. In each member at least one positively charged residue, as lysine, arginine or histidine is present. Acetyl-CoA synthetase (2p2f) and Japanese firefly luciferase (2d1q) are **class I**; human tryptophanyl-tRNA synthetase (2qui) and lysyl-tRNA synthetase (1e24) are **class II**; AcsD and AlcC are **class III** adenylate-forming enzymes. ATP superposed in line representation is ATP of AcsD.

Class I enzymes use either Arg (Acetyl-CoA synthetase (Gulick et al. 2003)), Lys (GrsA and firefly luciferases) or His (firefly luciferase) near the α phosphate. Here a highly conserved threonine is also found in all these enzymes (Figure 4.5a). In class III two conserved residues, His and Arg (Figure 4.5c), are close to the α phosphate but neither of these residues do superpose with those in class I. A strictly conserved arginine found in all class IIa or IIb type tRNA synthetases (R262 in LysU, Figure 4.5b) does superimpose with R305 in AcsD or R312 in AlcC. These residues (along with metal ion(s)) polarize the α phosphate of ATP enhancing the electrophilicity of the phosphorus. The positive charge will also polarize the negative charge of the

incoming carboxylate increasing its nucleophilicity. The surrounding positive charges crucially will stabilize the negatively charged pentavalent transition state.

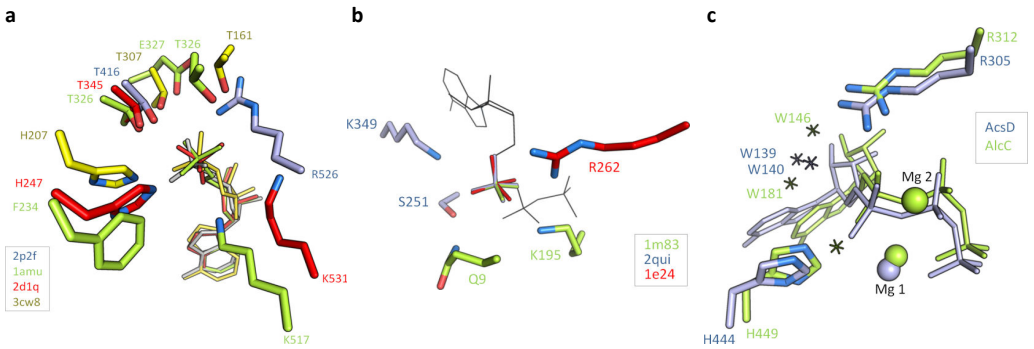


Figure 4.5: Active site alignment of different classes of adenylate-forming enzymes. Alignment was carried out on α phosphate of ATP or AMP. **a** Alignment of class I enzymes of Acetyl-CoA synthetase (2p2f), gramicidin S synthetase 1 (1amu), Japanese firefly luciferase (2d1q) and 4-Chlorobenzoyl-CoA ligase (3cw8). **b** Alignment of class II enzymes tryptophanyl-tRNA synthetase (1m83), human tryptophanyl-tRNA synthetase (2qui) and lysyl-tRNA synthetase (1e24), alignment of α phosphate of ATP is shown, ATP in grey line representation is ATP of 2qui. **c** Alignment of class III enzyme: AcsD colored in light blue (2w02) and AlcC colored in green, Mg 1 ion in AcsD and AlcC show an α-γ phosphate coordination, while AlcC coordinates a second Mg ion (Mg 2) in the α-β position (green). In contrast to class I and II water in class III enzymes is close to the α phosphate.

In all three classes point mutations of these positively charged residues show decreased enzyme activity. In class I enzymes mutations of the conserved lysine (K529R or K529Q in firefly luciferase) facing the α phosphate of ATP caused a ~1650 fold reduction of activity when changed into glutamine (Branchini et al. 1999). In class II aaRSs K233 mutants (second lysine in the conserved ATP binding KMSKS motif) of tyrosyl-tRNA synthetase, have been reported to deteriorate the ability to form tyrosine-adenylate (Fersht et al. 1988; First and Fersht 1993). Similar results were obtained for AcsD, a class IIIa enzyme, loosing activity after mutation of positively charged R305 and H444 (see section 2.3.7).

4.2.5 The release of pyrophosphate

The pyrophosphate leaving group in class III appears to be retained after adenylate formation in a pocket which extends beyond the γ phosphate. In class I enzymes pyrophosphate is released after forming the intermediate. This was examined in the adenylation domain of gramicidin S synthetase (Conti et al. 1996). In class II leucyl- or tyrosyl-tRNA synthetases, pyrophosphate is reported to be released after the amino acid adenylate was formed to mediate binding of tRNA which initiates the second step of tRNA amino-alkylation (Lin et al. 1975; Kuratani et al. 2006).

4.3 Conclusion

Adenylate-forming enzymes have important and diverse roles in metabolic pathways of prokaryotes and eukaryotes. We suggest that an inclusive definition of the superfamily based on the commonality of the adenylation chemistry may be more useful than one limited to three dimensional structures. Adenylating enzymes show that the same chemistry can be catalyzed by very different structures. Despite the profound differences in structure and active site architecture, there are clear chemical similarities in the environment that surrounds the α phosphate. This chemical conservation is of course driven by the requirement to stabilize the pentavalent negatively charged transition state. From a chemical standpoint, the nucleophilic attack of the α or γ phosphate by a hydroxyl are closely related, because in both cases a negatively charged nucleophile attacks. For instance, class III enzymes and kinases (as discussed in chapter 2) share a similar coordination of the ATP tri-phosphates, but not the same binding sites for the adenosine moieties, which are located opposite each other. This aligns the α-phosphorous of class III enzymes and the γ-phosphorous of kinases in identical position and hence allows a similar nucleophilic attack at the phosphate.

Chemical conservation has been observed in other enzyme families such as proteases. Though the different families of proteases use different ways to activate a water molecule or another nucleophile to achieve the cleavage of peptide bonds, they all need to stabilize the tetrahedral transition state.

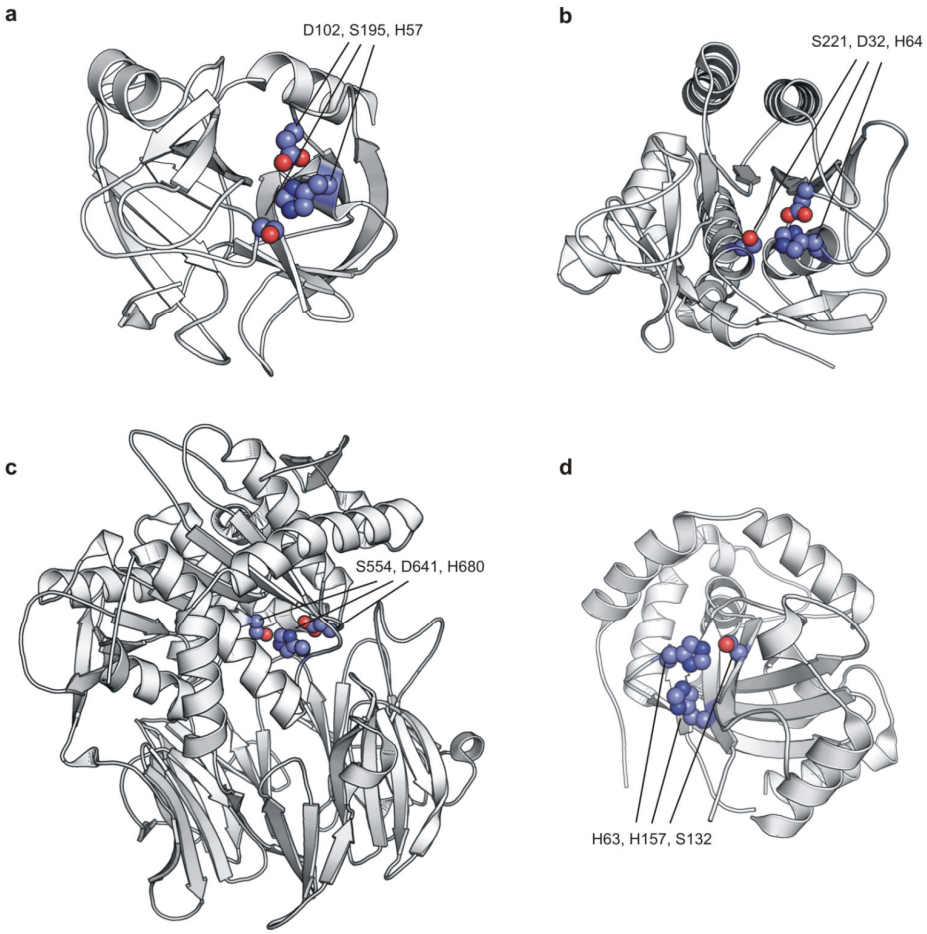


Figure 4.6: Chemical conservation of the serine protease triad. Serine proteases share a catalytic triad (His-Asp-Ser), but are structurally and evolutionary distinct. (a) Trypsin protease (1S81, pork) with catalytic residues H57, D102 and S195. (b) Subtilisin (2ST1, *Bacillus amyloliquefaciens*) with catalytic residues D32, H64 and S221. (c) Prolyl oligopeptidase (3EQ7, pork) with residues S554, D641 and H680. (d) HCMV protease (1NJU, human herpesvirus 5) with a different catalytic triad: H63, S132 and H157.

Serine proteases, which account for one third of all proteases, for instance comprise the well known catalytic triad (Asp-His-Ser). However serine proteases are structurally diverse as a comparison of trypsin, subtilisin and prolyl oligopeptidase shows (Figure 4.6a-c) (Page and Di Cera 2008). Interestingly, variations of the catalytic triad (e.g. His-Ser-His or Asp-Glu-Ser) are also found (Figure 4.6d) (Polgar 2005). The nucleophilic serine attacks the carbonyl of the substrate (protein or peptide) and forms a covalent acyl-ester intermediate (Hedstrom 2002). The two

assisting residues in the triad are needed to act as a general base (histidine) and an acid (aspartic acid). During reaction the base accepts a proton from the serine hydroxyl moiety and the acid protonates the amino leaving group of the substrate. This arrangement and the oxyanion hole (formed by S221 and N155 in subtilisin) stabilize the formed negative charge of the tetrahedral intermediate (Polgar 2005).

Akin to proteases, adenylating enzymes have utilized different spatial arrangements of residues to stabilize the transition state. This is primarily accomplished by the use of one or two positively charged residues such as His, Lys or Arg and the coordination of up to three Mg ions. It is likely that the requirement to activate carboxylic acids for condensation is particularly ancient (thioester, amide and ester bonds are extremely common chemical building blocks) and this may have driven the convergent evolution seen in the structures of the adenylate forming enzymes (Schmelz and Naismith 2009). Furthermore adenylate forming enzymes are not only interesting enzymes, but have utilities in commercial significant biotransformation or biosynthetic chemistry (Watanabe et al. 2006; Challis 2008; Fisch et al. 2009; Meier and Burkart 2009).

4.4 Future work

Class III adenylate-forming enzymes have a three domain topology. Since only two structures (AcsD and AlcC) have been solved, it remains to be proven whether this topology is conserved or not. It therefore is essential to solve more structures of class III enzymes. So far preliminary work was carried out to obtain crystals for AcsA and DesD (see appendix F and G). DesD is another class IIIc and AcsA a class IIIb enzyme. The structure of AcsA would be interesting, since it would be the first structure of a class IIIb enzyme and would reveal whether class IIIb enzymes share a similar topology with class IIIa and IIIc enzymes. Another interesting, yet unanswered, point is the ability of AlcC (as a processive enzyme) to catalyze two or three reactions while AcsD, which shares a homolog topology, just catalyzes one. It is important to take a closer look at the substrate binding sites to identify which residues turns one enzyme into a processive and the other into a single reaction catalyzing one.

Modification of the substrate specificity of other adenylate forming enzymes such as class I enzymes is interesting. For AcsD, a class III enzyme, it was shown that it is capable of utilizing a variety of different types of nucleophiles (-OH, -NH_2 and –SH) and that mutation can shift its substrate specificity. The enzymatic synthesized products may have use in biotransformation. The question is whether class I enzymes can be modified to utilize artificial compounds as well. One focus might be on a class I enzyme: the benzoate CoA ligase from *Burkholderia xenovorans* (Bains and Boulanger 2007). Bains and Boulanger demonstrated that benzoate CoA ligase is

also capable of utilizing fluorinated benzoates but not chlorobenzoates. Modification of its active site might adapt its substrate specificity to allow oxidation of chlorobenzoates as well. Another potential enzyme could be 4-Chlorobenzoate ligase from *Pseudomonas* species (Chang et al. 1997). Engineered microorganisms "equipped" with such enzymes for instance would be valuable tools in bacteriological treatment approaches of chemically polluted areas.

Bibliography

Abdalla, M. (2001). "Sudden decline of date palm trees caused by *Erwinia chrysanthemi*." Plant Disease **85**(1): 24-6.

Abe, T., Y. Hashimoto, H. Hosaka, K. Tomita-Yokotani and M. Kobayashi (2008). "Discovery of amide (peptide) bond synthetic activity in Acyl-CoA synthetase." J Biol Chem **283**(17): 11312-21.

Airas, R. K. (2007). "Magnesium dependence of the measured equilibrium constants of aminoacyl-tRNA synthetases." Biophys Chem **131**(1-3): 29-35.

Atkin, C. L., L. Thelander, P. Reichard and G. Lang (1973). "Iron and free radical in ribonucleotide reductase. Exchange of iron and Mossbauer spectroscopy of the protein B2 subunit of the Escherichia coli enzyme." J Biol Chem **248**(21): 7464-72.

Austin, J. and E. A. First (2002). "Potassium functionally replaces the second lysine of the KMSKS signature sequence in human tyrosyl-tRNA synthetase." J Biol Chem **277**(23): 20243-8.

Bains, J. and M. J. Boulanger (2007). "Biochemical and structural characterization of the paralogous novel benzoate oxidation (box) pathway." J Mol Biol **373**(4): 965-77.

Baldwin, T. O. (1996). "Firefly luciferase: the structure is known, but the mystery remains." Structure **4**(3): 223-8.

Beaumont, F. C., H. Y. Kang, T. J. Brickman and S. K. Armstrong (1998). "Identification and characterization of alcR, a gene encoding an AraC-like regulator of alcaligin siderophore biosynthesis and transport in Bordetella pertussis and Bordetella bronchiseptica." J Bacteriol **180**(4): 862-70.

Belrhali, H., A. Yaremchuk, M. Tukalo, C. Berthet-Colominas, B. Rasmussen, P. Bosecke, O. Diat and S. Cusack (1995). "The structural basis for seryl-adenylate and Ap4A synthesis by seryl-tRNA synthetase." Structure **3**(4): 341-52.

Beri, R. and R. Chandra (1993). "Chemistry and biology of heme. Effect of metal salts, organometals, and metalloporphyrins on heme synthesis and catabolism, with special reference to clinical implications and interactions with cytochrome P-450." Drug Metab Rev **25**(1-2): 49-152.

Berti, A. D. and M. G. Thomas (2009). "Analysis of achromobactin biosynthesis by *Pseudomonas syringae pv. syringae B728a."* J Bacteriol **191**(14): 4594-604.

Branchini, B. R., R. A. Magyar, M. H. Murtiashaw, S. M. Anderson, L. C. Helgerson and M. Zimmer (1999). "Site-directed mutagenesis of firefly luciferase active site amino acids: a proposed model for bioluminescence color." Biochem **38**(40): 13223-30.

Brickman, T. J., M. T. Anderson and S. K. Armstrong (2007). "Bordetella iron transport and virulence." Biometals **20**(3-4): 303-22.

Brickman, T. J. and S. K. Armstrong (2009). "Temporal signaling and differential expression of Bordetella iron transport systems: the role of ferrimones and positive regulators." Biometals **22**(1): 33-41.

Brickman, T. J., T. Hanawa, M. T. Anderson, R. J. Suhadolc and S. K. Armstrong (2008). "Differential expression of *Bordetella pertussis* iron transport system genes during infection." Mol Microbiol **70**(1): 3-14.

Brown, N. R., M. E. Noble, J. A. Endicott and L. N. Johnson (1999). "The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases." Nat Cell Biol **1**(7): 438-43.

Cendrowski, S., W. MacArthur and P. Hanna (2004). "*Bacillus anthracis* requires siderophore biosynthesis for growth in macrophages and mouse virulence." Mol Microbiol **51**(2): 407-17.

Challis, G. L. (2005). "A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases." Chembiochem **6**(4): 601-11.

Challis, G. L. (2007). "A Widely Distributed Bacterial Pathway for Siderophore Biosynthesis Independent of Nonribosomal Peptide Synthetases." Chembiochem **8**(13): 1477.

Challis, G. L. (2008). "Mining microbial genomes for new natural products and biosynthetic pathways." Microbiology **154**(Pt 6): 1555-69.

Challis, G. L. and J. H. Naismith (2004). "Structural aspects of non-ribosomal peptide biosynthesis." Curr Opin Struct Biol **14**(6): 748-56.

Chance, B. (1967). "The reactivity of haemoproteins and cytochromes." Biochem J **103**(1): 1-18.

Chang, K. H., H. Xiang and D. Dunaway-Mariano (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." Biochem **36**(50): 15650-9.

Collaborative Computational Project Number 4. (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallographica Section D **50**(5): 760-3.

Conti, E., N. P. Franks and P. Brick (1996). "Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes." Structure **4**(3): 287-98.

Conti, E., T. Stachelhaus, M. A. Marahiel and P. Brick (1997). "Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S." Embo J **16**(14): 4174-83.

Crepin, T., A. Yaremchuk, M. Tukalo and S. Cusack (2006). "Structures of two bacterial prolyl-tRNA synthetases with and without a cis-editing domain." Structure **14**(10): 1511-25.

Crosa, J. H. and C. T. Walsh (2002). "Genetics and assembly line enzymology of siderophore biosynthesis in bacteria." Microbiol Mol Biol Rev **66**(2): 223-49.

Cusack, S. (1995). "Eleven down and nine to go." Nat Struct Biol **2**(10): 824-31.

Cusack, S. (1997). "Aminoacyl-tRNA synthetases." Curr Opin Struct Biol **7**(6): 881-9.

Cusack, S., C. Berthet-Colominas, M. Hartlein, N. Nassar and R. Leberman (1990). "A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 A." Nature **347**(6290): 249-55.

Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, 3rd, J. Snoeyink, J. S. Richardson and D. C. Richardson (2007). "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids." Nucleic Acids Res **35**(Web Server issue): W375-83.

de Lorenzo, V., A. Bindereif, B. H. Paw and J. B. Neilands (1986). "Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in *Escherichia coli* K-12." J Bacteriol **165**(2): 570-8.

de Lorenzo, V. and J. B. Neilands (1986). "Characterization of iucA and iucC genes of the aerobactin system of plasmid ColV-K30 in *Escherichia coli*." J Bacteriol **167**(1): 350-5.

Derewenda, Z. S. (2004). "Rational protein crystallization by mutational surface engineering." Structure **12**(4): 529-35.

Desogus, G., F. Todone, P. Brick and S. Onesti (2000). "Active site of lysyl-tRNA synthetase: structural studies of the adenylation reaction." Biochem **39**(29): 8418-25.

Dong, A., X. Xu, A. M. Edwards, C. Chang, M. Chruszcz, M. Cuff, M. Cymborowski, R. Di Leo, O. Egorova, E. Evdokimova, E. Filippova, J. Gu, J. Guthrie, A. Ignatchenko, A. Joachimiak, N. Klostermann, Y. Kim, Y. Korniyenko, W. Minor, Q. Que, A. Savchenko, T. Skarina, K. Tan, A. Yakunin, A. Yee, V. Yim, R. Zhang, H. Zheng, M. Akutsu, C. Arrowsmith, G. V. Avvakumov, A. Bochkarev, L. G. Dahlgren, S. Dhe-Paganon, S.

Dimov, L. Dombrovski, P. Finerty, Jr., S. Flodin, A. Flores, S. Graslund, M. Hammerstrom, M. D. Herman, B. S. Hong, R. Hui, I. Johansson, Y. Liu, M. Nilsson, L. Nedyalkova, P. Nordlund, T. Nyman, J. Min, H. Ouyang, H. W. Park, C. Qi, W. Rabeh, L. Shen, Y. Shen, D. Sukumard, W. Tempel, Y. Tong, L. Tresagues, M. Vedadi, J. R. Walker, J. Weigelt, M. Welin, H. Wu, T. Xiao, H. Zeng and H. Zhu (2007). "In situ proteolysis for protein crystallization and structure determination." Nat Methods **4**(12): 1019-21.

Drechsel, H. and G. Jung (1998). "Peptide siderophores." J Pept Sci **4**(3): 147-81.

Drew, D. E., G. von Heijne, P. Nordlund and J. W. de Gier (2001). "Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*." FEBS Lett **507**(2): 220-4.

Du, L., C. Sanchez, M. Chen, D. J. Edwards and B. Shen (2000). "The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase." Chem Biol **7**(8): 623-42.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.

Eriani, G., M. Delarue, O. Poch, J. Gangloff and D. Moras (1990). "Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs." Nature **347**(6289): 203-6.

Fersht, A. R., J. W. Knill-Jones, H. Bedouelle and G. Winter (1988). "Reconstruction by site-directed mutagenesis of the transition state for the activation of tyrosine by the tyrosyl-tRNA synthetase: a mobile loop envelopes the transition state in an induced-fit mechanism." Biochem **27**(5): 1581-7.

First, E. A. and A. R. Fersht (1993). "Mutation of lysine 233 to alanine introduces positive cooperativity into tyrosyl-tRNA synthetase." Biochem **32**(49): 13651-7.

Fisch, K. M., A. F. Gillaspy, M. Gipson, J. C. Henrikson, A. R. Hoover, L. Jackson, F. Z. Najar, H. Wagele and R. H. Cichewicz (2009). "Chemical induction of silent biosynthetic pathway transcription in *Aspergillus niger*." J Ind Microbiol Biotechnol. **36**(9): 1199-213.

Franza, T., B. Mahe and D. Expert (2005). "*Erwinia chrysanthemi* requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection." Mol Microbiol **55**(1): 261-75.

Franza, T., I. Michaud-Soret, P. Piquerel and D. Expert (2002). "Coupling of iron assimilation and pectinolysis in *Erwinia chrysanthemi* 3937." Mol Plant Microbe Interact **15**(11): 1181-91.

Fulda, M., E. Heinz and F. P. Wolter (1994). "The fadD gene of *Escherichia coli* K12 is located close to rnd at 39.6 min of the chromosomal map and is a new member of the AMP-binding protein family." Mol Gen Genet **242**(3): 241-9.

Galanis, E., A. S. King, P. Varughese and S. A. Halperin (2006). "Changing epidemiology and emerging risk groups for pertussis." CMAJ **174**(4): 451-2.

Gehring, A. M., I. Mori and C. T. Walsh (1998). "Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF." Biochemistry **37**(8): 2648-59.

Giege, R., B. Lorber, J. P. Ebel, D. Moras and J. C. Thierry (1980). "[Crystallization of the complex formed between yeast aspartyl tRNA and its specific aminoacyl tRNA synthetase]." C R Seances Acad Sci D **291**(4): 393-6.

Giege, R., B. Lorber, J. P. Ebel, D. Moras, J. C. Thierry, B. Jacrot and G. Zaccai (1982). "Formation of a catalytically active complex between tRNAAsp and aspartyl-tRNA synthetase from yeast in high concentrations of ammonium sulphate." Biochimie **64**(5): 357-62.

Ginder, N. D., D. J. Binkowski, H. J. Fromm and R. B. Honzatko (2006). "Nucleotide complexes of Escherichia coli phosphoribosylaminoimidazole succinocarboxamide synthetase." J Biol Chem **281**(30): 20680-8.

Glanz, J. M., D. L. McClure, D. J. Magid, M. F. Daley, E. K. France, D. A. Salmon and S. J. Hambidge (2009). "Parental refusal of pertussis vaccination is associated with an increased risk of pertussis infection in children." Pediatrics **123**(6): 1446-51.

Goldschmidt, L., D. R. Cooper, Z. S. Derewenda and D. Eisenberg (2007). "Toward rational protein crystallization: A Web server for the design of crystallizable protein variants." Protein Sci **16**(8): 1569-76.

Goodnow, R. A. (1980). "Biology of *Bordetella bronchiseptica*." Microbiol Rev **44**(4): 722-38.

Gulick, A. M., V. J. Starai, A. R. Horswill, K. M. Homick and J. C. Escalante-Semerena (2003). "The 1.75 A crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A." Biochem **42**(10): 2866-73.

Hedstrom, L. (2002). "Serine protease mechanism and specificity." Chem Rev **102**(12): 4501-24.

Hellenbrand, W., D. Beier, E. Jensen, M. Littmann, C. Meyer, H. Oppermann, C. H. Wirsing von Konig and S. Reiter (2009). "The epidemiology of pertussis in Germany: past and present." BMC Infect Dis **9**: 22.

Higgins, I. J., D. J. Best, R. C. Hammond and D. Scott (1981). "Methane-oxidizing microorganisms." Microbiol Rev **45**(4): 556-90.

Hisanaga, Y., H. Ago, N. Nakagawa, K. Hamada, K. Ida, M. Yamamoto, T. Hori, Y. Arii, M. Sugahara, S. Kuramitsu, S. Yokoyama and M. Miyano (2004). "Structural basis of the substrate-specific two-step catalysis of long chain fatty acyl-CoA synthetase dimer." J Biol Chem **279**(30): 31717-26.

Holm, L. and C. Sander (1993). "Protein structure comparison by alignment of distance matrices." J Mol Biol **233**(1): 123-38.

Hua, Y., R. F. Keep, J. T. Hoff and G. Xi (2008). "Deferoxamine therapy for intracerebral hemorrhage." Acta Neurochir Suppl **105**: 3-6.

Hugouvieux-Cotte-Pattat, N., G. Condemine, W. Nasser and S. Reverchon (1996). "Regulation of pectinolysis in *Erwinia chrysanthemi*." Annu Rev Microbiol **50**: 213-57.

Irwin, M. J., J. Nyborg, B. R. Reid and D. M. Blow (1976). "The crystal structure of tyrosyl-transfer RNA synthetase at 2-7 A resolution." J Mol Biol **105**(4): 577-86.

Jogl, G. and L. Tong (2004). "Crystal structure of yeast acetyl-coenzyme A synthetase in complex with AMP." Biochem **43**(6): 1425-31.

Kabsch, W. (1993). "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants." J Appl Crystallogr **26**(6): 795-800.

Kadi, N., S. Arbache, L. Song, D. Oves-Costales and G. L. Challis (2008). "Identification of a gene cluster that directs putrebactin biosynthesis in Shewanella species: PubC catalyzes cyclodimerization of N-hydroxy-N-succinylputrescine." J Am Chem Soc **130**(32): 10458-9.

Kadi, N. and G. L. Challis (2009). "Chapter 17. Siderophore biosynthesis a substrate specificity assay for nonribosomal peptide synthetase-independent siderophore synthetases involving trapping of acyl-adenylate intermediates with hydroxylamine." Methods Enzymol **458**: 431-57.

Kadi, N., D. Oves-Costales, F. Barona-Gomez and G. L. Challis (2007). "A new family of ATP-dependent oligomerization-macrocyclization biocatalysts." Nat Chem Biol **3**(10): 652-6.

Kang, H. Y. and S. K. Armstrong (1998). "Transcriptional analysis of the *Bordetella alcaligin* siderophore biosynthesis operon." J Bacteriol **180**(4): 855-61.

Kelly, S. M. and N. C. Price (1997). "The application of circular dichroism to studies of protein folding and unfolding." Biochim Biophys Acta **1338**(2): 161-85.

Kiese, M. (1966). "The biochemical production of ferrihemoglobin-forming derivatives from aromatic amines, and mechanisms of ferrihemoglobin formation." Pharmacol Rev **18**(3): 1091-161.

Kochan, G., E. S. Pilka, F. von Delft, U. Oppermann and W. W. Yue (2009). "Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A." J Mol Biol **388**(5): 997-1008.

Köster, W. (2001). "ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12." Res Microbiol **152**(3-4): 291-301.

Köster, W. (2005). "Cytoplasmic membrane iron permease systems in the bacterial cell envelope." Front Biosci **10**: 462-77.

Krissinel, E. and K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallogr D Biol Crystallogr **60**: 2256-68.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." J Mol Biol **372**(3): 774-97.

Kuratani, M., H. Sakai, M. Takahashi, T. Yanagisawa, T. Kobayashi, K. Murayama, L. Chen, Z. J. Liu, B. C. Wang, C. Kuroishi, S. Kuramitsu, T. Terada, Y. Bessho, M. Shirouzu, S. Sekine and S. Yokoyama (2006). "Crystal structures of tyrosyl-tRNA synthetases from Archaea." J Mol Biol **355**(3): 395-408.

Lautru, S. and G. L. Challis (2004). "Substrate recognition by nonribosomal peptide synthetase multi-enzymes." Microbiology **150**(Pt 6): 1629-36.

Lin, C. S., R. Irwin and J. G. Chirikjian (1975). "Kinetic studies of leucyl transfer RNA synthetase from bakers' yeast. Order of addition of substrates and release of products." J Biol Chem **250**(24): 9299-303.

Liu, H. and J. H. Naismith (2008). "An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol." BMC Biotechnol **8**: 91.

Lorber, B., R. Giege, J. P. Ebel, C. Berthet, J. C. Thierry and D. Moras (1983). "Crystallization of a tRNA . aminoacyl-tRNA synthetase complex. Characterization and first crystallographic data." J Biol Chem **258**(13): 8429-35.

Mandal, K. and S. Maiti (2005). "Bacterial soft rot of aloe caused by *Pectobacterium chrysanthemi*: a new report from India." Plant Pathology **54**(4): 573.

Marahiel, M. A., L. O. Essen and A. H. David (2009). Chapter 13 Nonribosomal Peptide Synthetases: Mechanistic and Structural Aspects of Essential Domains. Methods Enzymol, Academic Press. **Volume 458:** 337-51.

Marahiel, M. A., T. Stachelhaus and H. D. Mootz (1997). "Modular peptide synthetases involved in nonribosomal peptide synthesis." Chem Rev **97**(7): 2651-73.

Mascarenhas, A. P. and S. A. Martinis (2008). "Functional segregation of a predicted "hinge" site within the beta-strand linkers of *Escherichia coli* leucyl-tRNA synthetase." Biochem **47**(16): 4808-16.

May, J. J., N. Kessler, M. A. Marahiel and M. T. Stubbs (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**(19): 12120-5.

McMahon, S. A., M. Oke, H. Liu, K. A. Johnson, L. Carter, N. Kadi, M. F. White, G. L. Challis and J. H. Naismith (2008). "Purification, crystallization and data collection of *Pectobacterium chrysanthemi* AcsD, a type A siderophore synthetase." Acta Crystallogr F Struc Biol and Cryst Com **64**: 1052-5.

Meier, J. L. and M. D. Burkart (2009). "The chemical biology of modular biosynthetic enzymes." Chem Soc Rev **38**(7): 2012-45.

Melo, N., A. C. Dias, L. Isidoro and R. Duarte (2009). "*Bordetella pertussis*, an agent not to forget: a case report." Cases J **2**(1): 128.

Miller, D. A., L. Luo, N. Hillson, T. A. Keating and C. T. Walsh (2002). "Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*." Chem Biol **9**(3): 333-44.

Moore, C. H., L. A. Foster, D. G. Gerbig, Jr., D. W. Dyer and B. W. Gibson (1995). "Identification of alcaligin as the siderophore produced by Bordetella pertussis and B. bronchiseptica." J Bacteriol **177**(4): 1116-8.

Münzinger, M., H. Budzikiewicz, D. Expert, C. Enard and J. M. Meyer (2000). "Achromobactin, a new citrate siderophore of *Erwinia chrysanthemi*." Z Naturforsch C **55**(5-6): 328-32.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-55.

Nakatsu, T., S. Ichiyama, J. Hiratake, A. Saldanha, N. Kobashi, K. Sakata and H. Kato (2006). "Structural basis for the spectral difference in luciferase bioluminescence." Nature **440**(7082): 372-6.

Nishio, T., N. Tanaka, J. Hiratake, Y. Katsube, Y. Ishida and J. Oda (1988). "Isolation and structure of the novel dihydroxamate siderophore alcaligin." J Am Chem Soc **110**(26): 8733-4.

O'Donoghue, P. and Z. Luthey-Schulten (2003). "On the evolution of structure in aminoacyl-tRNA synthetases." Microbiol Mol Biol Rev **67**(4): 550-73.

Osman, K. T., L. Du, Y. He and Y. Luo (2009). "Crystal structure of *Bacillus cereus* D-alanyl carrier protein ligase (DltA) in complex with ATP." J Mol Biol **388**(2): 345-55.

Otwinowski, Z. and W. Minor (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymol. J. R. M. Sweet and Eds. New York, Academic Press. **276:** 307-26.

Oves-Costales, D., N. Kadi, M. J. Fogg, L. Song, K. S. Wilson and G. L. Challis (2007). "Enzymatic logic of anthrax stealth siderophore biosynthesis: AsbA catalyzes ATP-dependent condensation of citric acid and spermidine." J Am Chem Soc **129**(27): 8416-7.

Owen, D. J., M. E. Noble, E. F. Garman, A. C. Papageorgiou and L. N. Johnson (1995). "Two structures of the catalytic domain of phosphorylase kinase: an active protein kinase complexed with substrate analogue and product." Structure **3**(5): 467-82.

Page, M. J. and E. Di Cera (2008). "Serine peptidases: classification, structure and function." Cell Mol Life Sci **65**(7-8): 1220-36.

Painter, J. and E. A. Merritt (2006). "TLSMD web server for the generation of multi-group TLS models." J App Cryst **39**(1): 109-11.

Palacio-Bielsa, A., M. A. Cambra and M. M. Lopez (2006). "Characterisation of potato isolates of *Dickeya chrysanthemi* in Spain by a microtitre system for biovar determination." Annals App Biol **148**(2): 157-64.

Perombelon, M. C. M. and A. Kelman (1980). "Ecology of the Soft Rot Erwinias." Annu Rev Phytopath **18**: 361-87.

Perona, J. J. and Y. M. Hou (2007). "Indirect readout of tRNA for aminoacylation." Biochem **46**(37): 10419-32.

Polgar, L. (2005). "The catalytic triad of serine peptidases." Cell Mol Life Sci **62**(19-20): 2161-72.

Rapp, M., D. Drew, D. O. Daley, J. Nilsson, T. Carvalho, K. Melen, J. W. De Gier and G. Von Heijne (2004). "Experimentally based topology models for E. coli inner membrane proteins." Protein Sci **13**(4): 937-45.

Rees, D. C. (2002). "Great metalloclusters in enzymology." Annu Rev Biochem **71**: 221-46.

Register, K. B., T. F. Ducey, S. L. Brockmeier and D. W. Dyer (2001). "Reduced virulence of a *Bordetella bronchiseptica* siderophore mutant in neonatal swine." Infect Immun **69**(4): 2137-43.

Rouault, T. A. (2003). "How mammals acquire and distribute iron needed for oxygen-based metabolism." PLoS Biol **1**(3): E79.

Schmelz, S., N. Kadi, S. A. McMahon, L. J. Song, D. Oves-Costales, M. Oke, H. T. Liu, K. A. Johnson, L. G. Carter, C. H. Botting, M. F. White, G. L. Challis and J. H. Naismith (2009). "AcsD catalyzes enantioselective citrate desymmetrization in siderophore biosynthesis." Nat Chem Biol **5**(3): 174-82.

Schmelz, S. and J. H. Naismith (2009). "Adenylate-forming enzymes." Curr Opin Struct Biol.

Shen, N., M. Zhou, B. Yang, Y. Yu, X. Dong and J. Ding (2008). "Catalytic mechanism of the tryptophan activation reaction revealed by crystal structures of human tryptophanyl-tRNA synthetase in different enzymatic states." Nucleic Acids Res **36**(4): 1288-99.

Skamnaki, V. T., D. J. Owen, M. E. Noble, E. D. Lowe, G. Lowe, N. G. Oikonomakos and L. N. Johnson (1999). "Catalytic mechanism of phosphorylase kinase probed by mutational studies." Biochem **38**(44): 14718-30.

Slawiak, M., E. Lojkowska and J. M. van der Wolf (2009). "First report of bacterial soft rot on potato caused by *Dickeya* sp (syn. *Erwinia chrysanthemi*) in Poland." Plant Patho **58**(4): 794.

Steinbacher, S., P. Hof, L. Eichinger, M. Schleicher, J. Gettemans, J. Vandekerckhove, R. Huber and J. Benz (1999). "The crystal structure of the *Physarum polycephalum* actin-fragmin kinase: an atypical protein kinase with a specialized substrate-binding domain." Embo J **18**(11): 2923-9.

Suzuki, T. (2008). "[Iron overload and iron chelation therapy in transfusion-dependent patients]." Nippon Rinsho **66**(3): 563-8.

Tan, T., E. Trindade and D. Skowronski (2005). "Epidemiology of pertussis." Pediatr Infect Dis J **24**(5 Suppl): S10-8.

Taylor, S. S., E. Radzio-Andzelm, Madhusudan, X. Cheng, L. Ten Eyck and N. Narayana (1999). "Catalytic subunit of cyclic AMP-dependent protein kinase: structure and dynamics of the active site cleft." Pharmacol Ther **82**(2-3): 133-41.

Thoden, J. B., S. M. Firestine, S. J. Benkovic and H. M. Holden (2002). "PurT-encoded glycinamide ribonucleotide transformylase. Accommodation of adenosine nucleotide analogs within the active site." J Biol Chem **277**(26): 23898-908.

van Wageningen, A. M., P. N. Kirkpatrick, D. H. Williams, B. R. Harris, J. K. Kershaw, N. J. Lennard, M. Jones, S. J. Jones and P. J. Solenberg (1998). "Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic." Chem Biol **5**(3): 155-62.

Walker, E. H., M. E. Pacold, O. Perisic, L. Stephens, P. T. Hawkins, M. P. Wymann and R. L. Williams (2000). "Structural determinants of phosphoinositide 3-kinase inhibition by wortmannin, LY294002, quercetin, myricetin, and staurosporine." Mol Cell **6**(4): 909-19.

Warner, P. J., P. H. Williams, A. Bindereif and J. B. Neilands (1981). "ColV plasmid-specific aerobactin synthesis by invasive strains of *Escherichia coli*." Infect Immun **33**(2): 540-5.

Watanabe, K., K. Hotta, A. P. Praseuth, K. Koketsu, A. Migita, C. N. Boddy, C. C. Wang, H. Oguri and H. Oikawa (2006). "Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*." Nat Chem Biol **2**(8): 423-8.

Weber, G., K. Schorgendorfer, E. Schneider-Scherzer and E. Leitner (1994). "The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame." Curr Genet **26**(2): 120-5.

Weimer, K. M., B. L. Shane, M. Brunetto, S. Bhattacharyya and S. Hati (2009). "Evolutionary basis for the coupled-domain motions in Thermus thermophilus leucyl-tRNA synthetase." J Biol Chem **284**(15): 10088-99.

Wilson, M. K., R. J. Abergel, K. N. Raymond, J. E. Arceneaux and B. R. Byers (2006). "Siderophores of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*." Biochem Biophys Res Commun **348**(1): 320-5.

Winn, M. D., M. N. Isupov and G. N. Murshudov (2001). "Use of TLS parameters to model anisotropic displacements in macromolecular refinement." Acta Crystallogr D Biol Crystallogr **57**(Pt 1): 122-33.

Woese, C. R., G. J. Olsen, M. Ibba and D. Soll (2000). "Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process." Microbiol Mol Biol Rev **64**(1): 202-36.

Wu, M. X. and K. A. Hill (1993). "A continuous spectrophotometric assay for the aminoacylation of transfer RNA by alanyl-transfer RNA synthetase." Anal Biochem **211**(2): 320-3.

Wu, R., A. S. Reger, X. Lu, A. M. Gulick and D. Dunaway-Mariano (2009). "The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate: coenzyme A ligase." Biochem **48**(19): 4115-25.

Xie, W., L. A. Nangle, W. Zhang, P. Schimmel and X. L. Yang (2007). "Long-range structural effects of a Charcot-Marie-Tooth disease-causing mutation in human glycyl-tRNA synthetase." Proc Natl Acad Sci U S A **104**(24): 9976-81.

Xu, R. M., G. Carmel, R. M. Sweet, J. Kuret and X. Cheng (1995). "Crystal structure of casein kinase-1, a phosphate-directed protein kinase." Embo J **14**(5): 1015-23.

Yang, J., L. F. Ten Eyck, N. H. Xuong and S. S. Taylor (2004). "Crystal structure of a cAMP-dependent protein kinase mutant at 1.26A: new insights into the catalytic mechanism." J Mol Biol **336**(2): 473-87.

Yaremchuk, A., M. Tukalo, M. Grotli and S. Cusack (2001). "A succession of substrate induced conformational changes ensures the amino acid specificity of *Thermus thermophilus* prolyl-tRNA synthetase: comparison with histidyl-tRNA synthetase." J Mol Biol **309**(4): 989-1002.

Yonus, H., P. Neumann, S. Zimmermann, J. J. May, M. A. Marahiel and M. T. Stubbs (2008). "Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains." J Biol Chem **283**(47): 32484-91.

Yu, S., E. Fiss and W. R. Jacobs, Jr. (1998). "Analysis of the exochelin locus in Mycobacterium smegmatis: biosynthesis genes have homology with genes of the peptide synthetase family." J Bacteriol **180**(17): 4676-85.

Zheng, J., E. A. Trafny, D. R. Knighton, N. H. Xuong, S. S. Taylor, L. F. Ten Eyck and J. M. Sowadski (1993). "2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor." Acta Crystallogr D Biol Crystallogr **49**(Pt 3): 362-5.

Appendix

A. AcsD substrate coordination

Table 4.1: Interactions between AcsD and ATP

ATP atom	phosphate moiety	AcsD residue and atom	Distance Å	Domain
O1G/1	γ	Gln 446 NE2	3.17	3 – loop9
O1G/1	γ	Gln 446 OE1	3.20	3 – loop9
O1G/1	γ	Mg ion	2.04	
O1G/1	γ	Asp 464 OD2	3.22	3 – loop10
O2G/1	γ	Arg 369 NH2	2.65	2 – β12
O2G/1	γ	Lys 293 NZ	2.84	2 - β10
O2G/1	γ	Asp 464 OD1	2.63	3 – loop10
O2G/1	γ	Asp 464 OD2	3.36	3 – loop10
O3G/1	γ	Arg 369 NH1	3.03	2 – β12
O3G/1	γ	Ser 279 OG	2.90	2 - loop5
O3G/1	γ	Thr 282 OG1	2.74	2 - loop5
O3G/1	γ	H_2O 123	3.02	
O3B/1	γ	Ser 279 OG	3.26	2 - loop5
O1B/1	β	Arg 281 NH2	3.24	2 - loop5
O1B/1	β	Arg 281 NE	2.76	2 - loop5
O1B/1	β	H_2O 239	2.91	2
O1B/1	β	Asp 464 OD2	3.39	3 – loop10
O2B/1	β	Asp 464 OD2	2.91	3 – loop10
O2B/1	β	Asp 464 OD1	3.27	3 – loop10
O2B/1	β	Lys 293 NZ	2.90	2 - β10
O2B/1	β	Glu 466 OE1	3.47	3 – loop10
O3A/1	β	Asp 464 OD2	3.47	3 – loop10
O3A/1	β	Mg ion	2.51	
O1A/1	α	Arg 305 NH2	3.04	2 – loop7
O1A/1	α	H_2O 140	3.02	
O2A/1	α	Asp 464 OD2	3.19	3 – loop10
O2A/1	α	His 444 NE2	2.92	3 – loop9
O2A/1	α	Mg ion	2.45	
O5/1	α	Arg 305 NH2	3.01	2 – loop7
O5/1	α	H_2O 139	3.07	
O4/1	ribose	Gln 446 CB	3.23	3 – loop9
O2/1	ribose	Thr 301 OG1	2.85	2 – loop7
O2/1	ribose	Thr 301 CG2	3.04	2 – loop7
O2/1	ribose	H_2O 6	3.08	
N1/1	base	Asn 509 ND2	3.11	3 - α13
N1/1	base	Asn 509 ND1	3.48	3 - α13
C2/1	base	His 444 ND1	3.31	3 – loop9
C2/1	base	His 444 CE1	3.49	3 – loop9
N6/1	base	Asn 509 OD1	3.45	3 - α13
N6/1	base	H_2O 56	2.52	
N6/1	base	Leu 445 CD1	3.47	3 – loop9
N7/1	base	H_2O 56	2.76	
N7/1	base	His 170 CA	3.41	2 - loop4

Table 4.2: Interactions between AcsD and citrate (overlap with ATP)

atom label in structure	citrate atom	AcsD residue and atom	Distance Å	Domain
CGC /0	C1	Asp 464OD2	3.64	3 – loop 10
CB /0	C3	Asp 464OD2	3.22	3 – loop 10
CA /0	C4	Asp 464OD2	3.18	3 – loop 10
CA /0	C4	H_2O 42	3.33	
CA /0	C4	Ser 279CB	3.68	2 - loop5
CAC /0	C5	Asp 464OD2	3.54	3 – loop 10
CBC /0	C6	Arg 281NE	3.37	2 - loop5
CBC /0	C6	Arg 281NH2	3.54	2 - loop5
CBC /0	C6	Ser 279CB	3.57	2 - loop5
OG 2/0	O1	Arg 305 NH2	2.93	2 – loop7
OG 1/0	O2	Asp 464OD2	2.68	3 – loop 10
OG 1/0	O2	His 444NE2	3.65	3 – loop9
OA 1/0	O3	Arg 369NH1	3.65	2 – β 12
OA 1/0	O3	Lys 293CE	3.46	2 - β 10
OA 1/0	O3	Thr 282OG1	3.16	2 - loop5
OA 2/0	O4	Arg 369NH1	3.57	2 – β 12
OB 1/0	O5	Arg 281CZ	3.26	2 - loop5
OB 1/0	O5	Arg 281NE	2.80	2 - loop5
OB 1/0	O5	Arg 281NH2	3.09	2 - loop5
OB 1/0	O5	Lys 293CE	3.19	2 - β 10
OB 2/0	O6	Arg 281NE	3.36	2 - loop5
OB 2/0	O6	H_2O 222	2.48	
OB 2/0	O6	Ser 279CB	3.24	2 - loop5
OB 2/0	O6	Ser 279O	3.45	2 - loop5
OHB /0	O7	Asp 464CG	3.28	3 – loop 10
OHB /0	O7	Asp 464OD1	3.44	3 – loop 10
OHB /0	O7	Asp 464OD2	2.45	3 – loop 10
OHB /0	O7	Glu 466OE1	2.99	3 – loop 10

Table 4.3: Interactions between AcsD and citrate

atom label in structure	citrate atom	AcsD residue and atom	Distance Å	Domain
CGC /0	C1	His170 NE2	3.60	2 – loop4
CGC /0	C1	Tyr504 OH	3.32	3 – α13
CG /0	C2	ADN C2	3.60	
CG /0	C2	ADN N3	3.68	
CB /0	C3			
CA /0	C4	Arg576 NH1	3.38	3
CAC /0	C5	ADN C5’	3.08	
CAC /0	C5	Arg305 NH2	3.62	2 – loop7
CAC /0	C5	Glu442 OE2	3.59	3 – loop9
CBC /0	C6	Arg305 NH1	3.57	2 – loop7
CBC /0	C6	Thr301 OG1	3.12	2 – loop7
OG 2/0	O1	Lys563 CD	3.42	3 – β16
OG 2/0	O1	Lys563 CE	3.19	3 – β16
OG 2/0	O1	Lys563 NZ	3.25	3 – β16
OG 2/0	O1	Tyr504 CE2	3.57	3 – α13
OG 2/0	O1	Tyr504 OH	3.27	3 – α13
OG 1/0	O2	Asn565 ND2	3.60	3
OG 1/0	O2	Asn565 OD1	3.59	3
OG 1/0	O2	His170 CD2	3.63	2 – loop4
OG 1/0	O2	His170 CE1	3.58	2 – loop4
OG 1/0	O2	His170 NE2	2.64	2 – loop4
OG 1/0	O2	Tyr504 OH	2.59	3 – α13
OA 1/0	O3	ADN C5’	3.08	
OA 1/0	O3	ADN O5’	3.31	
OA 1/0	O3	Arg 305 NH2	2.75	2 – loop7
OA 2/0	O4	ADN C5’	3.47	
OA 2/0	O4	Glu442 OE2	3.27	3 – loop9
OA 2/0	O4	His444 CD2	3.34	3 – loop9
OA 2/0	O4	His444 CE1	3.58	3 – loop9
OA 2/0	O4	His444 NE2	2.62	3 – loop9
OB 1/0	O5	ADN C2’	3.33	
OB 1/0	O5	ADN C3’	3.09	
OB 1/0	O5	Arg305 CZ	3.21	2 – loop7
OB 1/0	O5	Arg305 NH1	2.88	2 – loop7
OB 1/0	O5	Arg305 NH2	2.80	2 – loop7
OB 1/0	O5	Thr301 OG1	3.00	2 – loop7
OB 2/0	O6	Asn302 CB	3.54	2 – loop7
OB 2/0	O6	Asn302 N	3.63	2 – loop7
OB 2/0	O6	Thr301 CB	3.26	2 – loop7
OB 2/0	O6	Thr301 OG1	2.50	2 – loop7
OHB /0	O7	Arg576 NH1	3.52	3
OHB /0	O7	Asn302 CB	3.46	2 – loop7
OHB /0	O7	Asn302 CG	3.40	2 – loop7
OHB /0	O7	Asn302 ND2	2.82	2 – loop7
OHB /0	O7	Lys563 NZ	3.63	3 – β16

Table 4.4: Interactions between AcsD and *N*-citryl-EDA

***N*-citryl-EDA atom**	**citrate atom**	**AcsD residue and atom**	**Distance Å**	**Domain**
CAM	C1	His170 NE2	3.67	2 – loop4
CAM	C1	Tyr504 OH	3.40	3 – α13
CAM	C1	H_2O 410	3.68	
CAJ	C2	His170 CD2	3.50	2 – loop4
CAJ	C2	His170 NE2	2.53	2 – loop4
CAP	C3			
CAK	C4			
CAN	C5	ATP O5’	3.58	
CAO	C6	Arg305 NH1	3.52	2 – loop7
CAO	C6	Thr301 OG1	3.27	2 – loop7
OAE	O1	Lys563 CD	3.68	3 – β16
OAE	O1	Lys563 CE	3.44	3 – β16
OAE	O1	Lys563 NZ	3.05	3 – β16
OAE	O1	Tyr504 CE2	3.67	3 – α13
OAE	O1	H_2O 410	2.48	
OAB	O2	Asn565 ND2	3.62	3
OAB	O2	His170 CD2	3.69	2 – loop4
OAB	O2	His170 NE2	2.87	2 – loop4
OAB	O2	Tyr504 OH	2.46	3 – α13
OAB	O2	Tyr504 CZ	3.40	3 – α13
OAB	O2	Tyr504 CE2	3.50	3 – α13
OAB	O2	Asn509 ND2	3.20	3 – α13
NAL	O3 (N1)	Glu442 OE2	3.07	3 – loop9
OAC	O4	Arg305 CZ	3.20	2 – loop7
OAC	O4	Arg305 NH1	2.68	2 – loop7
OAC	O4	Arg305 NH2	2.88	2 – loop7
OAC	O4	ATP O5’	3.39	
OAC	O4	ATP O1A	3.67	
OAD	O5	Arg305 CZ	3.33	2 – loop7
OAD	O5	Thr301 OG1	3.16	2 – loop7
OAD	O5	Thr301 CB	3.42	2 – loop7
OAD	O5	Asn302 N	2.93	2 – loop7
OAD	O5	Asn302 CA	3.50	2 – loop7
OAD	O5	Asn302 CB	2.99	2 – loop7
OAD	O5	Asn302 CG	3.25	2 – loop7
OAD	O5	Asn302 ND2	3.59	2 – loop7
OAF	O6	Arg305 NH1	3.45	2 – loop7
OAF	O6	ATP O2’	3.56	
OAF	O6	ATP C2’	3.22	
OAF	O6	Thr301 CB	3.35	2 – loop7
OAF	O6	Thr301 OG1	2.65	2 – loop7
OAG	O7	Asn302 ND2	3.52	2 – loop7
OAG	O7	Lys563 NZ	2.74	3 – β16
OAG	O7	Lys563 CE	3.42	3 – β16
CAI		ATP O1A	3.26	
CAH		Glu442 OE2	3.63	3 – loop9
CAH		Glu442 OE1	3.64	3 – loop9
NAA		Glu442 OE1	3.10	3 – loop9
NAA		H_2O 352	3.38	

B. Circular dichroism spectra of AcsD mutants

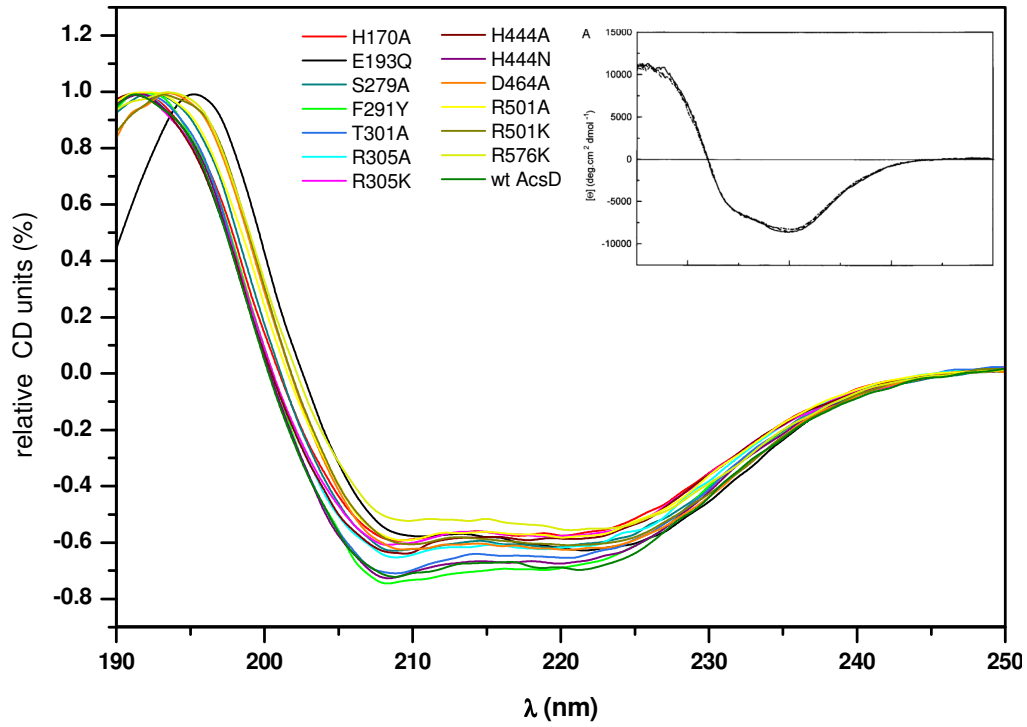


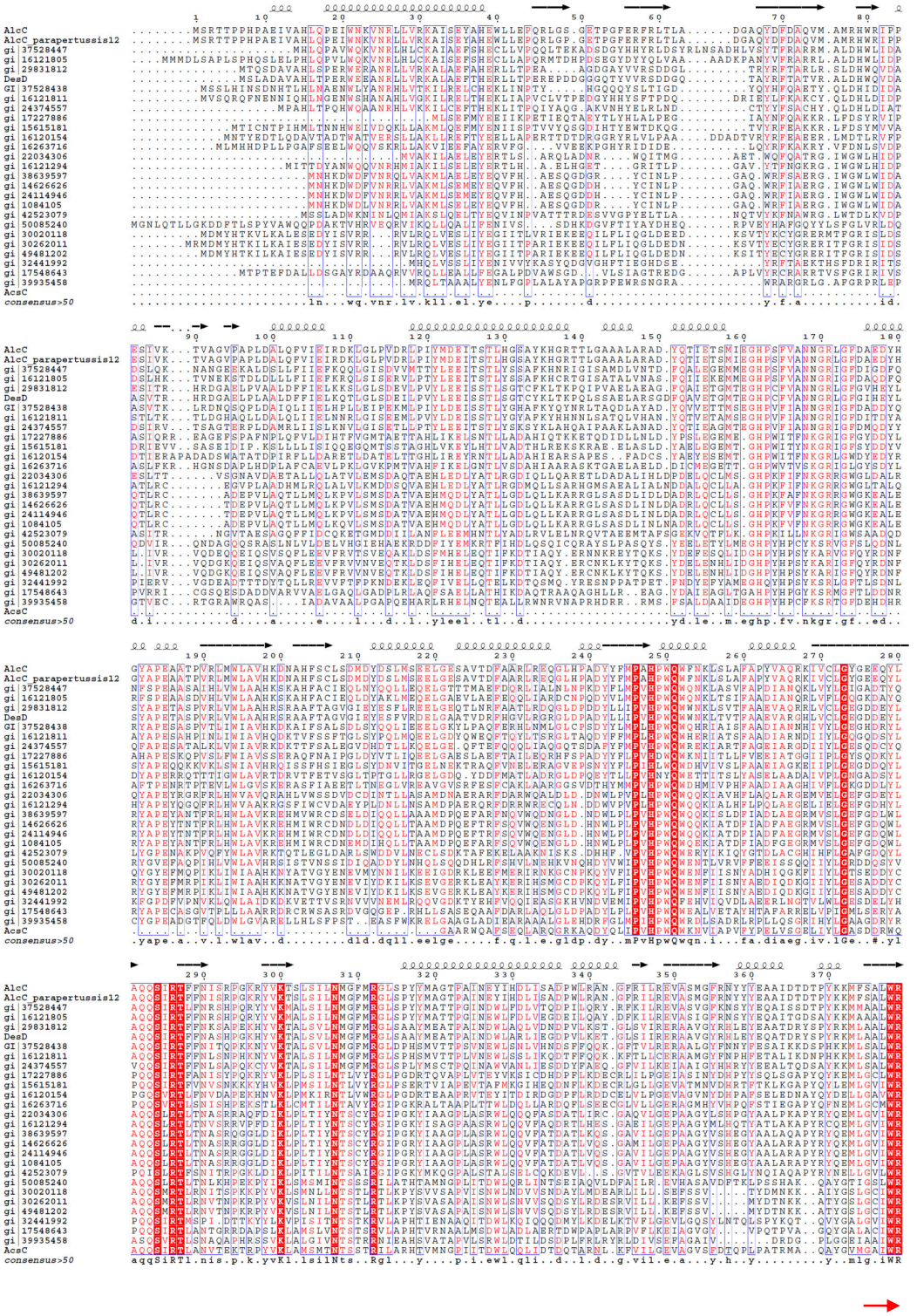
Figure 4.7: Circular dichroism spectra for AcsD mutants. Scans were carried out with 0.25-0.3 mg/ml protein dialysed in 25 mM phosphate buffer pH 7.5 using a JASCO J-810 CD spectropolarimeter across a 0.2 mm cuvette. All spectra have a characteristic shape for a folded protein as seen for type II dehydroquinase from *S. coelicolor* (see inset, taken from Kelly and Price 1997).

C. Sequence identity of AlcC and other NIS family members

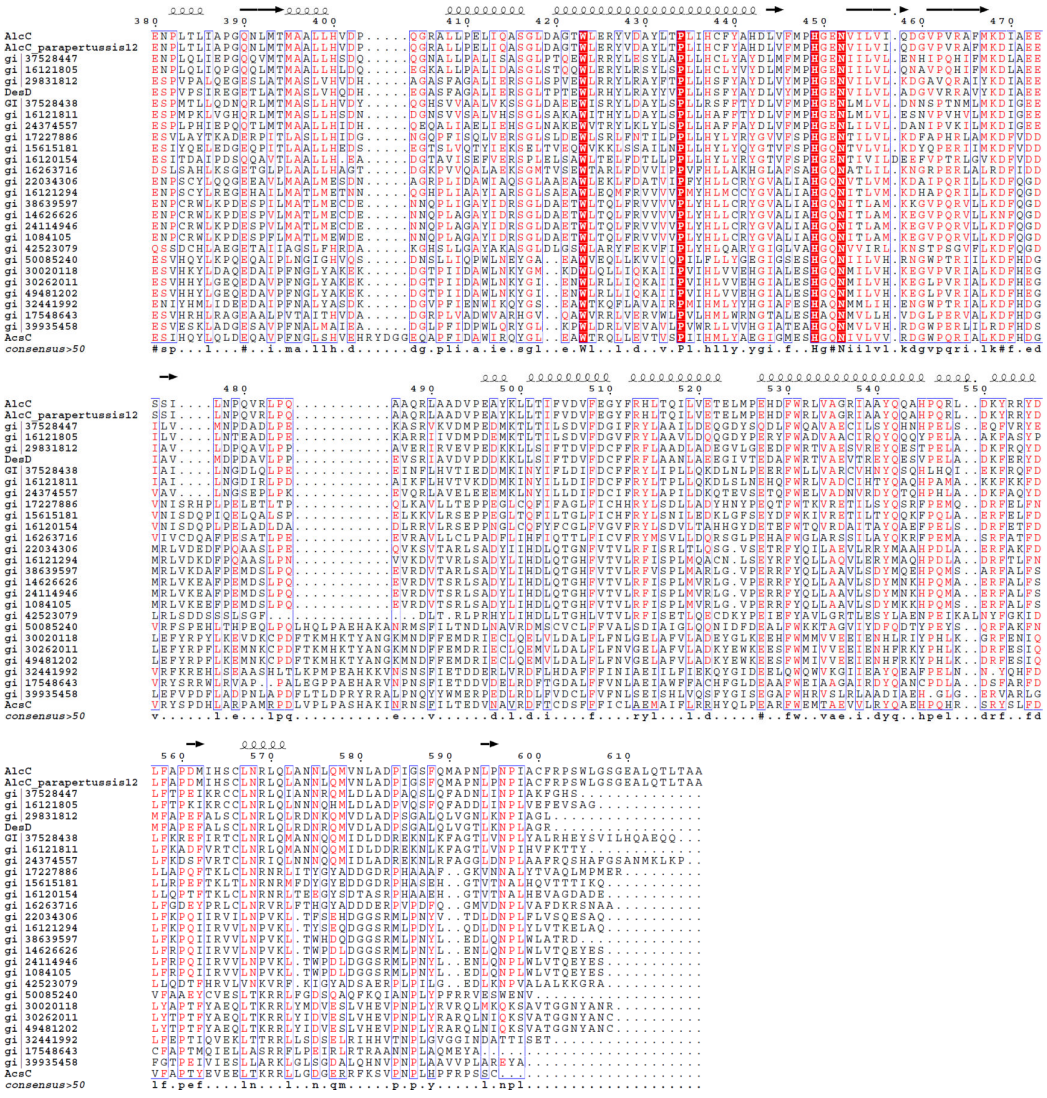
Table 4.5: Sequence identity and similarities of NIS synthetases in relation to AlcC

Family	GI number	Protein	Identity	Similarity	Organism
A	**16118732**	**AcsD**	**24.0**	**39.0**	**Pectobacterium chrysanthemi**
A	42523081	aerobactin siderophore biosynthesis protein iucA	20.0	37.0	Bdellovibrio bacteriovorus HD100
A	38453831	aerobactin synthetase	23.0	40.0	Vibrio mimicus
A	17227890	hypothetical protein all0394	22.0	40.0	Nostoc sp. PCC 7120
A	15615185	hypothetical protein BH2622	21.0	38.0	hypothetical protein [Bacillus halodurans C-125
A	38639599	hypothetical protein LV112	22.0	36.0	Klebsiella pneumoniae
A	21224145	hypothetical protein SCO5800	20.0	34.0	Streptomyces coelicolor A3(2)
A	21224549	hypothetical protein SCO6227	24.0	39.0	Streptomyces coelicolor A3(2)
A	29833864	iron transport protein	26.0	43.0	Streptomyces avermitilis MA-4680
A	29828545	iron transport protein	24.0	35.0	Streptomyces avermitilis MA-4680
A	29829008	iron transport protein	22.0	35.0	Streptomyces avermitilis MA-4680
A	21243906	iron transporter	21.0	36.0	Xanthomonas axonopodis pv. citri str. 306
A	21232484	iron transporter	21.0	35.0	Xanthomonas campestris pv. campestris str. ATCC 33913
A	22034304	IucA	23.0	39.0	Escherichia fergusonii
A	14626624	IucA	21.0	36.0	Shigella boydii
A	26109870	IucA protein	21.0	36.0	Escherichia coli CFT073
A	47569839	MW0093, putative	21.0	38.0	Bacillus cereus G9241
A	38637787	putative iron transport protein	25.0	41.0	Ralstonia eutropha H16
A	49484997	putative siderophore biosynthesis protein	22.0	39.0	Staphylococcus aureus subsp. aureus MSSA476
A	50035243	putative siderophore biosynthesis protein	18.0	39.0	Acinetobacter sp. ADP1
A	16261713	RhbC rhizobactin biosynthesis protein	23.0	38.0	Sinorhizobium meliloti 1021
A	39935459	rhizobactin siderophore biosynthesis protein rhbC	22.0	36.0	Rhodopseudomonas palustris CGA009
A	17548642	siderophore biosynthesis protein	23.0	41.0	Ralstonia solanacearum GMI1000
A	49481200	siderophore biosynthesis protein	22.0	39.0	Bacillus thuringiensis serovar konkukian str. 97-27
A	30020117	siderophore biosynthesis protein	22.0	39.0	Bacillus cereus ATCC 14579
A	24114944	siderophore biosynthesis protein	21.0	36.0	Shigella flexneri 2a str. 301
A	30262010	siderophore biosynthesis protein, putative	21.0	38.0	Bacillus anthracis str. Ames
A	23307118	vibrioferrin biosynthesis protein PvsD	22.0	40.0	Vibrio parahaemolyticus
B	**16118729**	**AcsA**	**22.0**	**39.0**	**Pectobacterium chrysanthemi**
B	50085239	hypothetical protein ACIAD2120	21.0	42.0	Acinetobacter sp. ADP1
B	21224548	hypothetical protein SCO6226	21.0	36.0	Streptomyces coelicolor A3(2)
B	27468688	hypothetical protein SE1770	21.0	42.0	Staphylococcus epidermidis ATCC 12228
B	27468690	hypothetical protein SE1772	20.0	39.0	Staphylococcus epidermidis ATCC 12228
B	23307116	ibrioferrin biosynthesis protein PvsD	22.0	40.0	Vibrio parahaemolyticus
B	50123036	putative siderophore biosynthesis protein	25.0	42.0	Pectobacterium atrosepticum
B	38637789	putative siderophore biosynthesis protein	23.0	39.0	Ralstonia eutropha H16
B	38233195d2	putative siderophore biosynthesis related protein	25.0	40.0	Corynebacterium diphtheriae NCTC 13129
B	38233195d1	putative siderophore biosynthesis related protein	25.0	40.0	Corynebacterium diphtheriae NCTC 13129
B	36958857	Siderophore biosynthesis protein	22.0	40.0	Rhizobium sp. NGR234
B	32441989	siderophore biosynthesis protein SbnC	24.0	41.0	Staphylococcus aureus
B	19749317	Y4xN	22.0	39.0	Sinorhizobium fredii
B	16520030	Y4xN	22.0	40.0	Rhizobium sp. NGR234
C	**33593441**	**AlcC**	**100.0**	**100.0**	**B. bronchiseptica Tohama I**
C	33597963	AlcC	99.0	99.0	B. parapertussis 12822
C	29831812	siderophore biosynthetic enzyme	52.0	70.0	Streptomyces avermitilis MA-4680
C	37528447	hypothetical protein plu4630	52.0	69.0	Photorhabdus luminescens subsp. laumondii TTO1
C	**21221236**	**hypothetical protein SCO2785, DesD**	**51.0**	**67.0**	**Streptomyces coelicolor A3(2)**
C	37528438	hypothetical protein plu4621	48.0	67.0	Photorhabdus luminescens subsp. laumondii TTO1
C	24374557	siderophore biosynthesis protein, putative	49.0	65.0	Shewanella oneidensis
C	17227886	hypothetical protein all0390	36.0	54.0	Nostoc sp. PCC 7120
C	15615181	hypothetical protein BH2618	31.0	50.0	Bacillus halodurans C-125
C	22034306	IucC	28.0	50.0	E.coli
C	16120154	IucC	32.0	48.0	Halobacterium sp. NRC-1
C	38639597	IucC	29.0	48.0	Klebsiella pneumoniae
C	14626626	IucC	28.0	48.0	Shigella boydii
C	24114946	siderophore biosynthesis protein	28.0	48.0	Shigella flexneri 2a str. 301
C	50085240	hypothetical protein ACIAD2121	26.0	48.0	Acinetobacter sp. ADP1
C	16263716	RhbF rhizobactin siderophore biosynthesis protein	29.0	47.0	Sinorhizobium
C	16118727	AcsC	28.0	46.0	Pectobacterium chrysanthemi
C	42523079	putative siderophore biosynthesis protein IucC	26.0	46.0	Bdellovibrio bacteriovorus HD100
C	32441992	SbnF	26.0	46.0	Staphylococcus aureus
C	17548643	siderophore biosynthesis protein	28.0	44.0	Ralstonia solanacearum GMI1000
C	39935458	rhizobactin siderophore biosynthesis protein RhsF	26.0	44.0	Rhodopseudomonas palustris CGA009
C	30020118	siderophore biosynthesis protein	24.0	44.0	Bacillus cereus ATCC 14579

D. Sequence alignment of NIS-family C



part 2



E. Analysis of dimer formation and macrocyclization

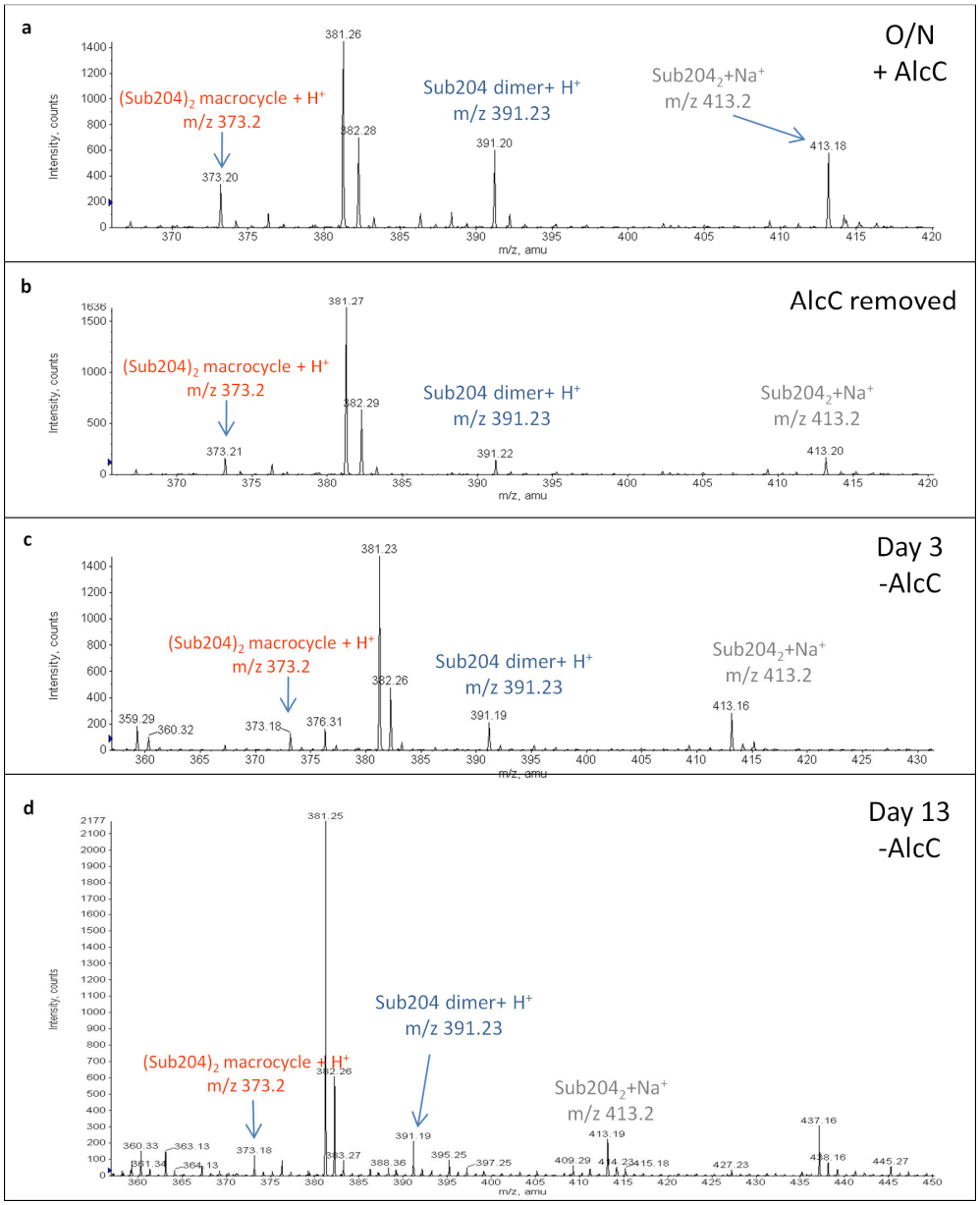


Figure 4.8: Analysis of dimer formation and macrocyclization. **(a)** O/N reactions of AlcC with Sub204 show peaks referring to dimeric Sub204 as well as $(Sub204)_2$ macrocycle. **(b)** Removing AlcC enzyme by filtration still shows the same result as in (a). **(c)** Stopped enzymatic reactions where AlcC was removed showed no significant change in dimer to macrocycle ratio after further incubation of three (c) or thirteen days (d). This indicates that the macrocyclization is enzyme catalyzed and does not occur autocatalytically.

F. Work on AcsA

F.1. Introduction

AcsA is a type B NIS enzyme involved in achromobactin biosynthesis. It performs the final production step of achromobactin by attaching two molecules of α-KG on each site of the achromobactin precursor (Figure 1.10). Alike AlcC, AcsA is proposed to be a processive enzyme. Here the initial work on expression, purification and obtaining AcsA crystals is summarized.

F.2. Results and discussion

F.2.1. Protein expression and purification

AcsA DNA (provided by Prof. Challis, Warwick) was transformed (1 min at 42 °C) into BL21(DE3) competent cells for protein expression. Cells were incubated for 1.5 h at 37 °C in 1 mL LB media and then plated on ampicillin agar plates O/N.

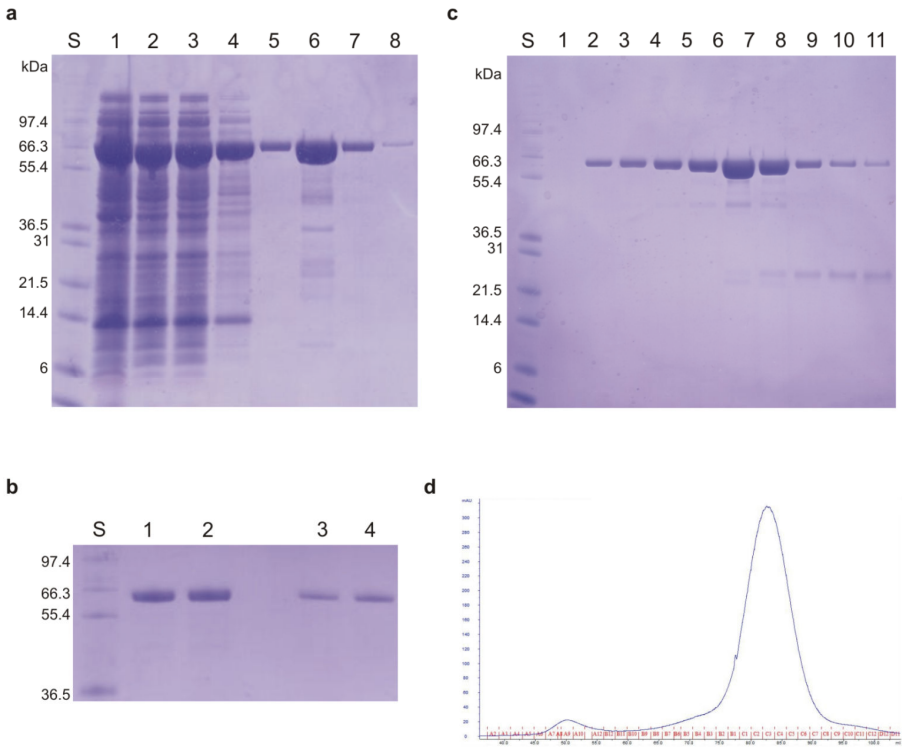


Figure F.1: Protein purification of AcsA. **(a)** IMAC purification of AcsA. **S**: standard; **1**: crude; **2**: S/N; **3**: flow through; **4**: wash 1: **5**: wash 2: **6**: elution 1; **7:** elution 2; **8:** elution 2. **(b)** TEV-cleavage with 1:10 TEV protease. TEV protease treatment was not successful. **1**: AcsA E1; **2**: AcsA E1 with TEV protease; **3**: AcsA E2; **4**: AcsA E2 with TEV protease. **(c)** SDS-PAGE of GF fractions. **S**: standard; **1**: A9; **2**: B4; **3**: B3; **4**: B3: **5**: B2: **6**: B1; **7:** C3; **8:** C5; **9**: C7; **10:** C8; **11:** C9. Fraction B4-C5 were combined and used in crystallization trails. **(d)** AcsA elution profile of GF.

Protein expression was carried out as described for AcsD (see chapter2). Cells were lysed by sonication (10 times for 30 s with 1 min breaks on ice). His_6-AcsA was purified via Ni-IMAC and eluted protein desalted (Figure F.1a). TEV cleavage with TEV protease at RT O/N was unsuccessful (Figure F.1b). DNA sequencing showed that the AcsA plasmid codes for a TEV site. His_6-AcsA protein was further purified by gel filtration as described for AcsD (Figure F.1c,d). Fractions B4-C5 were combined, concentrated to 6 mg/mL and used for crystallization trails.

F.2.2. Screening for AcsA crystallization conditions

Initial screening was performed with 6 mg/mL and 3 mg/ml AcsA protein, with and without 5-20 mM ATP, using commercial screens (Wizard, Index, Cryo, Crystal Screen I+II). Prior to droplet setup, the protein was incubated for 5 min with ATP and the precipitant was removed by centrifugation (5 min, 20'000 g, 4 °C). Additionally, a custom made stochastic screen (see Table F.1) was used for screening. Several conditions resulted in likely micro crystals or crystal spheroids:

1. Index screen (micro crystals)

well 15 (0.1 M HEPES pH 7.5m 0.5 M Mg formate dihydrate)

well 32 (1 M $(NH_4)SO_4$ 0.1 M BIS-TRIS pH 5.5, 1 % (w/v) polyethylene glycol 3350);

2. Wizard I+II with 5 mM ATP (micro crystals)

well 7 (10 % (w/v) PEG 8k, 0.1 M MES pH 6.0, 0.2 M $Zn(OAc)_2$)

well 55 (30 % (w/v) PEG 3k, 0.1 M TRIS pH 7.0, 0.2 M NaCl)

3. Stochastic screen (spheroids)

well C4 (0.053 M $MgCl_2$, 19.47 % (w/v) PEG MME 2k, 0.1 M NaCac pH 6.9)

4. Stochastic screen with 5 mM ATP (spheroids)

well A7 (0.091 mM $MnCl_2$, 11.9 % (w/v) PEG MME 5k, 0.1 M MOPS pH 7.2)

well B5 (0.082 mM $MnCl_2$, 24.8 % (w/v) PEG 8k, 0.1 M HEPES pH 7.5)

For optimization two new custom stochastic screens (Opt1 and Opt2) were prepared (Table F.2a, b). Crystal trails were set up with and without substrate (5 mM ATP, 5 mM α-KG or 5 mM ATP + 5 mM α-KG) with a final droplet size of 2 µL (1 µL AlcC protein + 1 µL precipitant). Protein concentrations were chosen as described above.

For condition 55 of the Wizard screen (30 % (w/v) PEG 3k, 0.1 M TRIS pH 7.0, 0.2 M NaCl) additional grid screens with TRIS and HEPES buffer, and different protein concentrations were prepared (Table F.3). Condition C1 (17 mM KCl, 28 % (w/v) PEG 3k, 0.1 M MES pH 6.0) of the Opt1 stochastic screen resulted in small crystals, which were further optimized in a new grid screen (Table F.4).

F.2.3. Identification of protein crystals

Likely crystals were obtained in following conditions:

1. 6 mg/mL AcsA with 20 mM ATP
7.9 % (v/v) MPD, 192 mM $CaCl_2$, 25.8 % (w/v) PEG 8K, 0.1 M HEPES pH7.5
- Not single and no good shape

2. 2.9 mg/mL AcsA with 5 mM ATP
189 mM $CaCl_2$, 13.8 % (w/v) PEG MME 2K, 0.1 M NaCac pH6.9, 18 mM ATP
- Two nice, small crystals

3. 2.9 mg/mL AcsA and no ATP
14.7 % (w/v) PEG 8K, 0.1 M HEPES pH7.8, 49 mM ATP
- Cluster of crystals

4. 2.9 mg/mL AcsA and 5 mM ATP
15.8 % (w/v) PEG MME 2K, 0.1 M NaCac pH6.7, 29 mM ATP
- Cluster of crystals

Most of the crystals were small or clustered and not suitable for data collection, but were tested to check their quality. Crystals were cryo-protected with 20 % PEG 400 or 20-25 % (v/v) glycerol and flash frozen in liquid nitrogen. Crystals did not diffract or showed no typical protein diffraction pattern as the example of condition 3 shows (Figure F.2). Unfortunately, the nice crystals found in condition 2 (2.9 mg/mL AcsA with 5 mM ATP) were identified as salt crystals. No further crystallization screens or optimizations have been carried out.

Table F1: Stochastic Screen used for inital screening of AcsA crystallization condition

A01 CaCl2 : 0.131 M PEG MME 2k : 15.422 % MOPS : pH 7.2 0.1 M Glycerol : 5.369 %	**A02** LiCl2 : 0.058 M PEG 8k : 26.79 % HEPES : pH 7.5 0.1 M ATP : 0.026 M	**A03** LiCl2 : 0.14 M PEG 2k : 21.527 % NaCac : pH 6.9 0.1 M	**A04** MgCl2 : 0.053 M PEG MME 5k : 13.369 % MOPS : pH 7.2 0.1 M Glycerol : 9.948 %	**A05** LiCl2 : 0.126 M PEG 8k : 17.264 % MOPS : pH 7.2 0.1 M	**A06** PEG MME 5k : 22.211 % MOPS : pH 7.2 0.1 M	**A07** MNCl2 2: 0.091 M PEG MME 5k : 11.895 % MOPS : pH 7.2 0.1 M	**A08** PEG MME 5k : 25.895 % HEPES : pH 7.2 0.1 M	**A09** LiCl2 : 0.091 M PEG MME 5k : 17.211 % HEPES : pH 7.8 0.1 M	**A10** MgCl2 : 0.103 M PEG MME 5k : 13.106 % NaCac : pH 6.7 0.1 M	**A11** LiCl2 : 0.076 M PEG 2k : 13.053 % HEPES : pH 7.5 0.1 M MPD : 1.601 %	**A12** LiCl2 : 0.052 M PEG 8k : 10.632 % HEPES : pH 7.8 0.1 M
B01 MNCl2 2: 0.107 M PEG MME 5k : 13.579 % HEPES : pH 7.5 0.1 M	**B02** MNCl2 2: 0.162 M PEG MME 2k : 24.211 % NaCac : pH 6.9 0.1 M	**B03** CaCl2 : 0.149 M PEG MME 2k : 15.264 % MOPS : pH 6.5 0.1 M MPD : 3.032 %	**B04** PEG 8k : 21.158 % HEPES : pH 7.8 0.1 M ATP : 0.03 M	**B05** MNCl2 2: 0.082 M PEG 8k : 24.79 % HEPES : pH 7.5 0.1 M	**B06** CaCl2 : 0.181 M PEG 8k : 24.685 % HEPES : pH 7.8 0.1 M Glycerol : 10.422 %	**B07** LiCl2 : 0.154 M PEG 2k : 21.843 % MOPS : pH 6.5 0.1 M Glycerol : 11.053 %	**B08** MgCl2 : 0.069 M PEG MME 5k : 10.106 % HEPES : pH 7.5 0.1 M	**B09** ATP : 0.042 M PEG MME 5k : 12.264 % HEPES : pH 7.8 0.1 M ATP : 0.012 M	**B10** LiCl2 : 0.14 M PEG 8k : 27.264 % MOPS : pH 6.5 0.1 M	**B11** CaCl2 : 0.076 M PEG 8k : 29.264 % NaCac : pH 6.9 0.1 M	**B12** CaCl2 : 0.173 M PEG MME 5k : 13.369 % HEPES : pH 7.5 0.1 M
C01 CaCl2 : 0.171 M PEG MME 5k : 25.106 % MOPS : pH 7.2 0.1 M	**C02** CaCl2 : 0.122 M PEG MME 5k : 10.737 % TRIS : pH 7.7 0.1 M	**C03** MgCl2 : 0.181 M PEG MME 2k : 24.106 % MOPS : pH 6.5 0.1 M Glycerol : 5.132 %	**C04** MgCl2 : 0.053 M PEG MME 2k : 19.474 % NaCac : pH 6.9 0.1 M	**C05** MgCl2 : 0.164 M PEG MME 5k : 21.843 % NaCac : pH 6.9 0.1 M	**C06** MNCl2 2: 0.122 M PEG MME 5k : 14.895 % HEPES : pH 7.5 0.1 M MPD : 3.453 %	**C07** PEG MME 5k : 14.422 % TRIS : pH 7.7 0.1 M	**C08** PEG MME 2k : 15.79 % NaCac : pH 6.7 0.1 M ATP : 0.029 M	**C09** PEG MME 2k : 14.264 % HEPES : pH 7.5 0.1 M	**C10** PEG MME 5k : 14.053 % HEPES : pH 7.2 0.1 M ATP : 0.038 M	**C11** PEG MME 5k : 13.264 % MOPS : pH 7.2 0.1 M ATP : 0.029 M	**C12** ATP : 0.042 M PEG 8k : 20.264 % MOPS : pH 6.5 0.1 M Glycerol : 5.527 %
D01 ATP : 0.044 M PEG MME 5k : 22.474 % HEPES : pH 7.8 0.1 M Glycerol : 10.185 %	**D02** MNCl2 2: 0.08 M PEG MME 5k : 19.264 % HEPES : pH 7.2 0.1 M Glycerol : 6.474 %	**D03** CaCl2 : 0.147 M PEG MME 2k : 21.316 % MOPS : pH 7.2 0.1 M ATP : 0.013 M	**D04** CaCl2 : 0.107 M PEG MME 2k : 18.737 % HEPES : pH 7.8 0.1 M ATP : 0.025 M	**D05** MNCl2 2: 0.086 M PEG MME 2k : 23.158 % HEPES : pH 7.5 0.1 M MPD : 2.443 %	**D06** MgCl2 : 0.145 M PEG MME 5k : 23.843 % NaCac : pH 6.9 0.1 M Glycerol : 6.237 %	**D07** LiCl2 : 0.12 M PEG MME 5k : 28.527 % HEPES : pH 7.5 0.1 M MPD : 1.685 %	**D08** ATP : 0.034 M PEG 400 : 33.769 % HEPES : pH 7.5 0.1 M Glycerol : 6.158 %	**D09** PEG 2k : 24.316 % TRIS : pH 7.7 0.1 M ATP : 0.017 M	**D10** LiCl2 : 0.115 M PEG 2k : 13.843 % MOPS : pH 7.2 0.1 M MPD : 7.411 %	**D11** LiCl2 : 0.122 M PEG MME 5k : 19.474 % NaCac : pH 6.9 0.1 M ATP : 0.047 M	**D12** PEG 8k : 18.0 % NaCac : pH 6.9 0.1 M ATP : 0.035 M
E01 MgCl2 : 0.112 M PEG MME 5k : 16.737 % NaCac : pH 6.7 0.1 M	**E02** CaCl2 : 0.194 M PEG 8k : 19.053 % MOPS : pH 7.2 0.1 M Glycerol : 10.737 %	**E03** ATP : 0.046 M PEG 400 : 25.601 % HEPES : pH 7.5 0.1 M MPD : 7.158 %	**E04** LiCl2 : 0.184 M PEG MME 5k : 11.053 % HEPES : pH 7.8 0.1 M	**E05** CaCl2 : 0.074 M PEG 400 : 23.495 % TRIS : pH 7.7 0.1 M	**E06** PEG 2k : 12.948 % HEPES : pH 7.8 0.1 M MPD : 10.19 %	**E07** PEG MME 2k : 22.211 % HEPES : pH 7.5 0.1 M	**E08** MgCl2 : 0.128 M PEG 8k : 28.106 % NaCac : pH 6.9 0.1 M MPD : 3.706 %	**E09** CaCl2 : 0.189 M PEG MME 2k : 13.79 % NaCac : pH 6.9 0.1 M ATP : 0.018 M	**E10** PEG 2k : 22.579 % HEPES : pH 7.8 0.1 M ATP : 0.035 M	**E11** MNCl2 2: 0.154 M PEG MME 2k : 22.474 % HEPES : pH 7.5 0.1 M	**E12** PEG 8k : 24.211 % MOPS : pH 7.2 0.1 M Glycerol : 7.264 %
F01 LiCl2 : 0.193 M PEG MME 5k : 17.0 % HEPES : pH 7.5 0.1 M ATP : 0.035 M	**F02** MgCl2 : 0.145 M PEG 8k : 15.158 % HEPES : pH 7.5 0.1 M Glycerol : 9.474 %	**F03** CaCl2 : 0.192 M PEG 8k : 25.843 % HEPES : pH 7.5 0.1 M MPD : 7.916 %	**F04** MNCl2 2: 0.074 M PEG 8k : 13.053 % HEPES : pH 7.2 0.1 M ATP : 0.014 M	**F05** ATP : 0.048 M PEG 2k : 12.422 % HEPES : pH 7.8 0.1 M Glycerol : 7.737 %	**F06** LiCl2 : 0.18 M PEG MME 5k : 29.053 % HEPES : pH 7.8 0.1 M Glycerol : 7.737 %	**F07** LiCl2 : 0.101 M PEG MME 2k : 20.106 % HEPES : pH 7.8 0.1 M None : 0.0 M	**F08** LiCl2 : 0.168 M PEG 400 : 27.369 % HEPES : pH 7.8 0.1 M Glycerol : 7.737 %	**F09** LiCl2 : 0.067 M PEG MME 2k : 24.316 % NaCac : pH 6.7 0.1 M None : 0.0 M	**F10** MgCl2 : 0.084 M PEG 2k : 11.843 % TRIS : pH 7.7 0.1 M None : 0.0 M	**F11** MNCl2 2: 0.189 M PEG 8k : 27.474 % NaCac : pH 6.9 0.1 M	**F12** MNCl2 2: 0.194 M PEG MME 2k : 18.158 % HEPES : pH 7.5 0.1 M
G01 ATP : 0.023 M PEG 8k : 15.422 % HEPES : pH 7.5 0.1 M MPD : 5.306 %	**G02** None : 0.0 M PEG MME 5k : 13.843 % HEPES : pH 7.5 0.1 M	**G03** MgCl2 : 0.192 M PEG 2k : 15.79 % HEPES : pH 7.8 0.1 M	**G04** None : 0.0 M PEG 8k : 14.527 % HEPES : pH 7.8 0.1 M Glycerol : 6.316 %	**G05** MgCl2 : 0.16 M PEG 8k : 13.211 % HEPES : pH 7.5 0.1 M Glycerol : 11.448 %	**G06** None : 0.0 M PEG 8k : 14.685 % HEPES : pH 7.8 0.1 M ATP : 0.049 M	**G07** MNCl2 2: 0.128 M PEG 8k : 22.579 % MOPS : pH 6.5 0.1 M None : 0.0 M	**G08** LiCl2 : 0.193 M PEG 8k : 18.0 % MOPS : pH 6.5 0.1 M	**G09** MNCl2 2: 0.152 M PEG 8k : 28.106 % HEPES : pH 7.5 0.1 M ATP : 0.04 M	**G10** MNCl2 2: 0.173 M PEG MME 5k : 12.0 % HEPES : pH 7.8 0.1 M	**G11** CaCl2 : 0.12 M PEG 400 : 24.927 % MOPS : pH 6.5 0.1 M	**G12** MgCl2 : 0.181 M PEG MME 5k : 29.632 % MOPS : pH 7.2 0.1 M MPD : 3.622 %
H01 MNCl2 2: 0.177 M PEG 400 : 35.874 % NaCac : pH 6.9 0.1 M Glycerol : 10.658 %	**H02** MNCl2 2: 0.185 M PEG MME 5k : 25.685 % HEPES : pH 7.8 0.1 M ATP : 0.032 M	**H03** LiCl2 : 0.154 M PEG 2k : 19.474 % HEPES : pH 7.8 0.1 M ATP : 0.038 M	**H04** LiCl2 : 0.155 M PEG 8k : 21.843 % HEPES : pH 7.8 0.1 M	**H05** LiCl2 : 0.129 M PEG MME 5k : 12.106 % TRIS : pH 7.7 0.1 M	**H06** LiCl2 : 0.109 M PEG MME 5k : 26.895 % HEPES : pH 7.8 0.1 M Glycerol : 5.211 %	**H07** CaCl2 : 0.12 M PEG 400 : 24.085 % MOPS : pH 6.5 0.1 M MPD : 3.369 %	**H08** None : 0.0 M PEG 8k : 26.106 % MOPS : pH 7.2 0.1 M	**H09** LiCl2 : 0.1 M PEG MME 5k : 20.369 % MOPS : pH 7.2 0.1 M Glycerol : 9.237 %	**H10** None : 0.0 M PEG 8k : 11.79 % HEPES : pH 7.8 0.1 M	**H11** MgCl2 : 0.154 M PEG MME 5k : 13.843 % NaCac : pH 6.9 0.1 M	**H12** MgCl2 : 0.152 M PEG 8k : 13.737 % TRIS : pH 7.7 0.1 M

Table F.2: Stochastic optimization screen Opt1 and Opt2. Conditions highlighted in red led to salt precipitate.

a

A1 MgCl2 : 0.098 M PEG MME 2k : 26.101 % HEPES : pH 7.2 0.1 M a-KG pH7.4 : 0.013 M	**A2** ZnAcet : 0.078 M PEG 3350 : 27.101 % TRIS : pH 7.5 0.1 M	**A3** ZnAcet : 0.025 M PEG MME 2k : 28.6 % NaCac : pH 6.9 0.1 M a-KG pH7.4 : 0.011 M	**A4** MgCl2 : 0.055 M PEG 4000 : 33.551 % MES : pH 6.5 0.1 M	**A5** PEG MME 2k : 17.2 % HEPES : pH 7.8 0.1 M	**A6** PEG 3350 : 32.05 % NaCac : pH 6.5 0.1 M a-KG pH7.4 : 0.015 M
B1 PEG 3000 : 21.7 % TRIS : pH 8.0 0.1 M	**B2** MgCl2 : 0.022 M PEG 3000 : 31.95 % NaCac : pH 6.9 0.1 M	**B3** PEG 3000 : 41.9 % HEPES : pH 7.8 0.1 M ATP pH7.4 : 0.015 M	**B4** MgCl2 : 0.045 M PEG MME 2k : 20.95 % TRIS : pH 8.0 0.1 M	**B5** LiCl2 : 0.047 M PEG MME 5k : 31.961 % MOPS : pH 7.2 0.1 M ATP pH7.4 : 0.013 M	**B6** ZnAcet : 0.04 M PEG MME 2k : 28.4 % MES : pH 6.0 0.1 M a-KG pH7.4 : 0.015 M
C1 KCl : 0.017 M PEG 3000 : 28.3 % MES : pH 6.0 0.1 M	**C2** PEG 4000 : 33.101 % MES : pH 6.0 0.1 M	**C3** PEG 8k : 16.321 % NaCac : pH 6.5 0.1 M	**C4** LiCl2 : 0.088 M PEG 8k : 22.401 % TRIS : pH 8.5 0.1 M	**C5** MgCl2 : 0.063 M PEG 3350 : 31.301 % MES : pH 6.0 0.1 M a-KG pH7.4 : 0.011 M	**C6** MgCl2 : 0.036 M PEG 3350 : 23.8 % MOPS : pH 6.7 0.1 M Glycerol : 10.501 %
D1 PEG MME 2k : 17.0 % HEPES : pH 7.8 0.1 M Glycerol : 9.225 %	**D2** PEG MME 2k : 16.901 % HEPES : pH 7.2 0.1 M a-KG pH7.4 : 0.013 M	**D3** PEG 8k : 26.241 % HEPES : pH 7.8 0.1 M	**D4** KCl : 0.115 M PEG MME 2k : 23.401 % HEPES : pH 7.2 0.1 M ATP pH7.4 : 0.013 M	**D5** LiCl2 : 0.088 M PEG 4000 : 21.15 % HEPES : pH 7.5 0.1 M ATP pH7.4 : 0.013 M	**D6** ZnAcet : 0.056 M PEG 8k : 17.601 % TRIS : pH 7.5 0.1 M

b

A1 LiCl2 : 0.068 M PEG 3350 : 33.051 % TRIS : pH 7.5 0.1 M ATP pH7.4 : 0.015 M	**A2** KCl : 0.031 M PEG MME 5k : 30.601 % NaCac : pH 6.9 0.1 M a-KG pH7.4 : 0.013 M	**A3** ZnAcet : 0.05 M PEG 8k : 22.52 % MOPS : pH 6.7 0.1 M a-KG pH7.4 : 0.013 M	**A4** ZnAcet : 0.042 M PEG 8k : 29.801 % Tris : pH 7.5 0.1 M	**A5** PEG MME 2k : 25.0 % TRIS : pH 8.0 0.1 M	**A6** MgCl : 0.036 M PEG 8k : 10.161 % MOPS : pH 7.0 0.1 M ATP pH7.4 : 0.013 M
B1 KCl : 0.03 M PEG 8k : 15.761 % NaCac : pH 6.9 0.1 M	**B2** PEG MME 2k : 29.8 % TRIS : pH 8.0 0.1 M	**B3** ZnAcet : 0.038 M PEG 8k : 11.201 % MES : pH 6.0 0.1 M	**B4** PEG 3350 : 21.25 % NaCac : pH 6.9 0.1 M Glycerol : 12.751 %	**B5** KCl : 0.048 M PEG 3000 : 26.5 % HEPES : pH 7.5 0.1 M	**B6** LiCl2 : 0.028 M PEG MME 5k : 28.2 % MOPS : pH 6.7 0.1 M Glycerol : 11.175 %
C1 MgCl : 0.045 M PEG 3000 : 22.101 % MES : pH 6.0 0.1 M ATP pH7.4 : 0.013 M	**C2** LiCl2 : 0.059 M PEG 4000 : 12.951 % HEPES : pH 7.2 0.1 M	**C3** MgCl : 0.082 M PEG 3000 : 30.301 % TRIS : pH 7.5 0.1 M	**C4** PEG 4000 : 21.15 % NaCac : pH 6.5 0.1 M	**C5** KCl : 0.07 M PEG 4000 : 30.15 % HEPES : pH 7.5 0.1 M	**C6** LiCl2 : 0.049 M PEG MME 5k : 32.08 % TRIS : pH 8.5 0.1 M ATP pH7.4 : 0.015 M
D1 ZnAcet : 0.076 M PEG MME 5k : 26.36 % NaCac : pH 6.5 0.1 M ATP pH7.4 : 0.013 M	**D2** ZnAcet : 0.09 M PEG MME 2k : 22.651 % NaCac : pH 6.5 0.1 M ATP pH7.4 : 0.013 M	**D3** MgCl : 0.014 M PEG 8k : 29.24 % NaCac : pH 6.9 0.1 M a-KG pH7.4 : 0.015 M	**D4** LiCl2 : 0.047 M PEG MME 5k : 31.0 % TRIS : pH 7.5 0.1 M Glycerol : 5.776 %	**D5** MnCl2 : 0.016 M PEG 3350 : 22.0 % MOPS : pH 7.0 0.1 M a-KG pH7.4 : 0.015 M	**D6** PEG MME 5k : 27.361 % MES : pH 6.5 0.1 M

Table F.3: Grid screen for condition 55 of Wizard I with TRIS and HEPES buffer.

25.0 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.0	27.5 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.0	30.0 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.0	32.5 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.0	35.0 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.0	37.5 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.0
25.0 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.5	27.5 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.5	30.0 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.5	32.5 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.5	35.0 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.5	37.5 % PEG3K 0.1 M NaCl2 0.1 M Tris pH 7.5
25.0 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.2	27.5 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.2	30.0 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.2	32.5 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.2	35.0 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.2	37.5 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.2
25.0 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.8	27.5 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.8	30.0 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.8	32.5 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.8	35.0 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.8	37.5 % PEG3K 0.1 M NaCl2 0.1 M HEPES pH 7.8

Site I: AcsA no ATP (6 + 3 mg/mL (small well))
Site II: AcsA (6 + 3 mg/mL (small well)) + 5 mM ATP
Site II: AcsA (6 + 3 mg/mL (small well)) + 5 mM ATP/ 5 mM α-KG

Table F.4: Grid screen II for condition C5 of the Opt1 screen

25 % PEG3K 0 M KCl 0.1 M MES	25 % PEG3K 0.017 M KCl 0.1 M MES	25 % PEG3K 0.050 M KCl 0.1 M MES	25 % PEG3K 0 M KCl 0.1 M MES	25 % PEG3K 0.017 M KCl 0.1 M MES	25 % PEG3K 0.050 M KCl 0.1 M MES
28 % PEG3K 0 M KCl 0.1 M MES	28 % PEG3K 0.017 M KCl 0.1 M MES	28 % PEG3K 0.050 M KCl 0.1 M MES	28 % PEG3K 0 M KCl 0.1 M MES	28 % PEG3K 0.017 M KCl 0.1 M MES	28 % PEG3K 0.050 M KCl 0.1 M MES
31 % PEG3K 0 M KCl 0.1 M MES	31 % PEG3K 0.017 M KCl 0.1 M MES	31 % PEG3K 0.050 M KCl 0.1 M MES	31 % PEG3K 0 M KCl 0.1 M MES	31 % PEG3K 0.017 M KCl 0.1 M MES	31 % PEG3K 0.050 M KCl 0.1 M MES
34 % PEG3K 0 M KCl 0.1 M MES	34 % PEG3K 0.017 M KCl 0.1 M MES	34 % PEG3K 0.050 M KCl 0.1 M MES	34 % PEG3K 0 M KCl 0.1 M MES	34 % PEG3K 0.017 M KCl 0.1 M MES	34 % PEG3K 0.050 M KCl 0.1 M MES

MES buffer: column 1-3 pH 6.0; column 4-6 pH 6.5
Site I: AcsA (6 .4 mg/mL)
Site II: AcsA (6 .0 mg/mL)
Site II: AcsA (3 mg/mL)
drop size 1.5+1.5µL

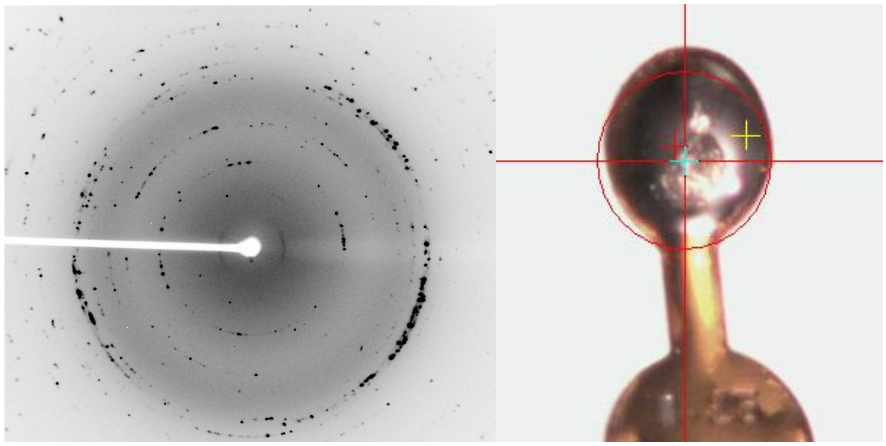


Figure F.2: Diffraction pattern of crystal 3 grown in 14.7 % (w/v) PEG 8k, 0.1 M HEPES pH 7.8, 49 mM ATP with 2.9 mg/mL AcsA in a 0.3 µL drop (150 + 150 nL).

F.3. Conclusion and future work

It was not possible to obtain good initial crystals. All crystals tested were not protein crystals or did not diffract when tested. Some of the crystals were identified as salt crystals. This His_6-tag may be one reason for lack of a good initial hit. This flexible series of amino acids may have interfered with crystallization. The His_6-tag could not be removed by TEV digestion, although DNA sequencing showed that a TEV-cleaving site is present. Hence a new vector with a longer spacer between TEV recognition site and the N-terminus may help. Construction of a C-terminal His_6-tag fusion (whether cleavable or not) should also be considered. Since the protein could be handled without problems (no degradation or precipitation) it might still be behave

well in different buffers and at lower salt concentration. This might decrease its solubility and may help in obtaining initial crystal condition. Another possibility would be in-situ proteolysis with chymotrypsin, by adding it in trace amounts to the crystal trails (Dong et al. 2007). Chymotrypsin cleaves at Y and F and is likely to cut off the His_6-tag which has the following sequence: MSYYHHHHHHDYDIPTTENL**Y|F**QG. This would be a cheap and quick method, since the remaining His_6-tagged protein can be used, and may lead to an initial condition.

G. Work on DesD

G.1. Introduction

DesD is the last enzyme in the desferrioxamine pathway of *Streptomyces coelicolor*. It catalyzes the trimerization and macrocyclization of *N*-hydroxy-*N*-succinyl-cadaverine, generating desferrioxamine E (Figure 3.30). As mentioned in Chapter 3 desferrioxamine is a metal complexing agent used in medical applications and a structure of DesD is of great interest. Here the initial work on DesD expression, purification and initial crystallization attempts is reported.

G.2. Results and discussion

G.2.1. Expression and purification of DesD

Expression problems

Expression trails in BL21(DE3) cells resulted in poor yield of full length as well as degraded His_6-DesD. Expression trails at different temperatures showed that less degradation occurs when His_6-DesD was expressed at temperatures lower than 20 °C. Western blot analysis with an anti-His_6 monoclonal antibody confirmed that these fragments result from DesD and are not *E. coli* proteins (Figure G.1). Whether these fragments result from protein degradation or are caused by incomplete biosynthesis within the cell is not yet clear. However, expression of His_6-DesD at 15 °C O/N did not increase the protein yield.

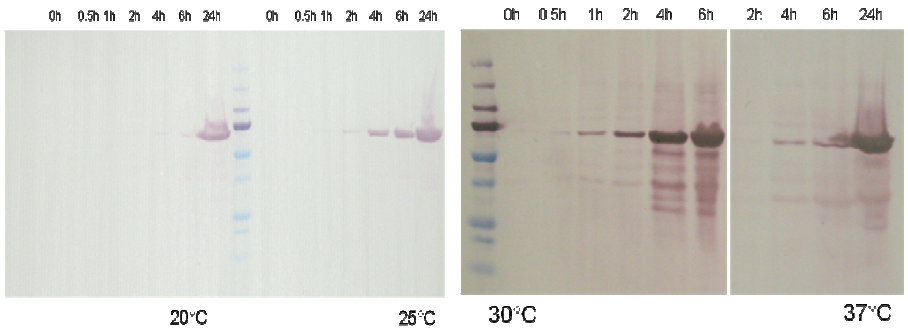


Figure G.1: Expression of DesD in BL21(DE3) competent cells at different temperatures. A Western blot with anti-His_6 monoclonal antibodies showed that expression at higher temperatures (>30 °C) results in more fragments of His_6-DesD and proves that these bands result from DesD (still His_6-tagged) and are not *E. coli* proteins.

Expression in Rosetta (DE3) competent cells

The DesD construct was transformed in Rosetta (DE3) competent cells. 1 L cultures were grown from 10 mL O/N starter cultures containing 30 µg/mL kanamycin and 40 µg/L chloramphenicol. Protein expression was induced at an OD_{600} of 0.8 with 0.5 mM IPTG and incubated for a further 15 h. This resulted in full length protein without degradation and yielded 17 mg protein/1 L LB media. Purification steps were performed as described for AcsD (chapter 2) (Figure G.2a). However TEV-cleavage with TEV protease was not successful. In a final step His_6-DesD was purified on a S200 Superdex gel filtration column (Figure G.2b). Protein was concentrated to 21.4 mg/ml, frozen in liquid nitrogen and stored at -80 °C.

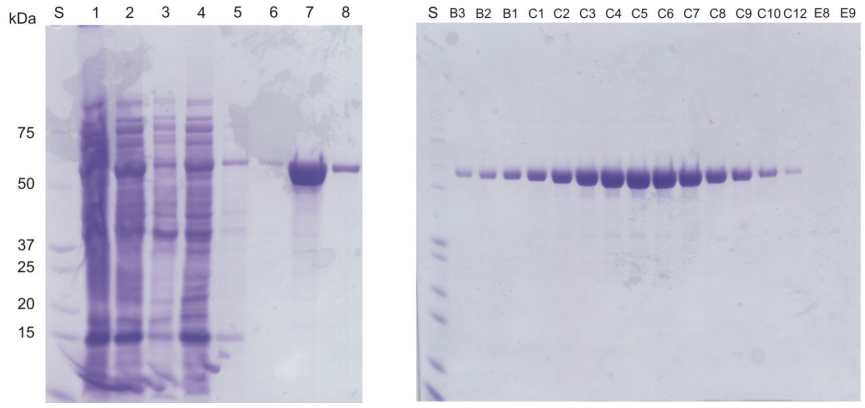


Figure G.2: DesD purification of protein expressed form Rosetta (DE3) cells. **(a)** IMAC purification of DesD. **S**: standard; **1**: crude; **2**: S/N; **3**: pellet; **4**: flow through; **5**: wash1, **6**: wash2; **7**: elution 1; **8:** elution 2. **(b)** SDS-PAGE of fractions B3 – E9 from His_6-DesD gel filtration.

G.2.2. DesD has no TEV-recognition site but thrombin cleaving site

Tests with different amounts of TEV protease revealed that the His_6-tag cannot be cleaved (Figure G.3a). This was surprising since the construct was thought to have a TEV recognition site. However DNA sequence analysis showed that this construct had a thrombin cleaving site instead. The N-terminal His_6-tag, including the recognition site for TEV and spacer was sequenced as MGSSHHHHHHSSG***LVPR***/***GS***SH. The His_6-tag could be removed by incubation with 3 units of thrombin per 100 µg DesD protein (Figure G.3b). 40 mg of DesD were treated with 350 units of thrombin for 30 h at RT. His_6-tag and non-cleaved protein were separated with a 5 mL pre-packed HisTrap column (GE Healthcare) and dialysed in 50 mM HEPES pH 7.5, 500 mM NaCl and 10 % (v/v) glycerol. DesD was concentrated and then purified on a S200 Superdex GF column (Figure G.3c). Purified protein was concentrated to 5.45 mg/mL, frozen in liquid nitrogen and stored at -80 °C. Cleaved DesD has four additional amino acids (GSSH) at the N-terminus, comprises 599 amino acids and has a total weight of 66957 Da.

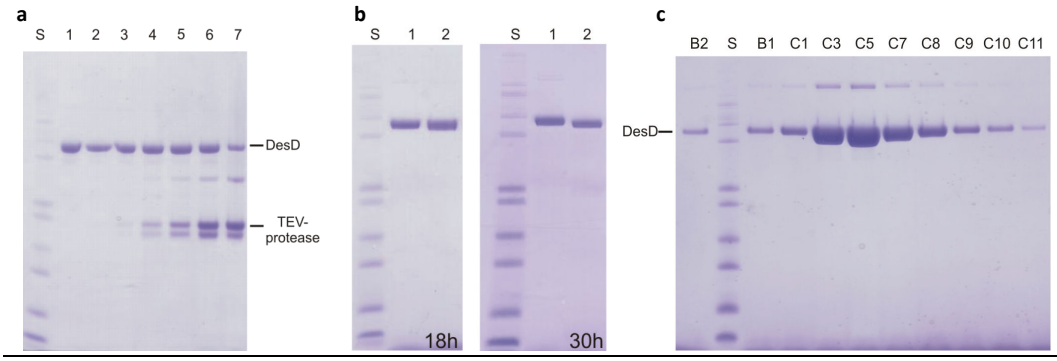


Figure G.3: TEV-protease cleaving and thrombin cleaving. **(a)** TEV-cleavage with different protease to DesD ratios. More protease does not cleave the tag. (1) 10 µg DesD; (2) 1:50 (TEV:DesD ratio); (3) 1:25; (4) 1:10; (6) 1:5; (7) 1:2; (8) 1:1. **(b)** Thrombin cleavage with 1 U/100 µg DesD led to partly cleaved protein after 18 h. Adding another 2 units thrombin/100 µg DesD and incubation for another 12 h (total 30 h) resulted in fully cleaved DesD protein. (1) DesD without thrombin, (2) DesD with thrombin incubated. **(c)** Gel filtration of DesD after His_6-tag was removed by thrombin cleavage and a second Ni-column.

G.2.3. DesD initial screening for crystallization condition

Neither with His_6-DesD nor with DesD an initial crystal condition could be obtained, despite the use of several commercial screens (JMAC, Classics, JCSG+, PEGs, Wizard, Crystal Screen or Cryo I+II). Screening with 10 mM ATP (added to the protein prior to droplet set up) and/or different protein concentrations were also not successful.

However, some condition resulted in spheroids as shown in Figure G.4, though their optimization failed.

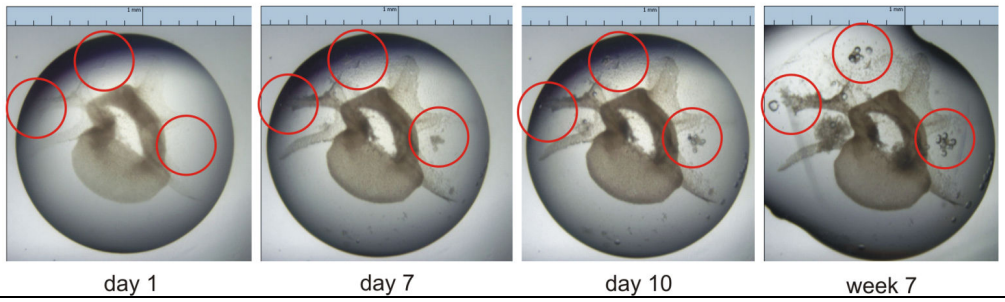


Figure G.4: DesD spheroids timeline. Spheroids started growing after several days (highlighted with red circles) and kept growing until their final size was reached after ~7 weeks. Crystal condition: Well 88 (H11) of the JMAc crystal screen with 20 mg/mL His_6-DesD and 10 mM ATP.

G.2.4. Methylation of DesD

DesD protein was methylated using the JBS methylation Kit (Jena Bioscience). Methylation efficiency was verified by mass spectrometry and confirmed that 100 % of the possible sites had been methylated. Methylated DesD (DesD-Met) protein was dialyzed in 50 mM TRIS-HCl pH 7.5, 250 mM NaCl and 10 % (v/v) glycerol to remove chemicals of the methylation step. Thereafter protein was concentrated to 5.8 mg/mL and used for crystallization screens (Wizard I+II, Index, Cryo I+II and Crystal screen I+II). Crystal screens were set up with 5.8 mg/mL and 2.9 mg/mL DesD-Met (incubated with or without 10 mM ATP). Some initial hits were obtained but the optimized conditions proved to be salt or non-diffracting crystals.

G.2.5. DesD-GFP-His_8 fusion construct

Since no crystallization condition was found, a C-terminal His_8-tag fusion protein of DesD and GFP (as reporter protein) was constructed using the pWaldo-GFPd vector (Drew et al. 2001; Rapp et al. 2004). DesD primers containing a Nde1 (forward primer) and a BamH1 (reverse primer) recognition site were synthesized and used in PCR reactions to introduce the restriction site (Figure G.5).

primer name	primer orientation	primer DNA sequence	N	GC No	AT No	Tm no	GC pp	Tm pp	Tm pp
DesDcTGFP_f	forward	GAAGAACATATGAGCCTCGCCGACGCCG	28	3	6	**24**	14	5	**66**
DesDcTGFP_r	reverse	GATTTCGGATCCCCGGCCTGCCAGG	25	6	6	**36**	11	2	**48**

Figure G.5: Primers for DesD-GFP-His_8 construct.

PCR was carried out with a temperature annealing gradient from 53.5-70.5 °C. Details of PCR reaction and the program schedule are shown in Table G.1a, b. The most PCR product was yielded when 400 nM primers were used with an annealing temperature of 70.5 °C. Four 50 µL PCR reactions were carried out with these optimized settings using the same PCR program as before. The amplified DesD DNA was separated on a preparative 0.9 % (w/v) argarose gel. The DesD PCR fragment was extracted using the Promega gel extraction kit.

Table G.1: Protocols for DesD-GFP-His_8 vector construct. **(a)** Reaction scheme for PCR. **(b)** PCR program with different annealing temperatures; T_1=70.5 °C, T_2=68 °C, T_3=64.2 °C, T_4=59.5 °C, T_5=55.9 °C and T_6=53.5 °C. **(c)** Restriction digest with Nde1 and BamH1 of eluted DesD DNA and pWaldo-GFPd vector at 37 °C for 4 h. After the first 2 h of incubation additional 2.5 µL BamH1 to each reaction was added, since enzyme is only 75 % active in restriction buffer D (Promega). **(d)** Protocol for ligation of prepared vector and DesD DNA with T4 ligase.

a

primer conc. (nM)			**50**	**150**	**250**	**400**
	stock	final				
PFU buffer (x)	10	1.00	35.0 µL	35.0 µL	35.0 µL	35.0 µL
dNTPs (mM)	10	0.25	8.8 µL	8.8 µL	8.8 µL	8.8 µL
DesD WT (ng)	65	32.50	1.0 µL	1.0 µL	1.0 µL	1.0 µL
primer fw (nM)	10000	x	1.8 µL	5.3 µL	8.8 µL	14.0 µL
primer re (nM)	10000	x	1.8 µL	5.3 µL	8.8 µL	14.0 µL
$MgSO_4$ (mM) +2 mM	50	3.00	21.0 µL	21.0 µL	21.0 µL	21.0 µL
DMSO (%)	100	2.00	7.0 µL	7.0 µL	7.0 µL	7.0 µL
water			273.8 µL	266.8 µL	259.8 µL	249.3 µL
Polymerase PFU			0.5 µL	0.5 µL	0.5 µL	0.5 µL

b

step	cycle	temp.	time
1	1x	95 °C	4 min
2	30x	95 °C	45 s
		72 °C 52 °C↑	45 s
		72 °C	4 min
3	1x	72 °C	10 min
		4 °C	∞

c

	DesD PCR	GFPd vector
Restriction buffer D (10x)	10.0 µL	10.0 µL
BSA (10 mg/mL)	1.0 µL	1.0 µL
DesD PCR product (1.26 µg)	56.0 µL	/
GFPd vector	/	5.0 µL
BamH1 (10 U/µL)	2.5 µL	2.5 µL
Nde1 (10 U/µL)	2.5 µL	2.5 µL
Water	28.0 µL	79.0 µL
volume tot.	100.0 µL	100.0 µL

d

	DesD	control
Vector DNA	150 ng	150 ng
DesD DNA	150 ng	/
Ligation buffer (10x)	1x	1x
T4 Ligase	1.5 U	1.5 U

Extracted DesD DNA and the pWaldo-GFPd vector were digested with Nde1 and BamH1 restriction enzymes for 4h at 37 °C. After the first 2 h of incubation more BamH1 (25 U) was added to get full cleavage of DNA and vector (Table G.1c). Thereafter one unit of shrimp alkaline phosphatase (SAP) was added to the pWaldo-GFPd reactions to dephosphorylate the 5'-phosphates of the vector. DNA of digested

DesD and the prepared vector were purified with the "MinElute Reaction Cleanup" kit (Qiagen). The ligation of vector and DesD was carried out with 1.5 units of T4 ligase and equal amounts of vector and DNA (150 ng each) at RT over night (Table G.1d). DH5α competent cells were transformed with 0.5 – 1 µL of this ligation product. DesD plasmid containing clones (grown on 30 µg/L kanamycin LB agar plates) were identified by colony PCR and a control digestion with BamH1 and Nde1, but finally checked by sequencing. The final DesD-GFP-His_8 fusion protein sequence comprised 855 amino acids with a His_8-tag and is displayed in Figure G.6.

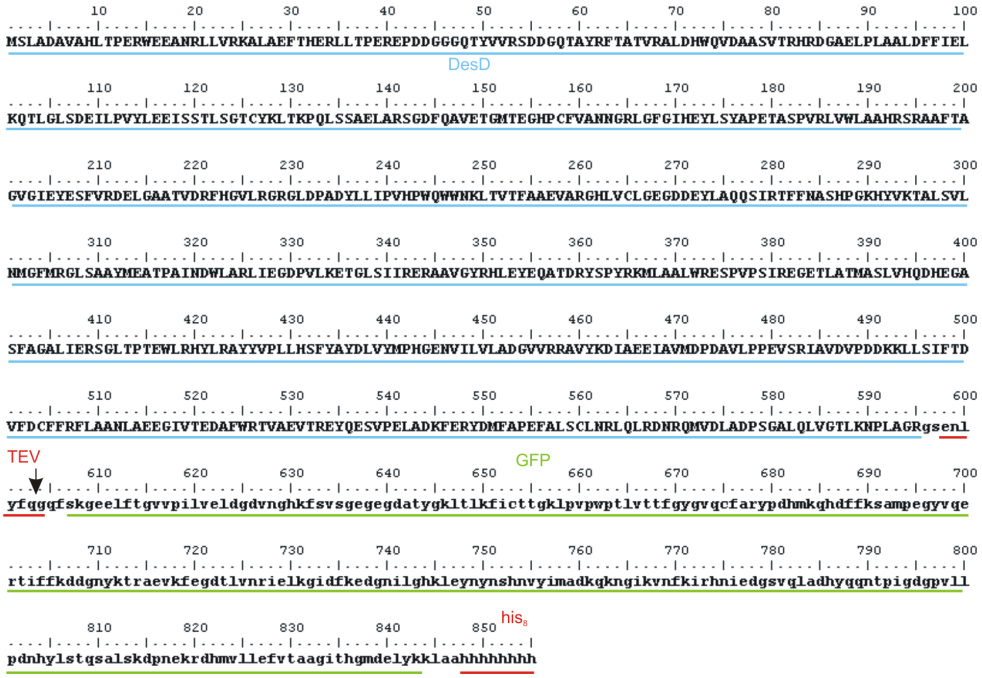


Figure G.6: Protein sequence of the final DesD-GFP-His_8 fusion construct.

G.2.6. DesD-GFP-His_8 protein expression

DesD-GFP-His_8 fusion vector was transformed into Rosetta (DE3) competent cells for protein expression. Cells were grown in 6x 1 L cultures, inoculated with 5 mL O/N starter cultures, containing 30 µg/L kanamycin and 40 µg/mL chloramphenicol. Protein expression was induced with 0.5 mM IPTG at culture density of OD_{600} =0.8 and incubated for a further 15 h. Cell lysis and protein purification were carried out as described for AcsD (chapter 2) (Figure G.7). GFP-His_8 tag of elution 1 and 3 was removed by incubation with TEV protease (1:50) (Figure G.7c).

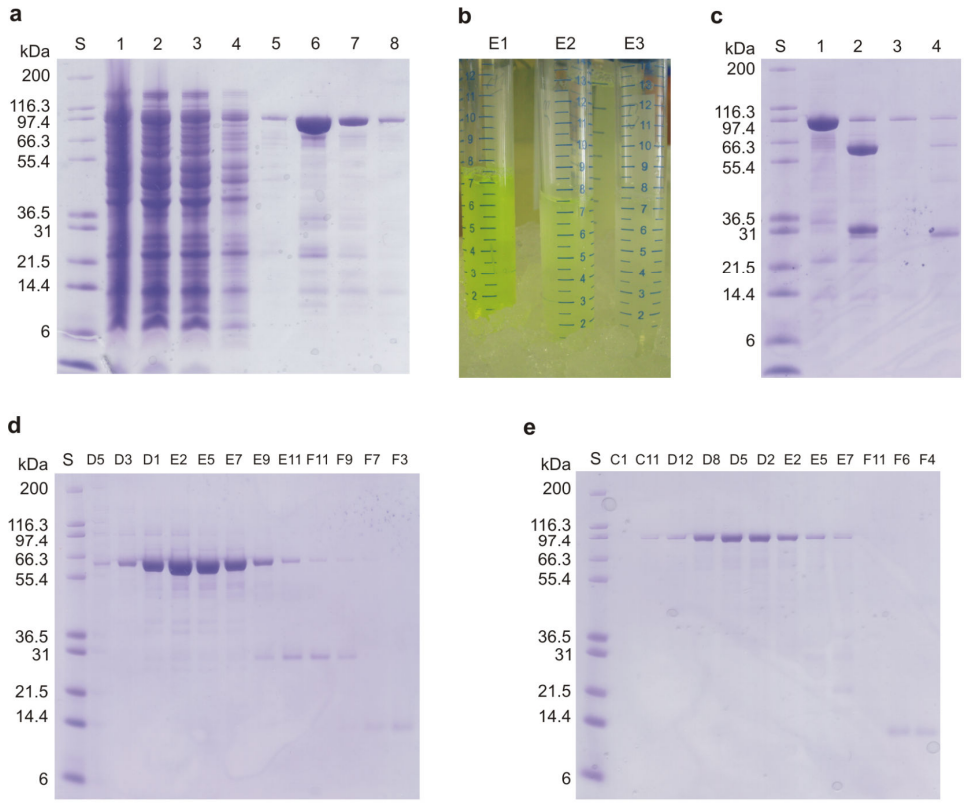


Figure G.7: DesD-GFP-His_8 purification. **(a)** IMAC purification of DesD-GFP-His_8. **S**: standard; **1**: crude; **2**: S/N; **3**: flow through; **4**: wash 1; **5**: wash 2, **6**: elution 1; **7**: elution 2; **8:** elution 3. **(b)** Elutions 1-3 contain DesD-GFP fusion protein. **(c)** TEV cleavage of DesD-GFP-His_8. **S**: standard; **1**: E1 not cleaved; **2**: E1 cleaved; **3**: E3 not cleaved; **4**: E3 cleaved. **(d)** Gel filtration of cleave E1 containing DesD. **(e)** Gel filtration of E2 containing DesD-GFP-His_8.

Cleaved GFP-His_8 and TEV protease were removed on a second Ni-column (see chapter 2). Elution 1 (DesD) and elution 2 (DesD-GFP-His_8) were further purified on a S200 Superdex GF column, concentrated to 6.5 mg/mL, frozen in liquid nitrogen and stored at -80 °C. The final protein has 8 additional C-terminal residues (GSENLYFG) and comprises 603 amino acids with a total weight of 67528 Da. The non-cleaved GFP fusion protein comprises 855 residues with a total weight of 96032 Da.

G.2.7. Screening with cleaved and non cleaved DesD-GFP-His_8

DesD-GFP-His_8 and cleaved DesD were used for screening with four commercial screens (Wizard I+II, Crystal screen I+II, JCSG+ and Cryo I+II) and one custom stochastic screen (screen 20 Run2 (10/09/08)). Wizard I condition 22 (0.1 M Tris pH 8.5, 10 % (v/v) 2-propanol) resulted in small crystals (needles) with 6.5 mg/mL DesD-GFP-His_8 protein. Since the crystals were quite small (Figure G.8) optimization of crystals will be needed. However, it has not yet been confirmed, whether these

crystals are protein crystals or not. But since no salt was used as precipitant or as an additive it is likely that these are DesD-GFP fusion crystals. The crystals are too small to judge whether they are stained, as GFP usually has a green color (see also Figure G.7b).

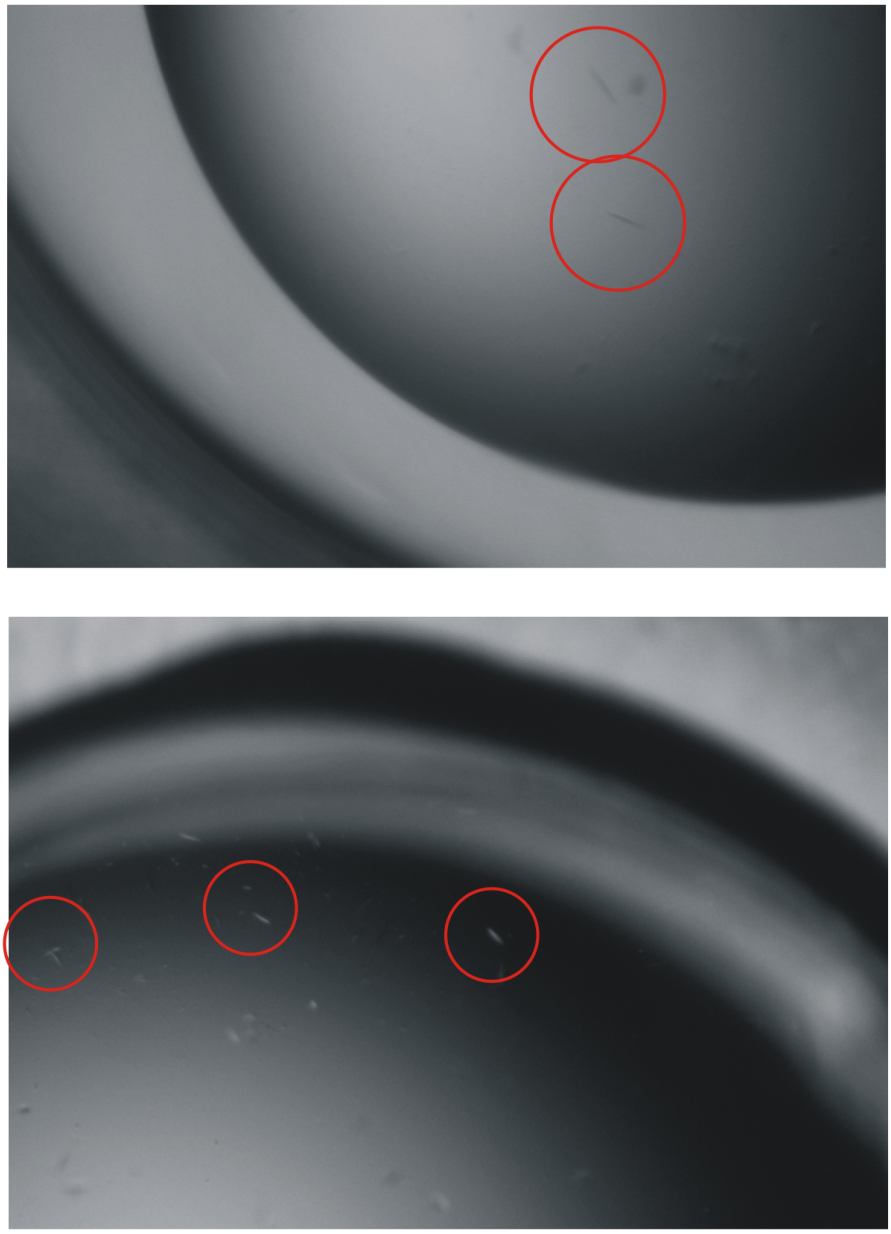


Figure G.5: Likely DesD-GFP-His_8 crystals. Crystals are tiny and have a “needle” morphology. They still have to be identified as protein crystals, but need to be improved to test them by x-ray or by staining with Coomassie blue.

G.3. Conclusion and future work

Initial expression and purification problems have been solved and pure N-terminal tagged DesD has been expressed and purified. It was shown that the used vector has no N-terminal TEV recognition site for cleaving the His_6-tag, but a thrombin recognition site. The tag-removal with thrombin was successful, but screening for an initial crystal condition remained unsuccessful. A C-terminal His_8-tagged DesD-GFP fusion vector was constructed, the fusion protein expressed, purified and successfully cleaved with TEV protease. Crystal trails with cleaved and uncleaved DesD-GFP were screened for a potential crystallization condition. One promising condition in the Wizard I screen was obtained. However crystals still have to be identified as protein crystals, and need further optimization. This could be achieved by systematic use of grid screens, which eventually might lead to good quality crystals.