Catalysis, substrate binding and specificity in the amidase from *Nesterenkonia* species

Serah Wangari Kimani

Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Molecular and Cell Biology
UNIVERSITY OF CAPE TOWN

Supervisor: Professor Trevor Sewell

December 2011

Abstract

"Catalysis, substrate binding and specificity in the amidase from *Nesterenkonia* species"

Serah Wangari Kimani
PhD thesis, Department of Molecular and Cell Biology, University of Cape Town
August 2011

Nitrilase superfamily amidases catalyze the conversion of amides to their corresponding acids and ammonia; they perform essential metabolic roles and are useful in the manufacture of fine chemicals and pharmaceuticals. A conserved Cys, Glu and Lys (CEK) catalytic triad has been identified in these enzymes, and catalysis proposed to follow a 'ping pong bi bi' mechanism in which the substrate forms a thioester with the catalytic cysteine. Details of each step in the pathway remain to be clarified. A 'second', structurally conserved glutamate that is not recognizable from sequence conservation alone has been found to be essential for activity in these enzymes, but its role in catalysis has never been elucidated. This work uses site-directed mutagenesis of the active site residues (E61, E139, C165 and K131) of the model aliphatic amidase from *Nesterenkonia* species (NitN), followed by mass spectrometric and X-ray crystallographic analysis of the mutants in the presence of a range of substrates to probe both the catalytic role of the glutamates and the basis of NitN substrate specificity. All mutants lacked enzymatic activity confirming the requirement of the four residues in the catalytic process. NitN has a preference for short aliphatic amides (SAA), of substrates tested previously, however the C165A mutants were found in this study to bind two SAA molecules in the active site, indicating possible capacity for larger substrates. Moreover, acrylamide (ACR) molecules were not adequately restrained by the active site pocket, as evidenced by their binding in multiple orientations and forming multiple covalent adducts. The amide group of the substrate at the catalytic site is positioned for the cysteine nucleophilic attack by hydrogen bonding with both glutamates and the catalytic lysine (K131). The amide carbonyl is located in an 'oxyanion hole' formed by the protonated amino group of K131 and the peptidic NH group of Y166, while the side chain moiety is stabilized through hydrophobic contacts. The structure of the C165A mutant in complex with propionitrile also revealed that the 'oxyanion hole' can be filled by a nitrile, which has implications for the mechanism of nitrilases. Instability of the E61Q/L and E139Q mutants suggests that the stability of the fold of NitN is dependent on hydrogen bonding interactions between the side chain of these glutamates and the lysine. Mass spectrometry

showed that a thioester enzyme-substrate intermediate could be trapped by mutating E61. The water substrate for the second phase of the reaction is positioned by hydrogen bonding to both glutamates. Incubation of the glutamate mutants with amide substrates resulted in the formation of two further types of adduct to the catalytic cysteine. An S_N2 substitution reaction involving the displacement of the fluorine in fluoroacetamide was observed with the E61Q/L mutants, while Michael addition was observed with ACR when either of the glutamate residues was mutated. These reactions not only emphasize the role that the complete catalytic assembly plays in positioning the substrates but also demonstrate that the 'first' glutamate (E61) is not necessary to catalyze nucleophilic attack by the cysteine. This is contrary to previous proposals that this glutamate could play the role of a general base catalyst in the acylation phase of the reaction. Mutation of K131 to glutamine resulted in the formation of a covalently modified expressed protein that appears to have an adipamide thioester intermediate trapped in the active site, once again emphasizing the need for the intact catalytic assembly to catalyze the hydrolysis of the acyl intermediate. In summary, this work has revealed, in detail, the role of the EKE triad which includes maintenance of the active site configuration, maintenance of the protein fold, positioning of the amide moiety for nucleophilic attack and general base catalysis role of the 'first' glutamate in the deacylation phase of the reaction. In addition, our results support a CEEK catalytic tetrad in the nitrilase superfamily enzymes, as opposed to the previously proposed CEK triad.

Dedication

I dedicate this thesis with love to my mama Mary Mukina.

Niundu wa wendo waku muingi, kwirutira, kwiheana na kwi-ima

Uri umwe thi-ini wa million maitu!!

Acknowledgements

I would like to thank my supervisor Prof. Trevor Sewell for guidance, enthusiastic support and input, and for creating an environment in which it has been a pleasure to learn.

I am grateful to Pro. Don Cowan of the Biotechnology department at the University of the Western Cape for providing us with the *Nesterenkonia* species amidase-containing plasmid.

Thanks to Dr. Andrew Thomson of Proxima I beamline at the Synchrotron Soleil (Paris, France) and to Prof. Wolf-Dieter Schubert of the Biotechnology department at the University of the Western Cape for assistance with diffraction data collection.

I am grateful to Prof. Roger Hunter of the chemistry department at the University of Cape Town (UCT) for many discussions and correspondences that helped make sense of chemistry.

Many thanks to Dr. Jean Watermeyer (Electron Microscope Unit (EMU), UCT) for her continued support, encouragement and unconditional friendship, and for assistance with proofreading most of the thesis. I am also grateful to Dr. Brandon Weber and Simon Broadley of the EMU (UCT) for proofreading parts of this thesis.

I am grateful to the staff of the EMU (UCT) for their support and assistance and to my fellow students for their inspiration and friendship.

I am indebted to my family for their patience, prayers and encouragement, and to my friends who have supported and encouraged me in the course of my studies.

I gratefully acknowledge the financial assistance of the South African National Research Foundation, Molecular and Cell Biology department (UCT), International Union of Crystallography, Roche Diagnostics (South Africa), the South African synchrotron fund and the University of Cape Town.

Finally, I acknowledge the hand of God in all these; He is the giver of knowledge and wisdom.

List of Abbreviations

2X YT	2x Yeast extract and Tryptone bacterial growth medium	**LB**	Luria Broth
3D	three dimensional	**LUMO**	Lowest Unoccupied Molecular Orbital
ACE	acetamide	**MALDI-TOF**	Matrix Assisted Laser Desorption Ionization - Time of Flight
ACR	acrylamide	**MR**	molecular replacement
ADM	adipamide	**MS**	mass spectrometry
AmiE	the branch 2 aliphatic amidase gene	**MTB**	*Mycobacterium tuberculosis*
AmiF	the formamidase gene from *Helicobacter pylori*	**MWCO**	molecular weight cut-off
Amp	ampicillin	**MW**	molecular weight
ANT	acetonitrile	**MWM**	molecular weight markers
ASU	asymmetric unit	**Na-nitroprusside**	Sodium nitroprusside
βAS	β-alanine synthase	**$NaAD^+$**	Nicotinic acid Adenine Dinucleotide
BMD	butyramide	**NAD^+**	Nicotinamide Adenine Dinucleotide
BNT	butyronitrile	**$NADsyn^{gln}$**	glutamine-dependant NAD^+ synthetase
BSA	bovine serum albumin	**NCS**	non-crystallographic symmetry
CCP4	Collaborative Computational Project number 4	**NHase**	nitrile hydratase
CEK	Cys, Glu and Lys catalytic triad	**Nit**	nitrilase
CEEK	Cys, Glu, Glu and Lys catalytic tetrad	**NitNFhit**	Nitrilase-Fhit rosetta stone protein from *C. elegans*
CFE	cell free extract	**NitN**	*Nesterenkonia* species amidase
Cm	chrolamphenicol	**PCR**	Polymerase Chain Reaction

C-N bond	carbon-nitrogen bond	**PDB**	Protein Data Bank
CPK	Corey-Pauling-Kultun scheme of coloring atoms	**PDB id**	Protein Data Bank identity code
DCase	bacterial N-carbamyl-D-amino acid amidohydrolase	**PMD**	propionamide
DmβAS	*Drosophila melanogaster* β-alanine synthase	**PNT**	propionitrile
dNTPs	dinucleotide triphosphates	**RAPc8 amidase**	the aliphatic amidase from *Geobacillus pallidus* RAPc8
DON	6-diazo-5-oxo-L-norleucine	**rmsd**	root mean square deviation
DTT	Dithiothreitol	**rpm**	revolutions per minute
FAE	fluoroacetamide	**SAA**	short aliphatic amides
FMD	formamide	**SAB**	sample application buffer
FNT	fluoroacetonitrile	**SEC**	size exclusion chromatography
GAT	glutamine amide transfer	**SDS-PAGE**	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
GF	gel filtration	**TF**	Trigger Factor
HOMO	Highest Occupied Molecular Orbital	**TMV**	tobacco mosaic virus
IPTG	Isopropyl-Beta-D-Thiogalactoside	**V_o**	void volume
Kan	kanamycin	**V_t**	total volume
kDa	Kilo Dalton	**V_m**	Matthews coefficient
LA	Luria Agar	**WT**	Wild-type
FNT	fluoroacetonitrile		

Table of contents

Chapter 1

Literature review

Chapter synopsis

This chapter covers the background information that is available in the literature concerning substrate specificity and the catalytic mechanism of the nitrilase superfamily amidases. It is divided into 5 sections. Section 1 gives a general introduction into activity, classification, structural fold and oligomerization of the nitrilase enzymes, while section 2 reviews the structural determinants of substrate specificity in amidases. Section 3 explores what is known about the catalytic mechanism of the nitrilase superfamily amidases (mainly), section 4 covers activity and the substrate profile of the model amidase from *Nesterenkonia* species (NitN), while section 5 presents study objectives and justification.

1.1 General Introduction

1.1.1 C-N hydrolases or the nitrilase superfamily enzymes

C-N hydrolases are a group of homologous enzymes that are found in plants, animals, fungi and many prokaryotes, where they catalyze hydrolysis or condensation of non-peptidic C-N bonds. The hydrolytic reactions of these enzymes yield a range of essential natural products such as β-alanine, biotin, NAD^+, auxin, *et cetera* and in some cases result in the deamination of protein and amino acid substrates, while the C-N bond condensation reactions are important in post-translational modifications of amino acids, proteins and other compounds (Pace and Brenner 2001).

Non-peptidic C-N bond-containing compounds are widespread in nature and they include among others, organic cyanides or nitriles (R-C≡N), inorganic cyanides (H-C≡N), acid amides [R-C(=O)-NH_2], secondary amides [R-C(=O)NH-R'] and N-carbamyl amides [R-NH-C(=O)-NH_2] (Brenner 2002). C-N hydrolases hydrolyze these compounds by attacking either the cyano carbon of a linear nitrile substrate or the planar carbon of an amide substrate, using a conserved cysteine residue in a covalent catalytic process (Bork and Koonin 1994;Stevenson et al. 1990). A small group of the nitrilase superfamily enzymes condense acyl moieties of some of the compounds mentioned above to the termini of proteins and other compounds during post-translational modifications that are essential for maturation and activity. Figure 1.1 summarizes the four distinct reactions that are catalysed by the members of the nitrilase superfamily.

Although N-terminal nucleophile (Ntn) hydrolases and triad amidotransferases (Zalkin and Smith 1998), signature amidases (Chebrou et al. 1996;Patricelli and Cravatt 2000) and nitrile hydratases (Huang et al. 1997;Kobayashi and Shimizu 1998) also act on the same C-N compounds listed above, they are structurally and mechanistically unrelated to the nitrilase superfamily enzymes (Brenner 2002).

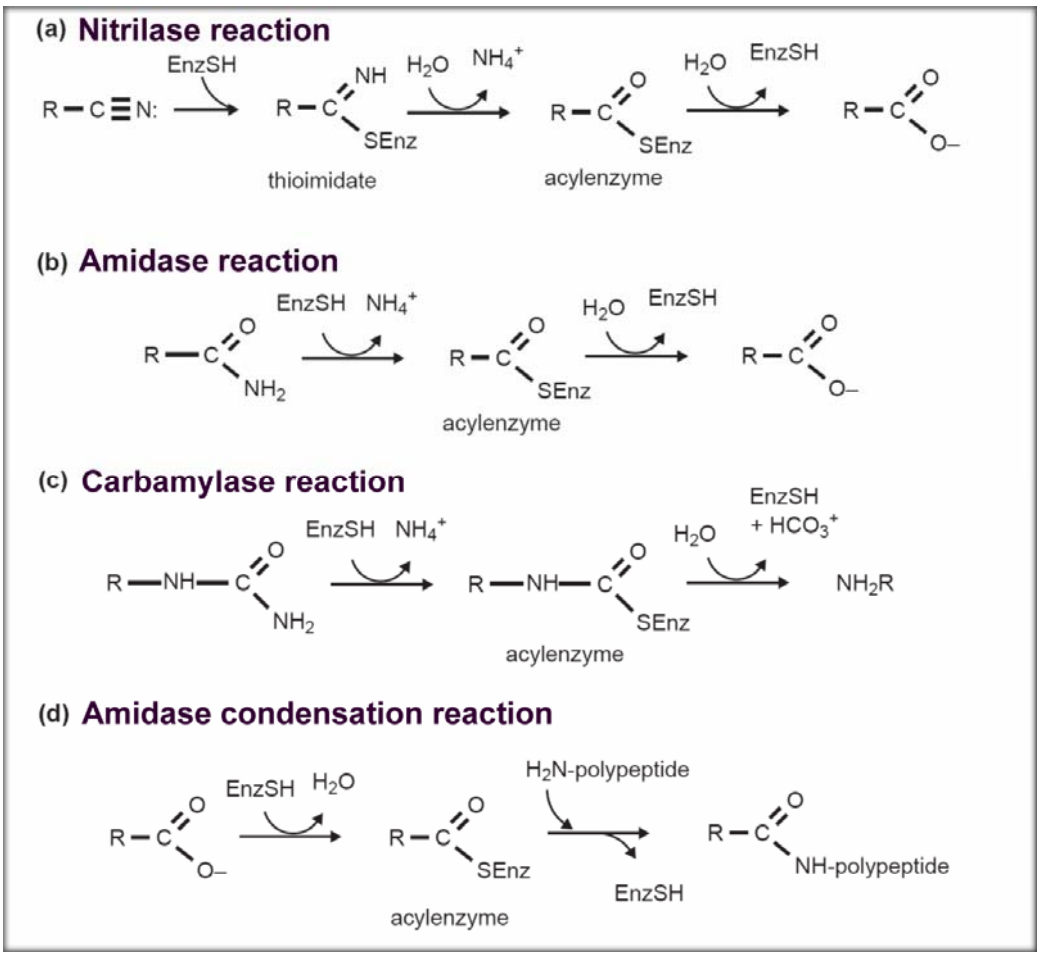


Fig. 1.1: Four distinct reactions that are catalyzed by the members of the nitrilase superfamily. Diagram taken from Pace and Brenner 2001.

1.1.2 Classification of the nitrilase superfamily C-N hydrolases

On the basis of global and structure-based sequence similarities, signature sequences in the vicinity of the catalytic residues and the presence of additional domains, Pace and Brenner (2001) classified the enzymes of the nitrilase superfamily into 13 branches. Members of the 7 branches have the nitrilase-like domain fused to domains of varied functions (Fig. 1.2), which is necessary for: linking ammonia production to ammonia consumption like in the case of the glutamine-dependent NAD^+ synthetases; localization of the nitrilase domain; and possibly linking of proteins involved in cellular signaling (Brenner 2002). Even though the superfamily is designated as 'nitrilase', only branch 1 enzymes have cyanide hydratase or nitrile-hydrolyzing activity, while most of the remaining branches (branches 2-11) are comprised of amidases with different physiological activities and substrate specificities as documented in figure 1.2 and summarized below.

Branch 1 nitrilases catalyze the hydrolysis of nitriles to their corresponding acids and ammonia. Although these enzymes are commonly found in bacterial species that are known to metabolize nitriles as a sole source of carbon and nitrogen, their physiological functions are still not clear (Banerjee et al. 2002). However, as certain bacterial species like *Rhodococcus* and *Bacillus* are capable of metabolizing aldoximes through nitrile intermediates, it has been speculated that nitrile-degrading enzymes could be components of complex pathways that control production and degradation of cyanogenic glycosides and related compounds where aldoximes are key intermediates (Banerjee et al. 2002;Kato et al. 1998;Kato et al. 2000). The nitriles generated from aldoximes are degraded through various metabolic routes including hydrolysis, oxidation (oxygenases; (Sawyer et al. 1984)) and reduction (nitrogenases; (Liu et al. 1997)) (Fig. 1.3), with nitrile hydrolysis in microbes following two pathways: (1) A single hydrolytic pathway by nitrilases resulting in the corresponding acids and ammonia; and (2), a bi-enzymatic pathway that involves hydration of nitriles to the corresponding amides by nitrile hydratases, followed by the conversion of amides to the corresponding organic acids and ammonia by amidases. C-N hydrolase enzymes, cyanide hydratase and cyanide dihydratase function as detoxifying agents for the toxic cyanide compounds that are formed abiotically from metal cyanides (Banerjee et al. 2002), by hydrolyzing them to formamide, and formic acid and ammonia respectively (Ingvorsen et al. 1991;Wang et al. 1992).

Branch 2 is comprised of highly homologous aliphatic amidases (EC 3.5.1.4) from prokaryotes that catalyze the hydrolysis of short aliphatic amides in addition to other *in vitro* activities (Fournand and Arnaud 2001). Although the physiological function of these enzymes is not understood, most of them are likely to be involved in the degradation of toxic nitriles (Banerjee et al. 2002;Kato et al. 1998) in the same pathways as nitrile hydratases (Kobayashi and Shimizu 1998;Yamada and Kobayashi 1996), as well as in carbon and nitrogen fixation in bacterial cells (Sharma et al. 2009). The ammonia produced from the breakdown of various amides by the two paralogous aliphatic amidases from the grastric pathogen, *Helicobacter pylori,* has been suggested to be important as a nitrogen source, as a compound for protecting the bacteria against gastric acidity and as a cytotoxic molecule (Bury-Mone et al. 2003;Skouloubris et al. 2001).

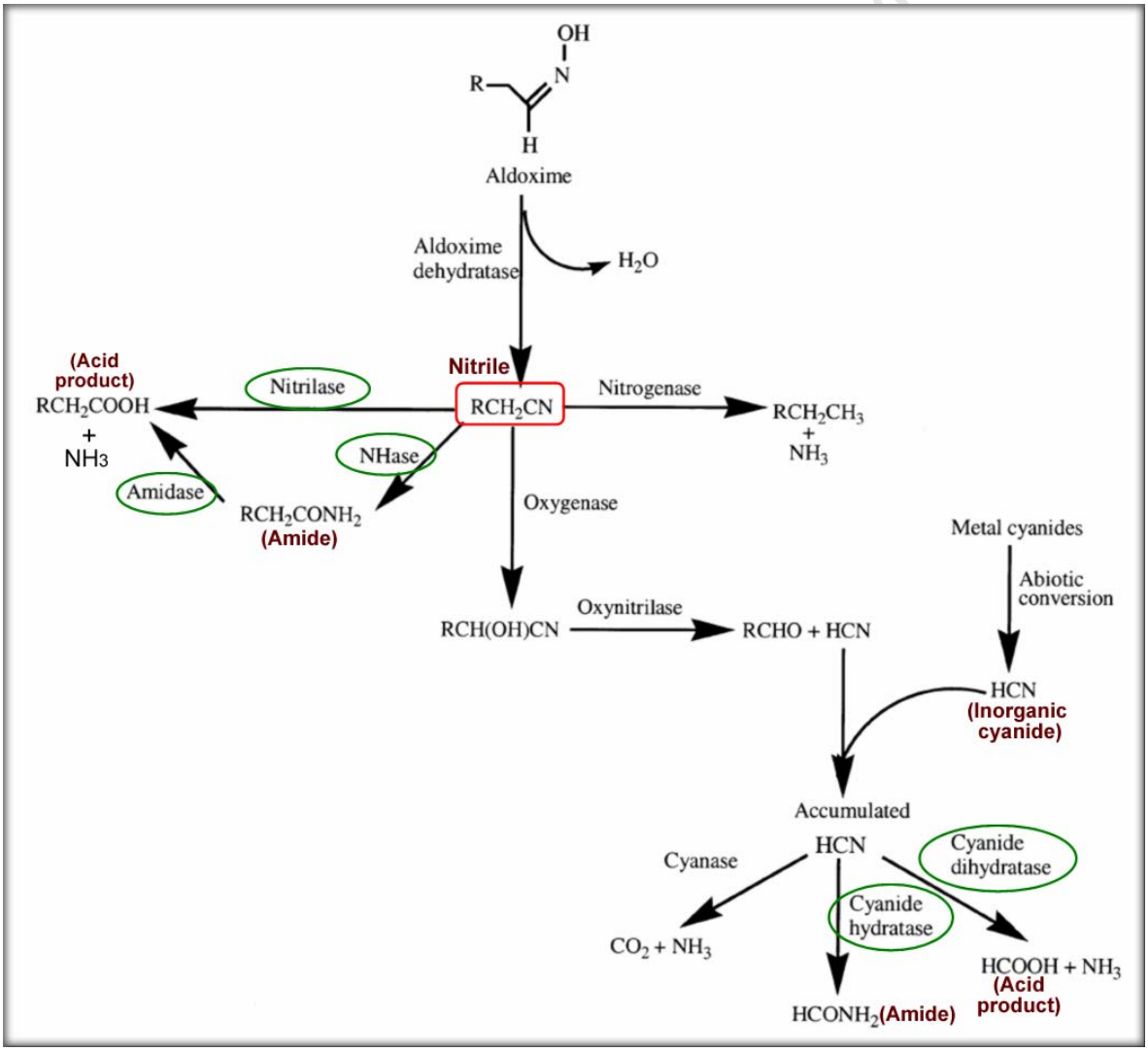
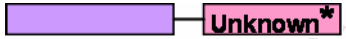
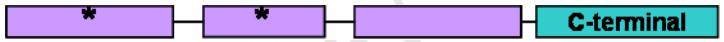


Fig. 1.3: Pathways for nitrile metabolism in micro-organisms. Enzymes from the nitrilase superfamily are circled in green and various nitrile metabolism products are labelled in red. The figure is from Banerjee et al. 2002 with slight modifications.

Nitrilase Branch number & name	Enzymatic reaction	Domain structure	Representative member of the branch Example of specific reaction
1-Nitrilase	Nitrilase		***Rhodococcus* Nit A** (Kobayashi et al. 1992a) Acrylonitrile + $2H_2O$ → Acrylic acid + NH_3
2-Aliphatic amidase	Amidase		***Pseudomonas* amiE** (Novo et al. 2002) Propionamide + H_2O → Propionic acid + NH_3
3-Amino-terminal amidase	Amidase		***Saccharomyces* NTA1** (Baker and Varshavsky 1995) N-terminal asparagine + H_2O → N-terminal aspartic acid + NH_3
4-Biotinidase and pantetheinase	Secondary amidase		**Human BTD** (Hymes and Wolf 1996) Biocytin + H_2O → Biotin + lysine
5-β-Ureidopropionase	Amidase		**Rat BAS** (Kvalnes-Krick and Traut 1993) N-carbamyl-β-alanine + $2H_2O$ → β-alanine + CO_2 + NH_3
6-Carbamylase	Amidase		***Agrobacterium* DCase** (Nakai et al. 2000) N-carbamyl-D-methionine + H_2O → D-methionine + CO_2 + NH_3
7-Prokaryotic Gln-dependent NAD^+ synthase	Amidase		***Mycobacterium* NAD^+ synthetase** (Bellinzoni et al. 2002) Glutamine + H_2O → Glutamic acid + NH_3

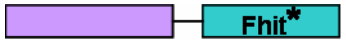
8-Eukaryotic Gln-dependent NAD^+ synthase	Amidase		**Human NAD^+ synthetase** (Hara et al. 2003) Glutamine + H_2O → Glutamic acid + NH_3
9-Apolipoprotein N-acetyltransferase	Reverse amidase		***Pseudomonas* cutE** (Piotrowski et al. 2001) N-terminal DAG-modified Cys + palmitate → N-palmitylated protein + H_2O
10-Nit and Nitfhit	ω-amidase		**Mammalian Nit2** (Krasnikov et al. 2009a) α-ketoglutaramate + H_2O → α-ketoglutarate + NH_3
11-N-carbamyl putrescine amidohydrolase	Amidase		***Pseudomonas* AguB** (Nakada et al. 2001) N-carbamyl putrescine + H_2O → Putrescine + CO_2 +NH_3
12-NB12	unconfirmed		***Sphingomonas* 1dhX** (Pace and Brenner 2001) Unknown reaction
13-Non-fused outliers	unknown		Unknown reaction

Fig 1.2: A summary of the reactions catalyzed by the 13 branches of the nitrilase superfamily enzymes and their domain organization. Enzymes in the 7 of the 13 branches have nitrilase-related domain (coloured purple) fused to other domains. Domains with a black star are only found in some members of the branch. The information in this figure was mainly consolidated from the nitrilase superfamily classification reviews: (Pace and Brenner 2001;Brenner 2002).

Branch 3 enzymes (N-terminal amidases) catalyze the conversion of asparagine or glutamine residues at the N-terminus of a polypeptide chain to the corresponding aspartate or glutamate residues respectively (Baker and Varshavsky 1995), and they function in the N-end rule pathway of protein degradation (Mogk et al. 2007;Varshavsky 1997) that is present in both prokaryotic and eukaryotic organisms for regulation of protein concentrations during various cellular functions (Eisele and Wolf 2008). Asparagine and glutamine are tertiary destabilizing N-degrons (proteolysis degradation signals) in *Saccharomyces cerevisiae* yeast (Bachmair and Varshavsky 1989;Baker and Varshavsky 1995) and mammalian cells (Gonda et al. 1989;Levy et al. 1996), and their deamination through the activity of N-terminal amidases constitutes the first step of ubiquitin-dependent N-end rule pathway. The yeast nitrilase-related N-terminal amidase is specific for both glutamine and asparagine (Gonda et al. 1989;Wang et al. 2009), while the two mammalian N-terminal amidases are highly specific for either asparagine or glutamine and they are not members of the nitrilase superfamily (Wang et al. 2009).

Branch 4 is comprised of secondary amidases (biotinidases and pantetheinases) that are found in mammalian organisms, where they play the role of recycling vitamins in the cells. Biotinidases (EC 3.5.1.12) catalyze the hydrolysis of biotinamides like biocytin, biotinylated peptides, biotin esters and simple biotinamides to form biotin (vitamin B7) and lysine at acidic pH (Hymes and Wolf 1996), while pantetheinases (EC 3.5.1.-; pantetheine hydrolases) catalyze the hydrolysis of D-pantetheine to pantothenic acid (vitamin B5) and cysteamine (Dupre et al. 1970). Biotinidases are also responsible for freeing biotin from biotin-bound dietary proteins (Wolf et al. 1985), as well as recycling it from biocytin that is formed upon degradation of biotinylated holocarboxylases (Wolf 2001). In addition to their biotinyl-hydrolase function, biotinidases also exhibit biotinyl-transferase activity at neutral to alkaline pH, which heavily suggests that, their most dominant physiological function is to transfer the biotin moiety from biocytin to suitable nucleophilic acceptor molecules, such as histones (Hymes et al. 1995). Pantetheinases on the other hand are key enzymes in the dissimilative pathway that liberates pantothenate from biological complexes such as coenzyme A (CoA) and acyl carrier protein (ACP), allowing its recycling in the cell (Dupre and Cavallini 1979).

Branch 5 and 6 enzymes are specialized amidases that convert N-carbamyl compounds to the corresponding primary amines, with the release of carbon dioxide and ammonia (Brenner 2002). Branch 5 is comprised of β-alanine synthases (βAS; also known as β-ureidoproprionases

or N-carbamyl-β-alanine amidohydrolases, EC 3.5.1.6) that are involved in pyrimidine bases catabolism pathway in eukaryotes (Piskur et al. 2007) while branch 6 is made up of the bacterial N-carbamyl-D-amino acid amidohydrolases (DCases), the biological function of which is still unknown (Nakai et al. 2000).

Branches 7 and 8 are glutamine-dependent NAD^+ synthetases ($NADsyn^{gln}$) that function in the last step of the NAD^+ biosynthetic pathways, where they catalyze the ATP-dependent conversion of nicotinic acid adenine dinucleotide ($NaAD^+$) to nicotinamide adenine dinucleotide (NAD^+) cofactors that are essential for various physiological functions in the cell (Belenky et al. 2007). These enzymes are characterized by the presence of two distinct domains (glutaminase and synthetase) on the same polypeptide chain (Bellinzoni et al. 2002) and they occur in all eukaryotic organisms (branch 8) as well as some prokaryotes (branch 7) including *Mycobacterium tuberculosis* (MTB), *Streptomyces* sp., *Plasmodium* sp. among others (Bellinzoni et al. 2002). In these enzymes, the amino-terminal glutamine amide transfer (GAT; also called glutaminase) domain utilizes L-glutamine as the sole nitrogen source to generate ammonia, which is then transferred to the $NaAD^+$ substrate by the NAD^+ synthetase domain (Fig. 1.4).

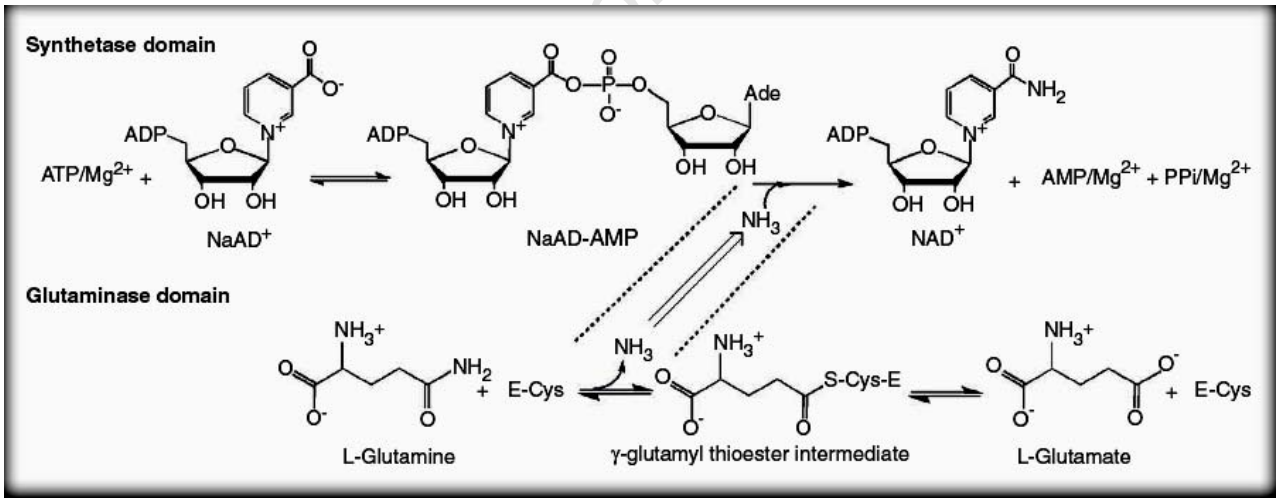


Fig. 1.4: Schematic representation of the coupled reactions that are catalyzed by the glutamine-dependent NAD^+ synthetases from eukaryotes and and some prokaryotes. The glutaminase domain hydrolyzes glutamine to generate ammonia, which is shuffled to the synthetase domain where NAD^+ cofactors are generated. The figure was taken from LaRonde-LeBlanc et al. 2009.

Branch 9 is comprised of the membrane-bound Apolipoprotein N-acyltransferases that are involved in lipid post-translational modification of lipoproteins in all gram-negative bacteria (Vidal-Ingigliardi et al. 2007) and *Mycobacteria* (Tschumi et al. 2009). These reverse amidases catalyze the transfer of an acyl group from phospholipids to the free N-terminal cysteine of apolipoproteins, resulting in mature acylated lipoproteins (Buddelmeijer and Young 2010).

Branch 10 is comprised of amidases from mammals, plants, bacteria and fungi, where among other possible physiological functions, some of the characterized members function in the nitrogen and sulfur metabolic and salvage pathways (Cobzaru et al. 2011;Krasnikov et al. 2009a). The mammalian genome encodes two branch 10 enzymes, namely nitrilase-1 (Nit1) and nitrilase-2 (Nit2), whose expression in a wide range of tissues has been shown to have tumor suppression properties (Barglow et al. 2008;Krasnikov et al. 2009b;Semba et al. 2006) (Lin et al. 2007). The Nit2 enzyme from rat tissues has recently been identified as an ω-amidodicarboxylate amidohydrolase (ω-amidase; EC 3.5.1.3) (Meister et al. 1952;Meister et al. 1955), which catalyzes the deamidation of α-ketoglutaramate (αKGM) and α-ketosuccinamate (αKSM) to α-ketoglutarate and oxaloacetate respectively, where αKGM and αKSM are *in vivo* metabolites that are generated through transamination of glutamine and asparagine respectively (Krasnikov et al. 2009a). The Nit2 ω-amidase activity is important in various processes like in the salvage of α-keto acids of essential amino acids in mammals, bacteria and plants (Cooper 2004;Cooper and Meister 1981), in the synthesis of aminocyclitol antibiotics like gentamycin and neomycin in bacteria (Lucher et al. 1989), in photorespiration in higher plants (Ta et al. 1985) and in nicotine catabolism in some gram-positive soil bacteria (Cobzaru et al. 2011). A schematic representation of a methionine salvage bi-cycle pathway involving glutamine transamination in mammals, plants and bacteria is presented in figure 1.5.

N-carbamoylputrescine amidohydrolases from branch 11 of the nitrilase superfamily are found in plants and bacteria only, where they are involved in the biosynthesis of polyamines (low molecular mass polycations including putrescine, spermidine and spermine) that are important for optimal cell growth (Nakada and Itoh 2003;Piotrowski et al. 2003).

Enzymes from branches 12 and 13 have unknown enzymatic activities, although branch 12 enzymes are predicted to have protein substrates due to their fusion with an amino-terminal acyl-transferase domain (Brenner 2002;Pace and Brenner 2001).

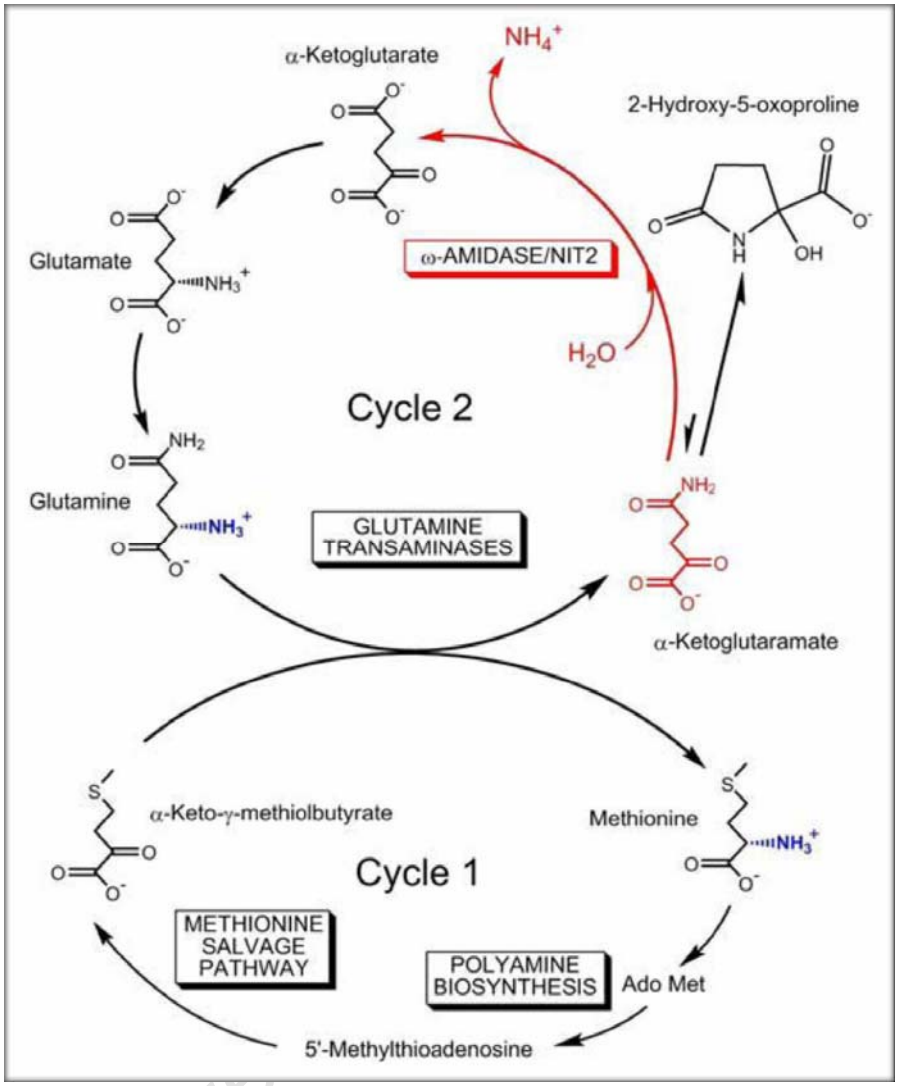


Fig. 1.5: A bi-cycle pathway for methionine (sulfur and methyl groups) recovery from polyamines. It involves transamination of α-keto-γ-methiolbutyrate with glutamine and the hydrolysis of αKGM by an ω-amidase. The figure was taken from Krasnikov et al. 2009a.

1.1.3 Applications of the nitrilase superfamily enzymes

In the past two decades, biotechnological industries have been exploring the enzymatic hydrolysis of nitriles and amides as a means of producing a broad spectrum of useful amides, amines, amidines, carboxylic acids, esters, drug intermediates and pharmaceuticals (Banerjee et al. 2002;Fournand and Arnaud 2001). This has been driven by attempts to find alternatives to traditional acid- or base-catalyzed amide and nitrile hydrolytic methods, which otherwise require harsh conditions of temperature and pH, and hence are incompatible with the labile structures of many industrially relevant compounds. These processes introduce unwanted by-

products, decrease product yield and increase the overall production costs (Banerjee et al. 2002;Fournand and Arnaud 2001). The observed chemo-, regio-, and enantio-selective properties (Banerjee et al. 2002;Yamamoto et al. 1990;Yamamoto et al. 1991) of these enzymes can be utilized to produce enantio-pure products. A good example of these is the bacterial D-stereospecific N-carbamyl-D-amino acid amidohydrolase that is employed in the second step of enzymatic production of optically active D-amino acids, which serve as useful intermediates in the production of physiologically active peptides and β-lactam antibiotics like cephalosporins and penicillins (Nakai et al. 2000;Olivieri et al. 1981). The nitrile hydratase from *Rhodococcus rhodochrous* J1 has been utilized by the Nitto Chemistry Industry Company Ltd in Japan to produce an average of 30 000 tons of acrylamide annually (Yamada and Kobayashi 1996), and the use of a nitrile hydratase-amidase couple in *Rhodococcus* sp. R312 strain to produce adipic acid from adiponitrile has also been reported (Moreau et al. 1993). Adipic acid is one of the raw materials in the manufacture of nylon-6:6.

Most nitrile compounds are reported to be highly toxic, carcinogenic and mutagenic (Pollak et al. 1991) and as industrial wastewater contaminates the environment, these toxic compounds pose a danger to both humans and animals. The use of micro-organisms containing nitrile-degrading enzymes would constitute a cost-effective way of detoxifying the environment. For example, the use of a mixed culture of bacteria containing nitrilases, nitrile hydratases and amidases to biodegrade acrylonitrile-containing effluent from acrylonitrile-manufacturing industries has been reported (Wyatt and Knowles 1995). Several soil micro-organisms have also been reported that degrade nitrile-containing herbicides in the soil, ensuring that these herbicides do not accumulate in foods, where they could result in disease conditions in humans (Freyssinet et al. 1996). An example of these is the soil bacterium, *Agrobacterium radiobacter,* which has the potential of degrading bromoxynil herbicide (Muller and Gabriel 1999).

Amidases from families that are only found in prokaryotes are potential drug targets; these include the branch 2 amidases from *Helicobacter pylori* and the branch 7 glutamine-dependent NAD^+ synthetases ($NADsyn^{gln}$). *H. pylori* is a human gastric pathogen that colonizes the mucous membrane of the stomach resulting in various stomach ailments including ulcers (Bury-Mone et al. 2003). *H. pylori* genome encodes two paralogous branch 2 amidases, AmiE (aliphatic amidase) and AmiF (formamidase) that are thought to play the role of protecting the bacterium from the harsh acidic environment of the stomach through ammonia production (Bury-Mone et al. 2003;Skouloubris et al. 1997;Skouloubris et al. 2001). The $NADsyn^{gln}$

enzyme functions in the NAD^+ co-factors biosynthetic pathway, and has been shown to be essential for cell growth and viability in pathogenic bacteria including *Mycobacterium tuberculosis* (MTB) (Sassetti et al. 2003). Since some NAD^+ recycling pathways in humans are not dependent on NAD^+ synthetase (Tempel et al. 2007), the MTB NAD^+ synthetase is an attractive target as the inhibitor selectivity issues may not be of so much concern.

1.1.4 The structural fold and oligomeric forms of amidases

The monomeric subunits of nitrilase superfamily members exhibit a four-layered α-β-β-α architectural fold as demonstrated by the *Geobacillus pallidus* RAPc8 amidase (RAPc8 amidase) subunit structure (Kimani et al. 2007) in figures 1.6 (A) and (B). The monomeric subunits associate across a conserved interface ('A-surface' in the notation of Sewell et al. 2005), resulting in an eight-layered α-β-β-α:α-β-β-α dimeric structure (Fig. 1.6 (C)) that is conserved in all known crystal structures and constitutes the basic building unit in the oligomeric enzymes with more than two subunits. An exception to this structural architecture is the glutaminase domain of the NAD^+ synthetase from *Cytophaga hutchinsonii* (PDB id, 3ilv; Palani et al. 2009), the biological unit of which is indicated to be a homodimer, but one that is formed by subunits that interact in a manner that does not form the conserved dimer interface between the glutaminase domains.

All characterized nitrilase superfamily amidases form active homo-oligomeric complexes in solution that include dimers, tetramers, hexamers and octamers, with each subunit having an active site pocket within which a Cys, Glu and Lys (CEK) catalytic triad is located. The branch 1 nitrilases on the other hand form elongated helical structures of varying lengths (Sewell et al. 2005). Twenty one crystal structures of the nitrilase superfamily amidases are available in the protein data bank (PDB; (Berman et al. 2000)), 10 of which are unique and the rest 11 are for mutants with or without substrate compounds, or structures at different resolutions. The oligomeric forms of the 10 unique amidase structures are summarized in table 1.1 and the quaternary structures for some of them are rendered in figure 1.7.

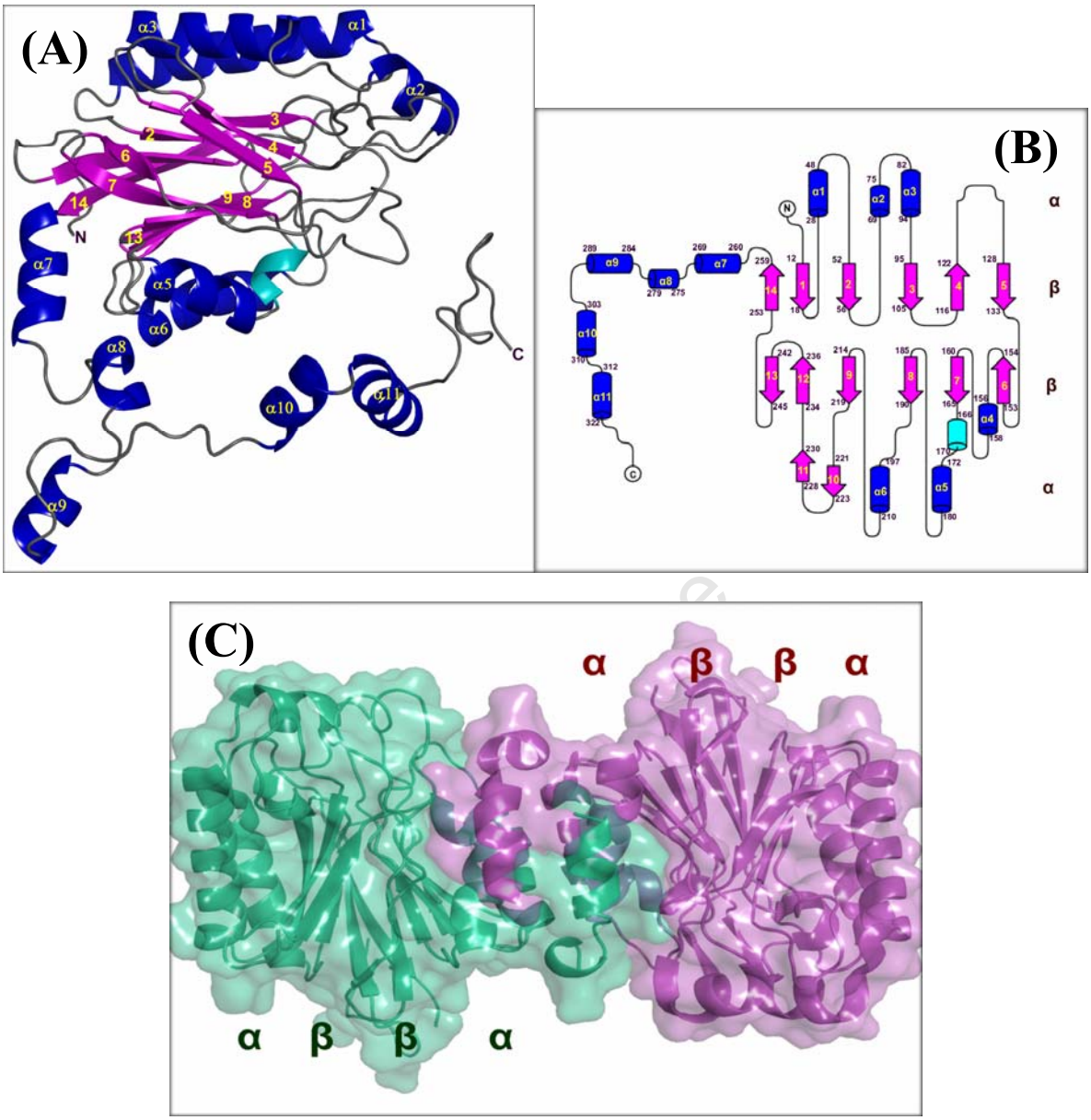
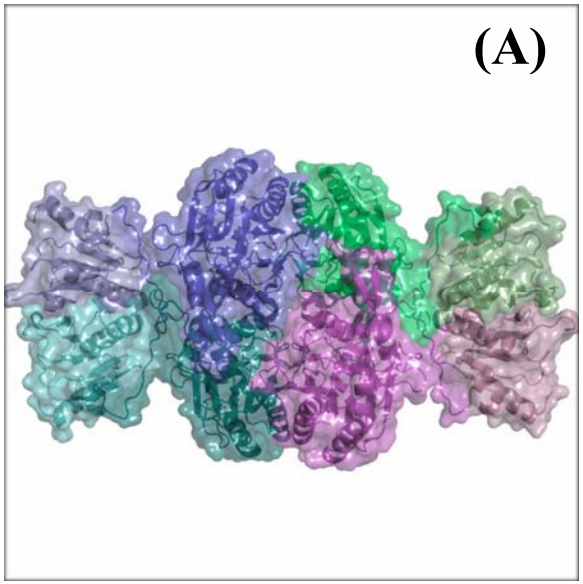


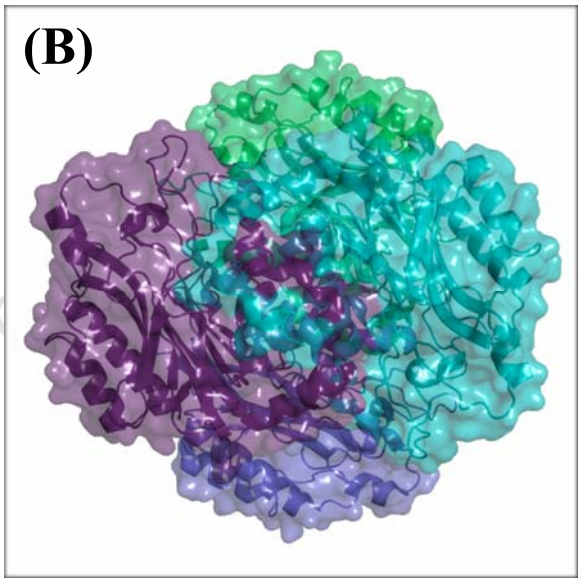
Fig. 1.6: The structural fold of the nitrilase superfamily enzymes. **(A)** A cartoon representation of the monomeric structure of RAPc8 amidase (PDB id, 2plq; (Kimani et al. 2007)) showing the 4-layered α-β-β-α sandwich fold. **(B)** The topology diagram of the secondary structure elements in the RAPc8 amidase subunit. The structural elements are coloured in the same colour scheme as the cartoon structure in panel (A), and they clearly show the α-β-β-α sandwich fold of the monomer. **(C)** Transparent surface and cartoon representation of the dimeric structure of the *Nesterenkonia* sp. amidase (PDB id, 3hkx; (Nel et al. 2011)) clearly showing the 8-layered α-β-β-α:α-β-β-α sandwich structure that is formed upon dimerization of monomeric subunits across the 'A-surface'. The monomers are coloured in different colours, and they are viewed down the 2-fold axis of the dimer interface. The pictures in panels (A) and (B) were taken from Kimani et al. 2007.

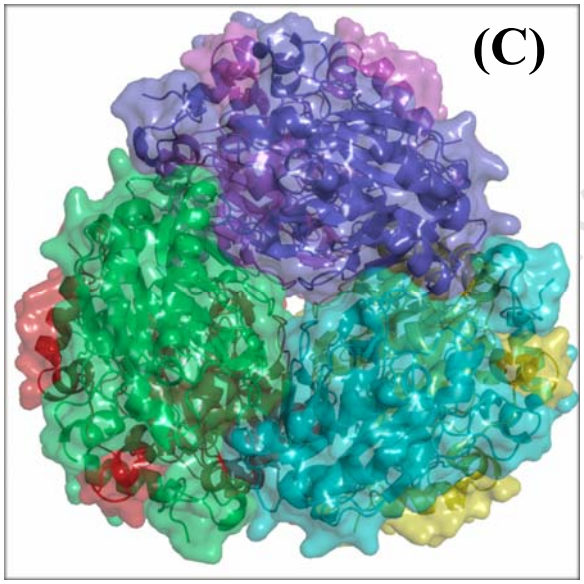
Table 1.1: Details of the 10 unique nitrilase superfamily amidases whose crystal structures are available in the PDB to-date.

PDB ID	Enzyme	Species	Chain length (Number of amino acids)	Nitrilase superfamily branch	Quaternary structure	Reference
1ems	Nit-Fragile histidine triad fusion protein (Nitfhit)	*Caenorhabditis elegans*	440	10 - Nit and Nitfhit	Homotetramer	Pace et al. 2000
1erz	N-carbamyl-D-amino acid amidohydrolase (DCase)	*Agrobacterium* sp.	303	6 - Carbamylases	Homotetramer	Nakai et al. 2000
2dyu	Formamidase (AmiF)	*Helicobacter pylori*	334	2 - Aliphatic amidases	Homohexamer	Hung et al. 2007
2plq	Aliphatic amidase (AmiE)	*Geobacillus pallidus* RAPc8	348	2 - Aliphatic amidases	Homohexamer	Kimani et al. 2007
2uxy	Aliphatic amidase (AmiE)	*Pseudomonas aeruginosa*	341	2 - Aliphatic amidases	Homohexamer	Andrade et al. 2007
2vhh	β-alanine synthase	*Drosophila melanogaster*	405	5 - β-Ureidoproprionases	Homooctamer[a]	Lundgren et al. 2008
2wlv	Nitrilase-2	*Mus musculus*	276	10 - Nit and Nitfhit	Homodimer	Barglow et al. 2008
3dla	Glutamine-dependent NAD^+ synthetase	*Mycobacterium tuberculosis*	680	7 - Prokaryotic Gln-dependent NAD^+ synthetase	Homooctamer	LaRonde-LeBlanc et al. 2009
3hkx	Amidase	*Nesterenkonia* sp.	283	2 - Aliphatic amidase[b]	Homodimer	Nel et al. 2011
3ilv	Glutamine-dependent NAD^+ synthetase	*Cytophaga hutchinsonii*	634	7 - Prokaryotic Gln-dependent NAD^+ synthetase	Homodimer[c]	Palani et al. 2009

[a]This enzyme forms various oligomeric complexes in solution, but only the octameric form was observed in the crystal.
[b]Classification of this amidase as an aliphatic amidase is tentative, and it is based on the currently known *in vitro* substrate specificity profile.
[c]The quaternary structure of this NAD^+ synthetase is not fully confirmed; a monomeric structure is present in the asymmetric unit of the $P2_12_12_1$ unit cell, while the biological unit is indicated to be a dimer in the PDB. The structure has not been published and the biochemical characterization of this enzyme has not been reported.







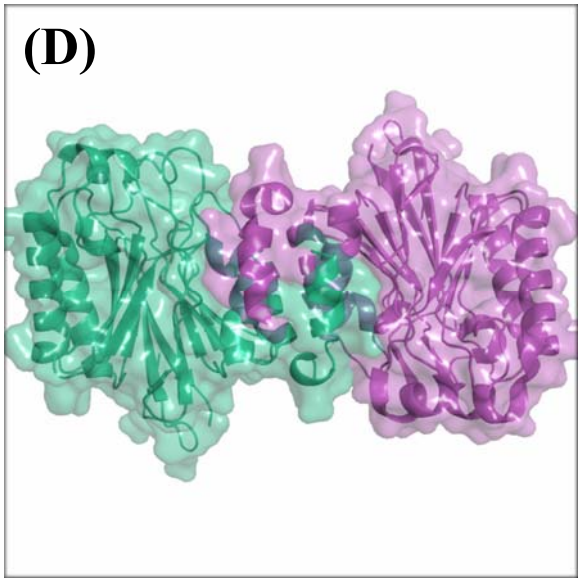
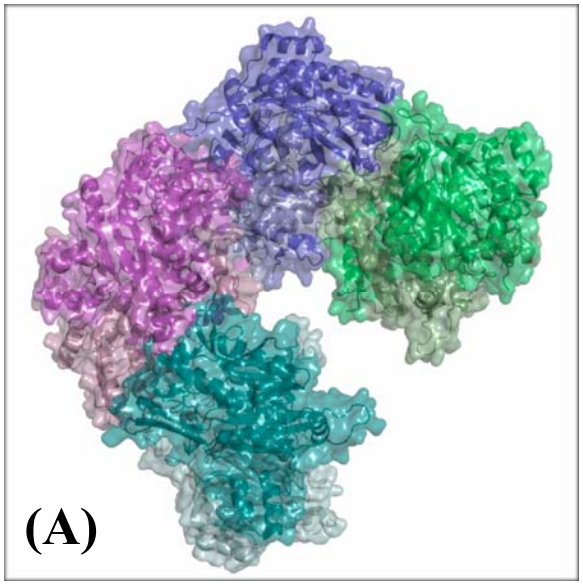
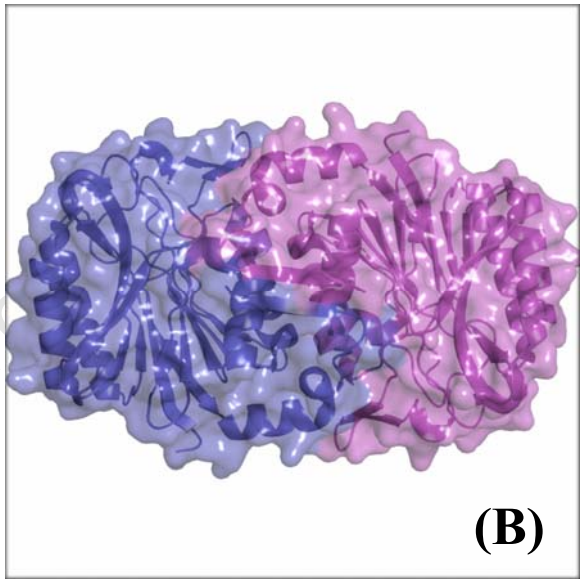
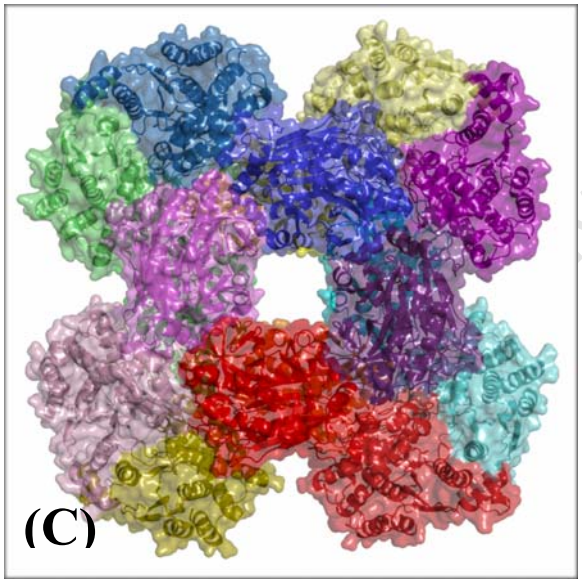


Fig. 1.7 (I): Cartoon and transparent surface representation of quaternary forms of the nitrilase superfamily amidase crystal structures. **(A)** The homotetrameric structure of the *C. elegans* Nitfhit protein (PDB id, 1ems). The middle part is the Nit domain tetramer (dimer of dimers; each monomer in a different color) that is attached to two dimers of the Fhit domain on each side. **(B)** The homo-tetrameric structure of the *Agrobacterium* sp. N-carbamoyl-D-amino acid amidohydrolase (DCase; PDB id, 1erz). The structure is made of a dimer of dimers and each monomer is rendered in a different color. **(C)** The homo-hexameric structure of the *G. pallidus* RAPc8 aliphatic amidase (RAPc8 amidase; PDB id, 2plq), with each subunit in a different color. The structure is made of a trimer of dimers that are arranged around a 3-fold axis. The 2-fold dimer axes are perpendicular to the 3-fold axis, giving rise to a complex with D3 point-group symmetry. The oligomeric structures of the formamidase from *Helicobacter pylori* (PDB id, 2dyu) and the aliphatic amidase from *Pseudomonas aeruginosa* (PDB id, 2uxy) are highly homologous to the RAPc8 amidase structure. **(D)** The homodimeric structure of the *Nesterenkonia* sp. amidase (NitN; PDB id, 3hkx).







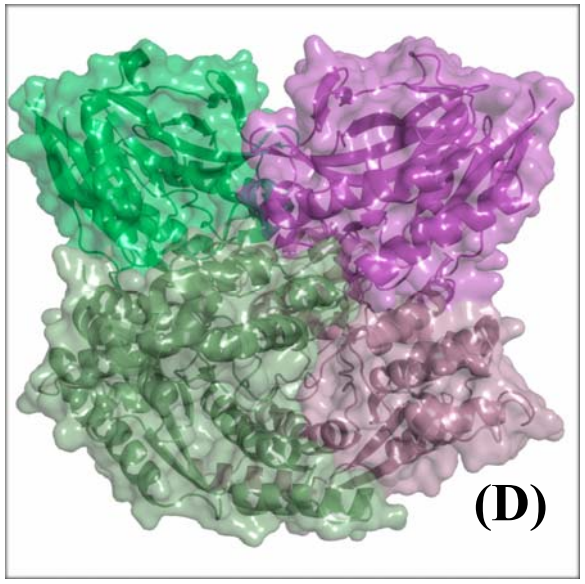


Fig. 1.7 (II): Cartoon and transparent surface representation of quaternary forms of the nitrilase superfamily amidase crystal structures. **(A)** The homooctameric structure of β-alanine synthase from *Drosophila melanogaster* (PDB id, 2vhh). The four dimers in the octamer are colored differently, with the monomers in each dimer having different shades of the same color. **(B)** The dimeric structure of Nitrilase-2 ω-amidase from *Mus musculus* (PDB id, 2wlv). **(C)** The homooctameric structure of the glutamine-dependent NAD^+ synthetase from *Mycobacterium tuberculosis* (MTB $NADsyn^{gln}$; PDB id, 3dla). The four dimers of the glutaminase domain form the central cube, while the four dimers of the synthetase domain are attached at the four corners, resulting in a square-like structure. **(D)** The homodimeric structure of the *Cytophaga hutchinsonii* glutamine- dependent NAD^+ synthetase (PDB id, 3ilv). The glutaminase domains of each monomer are rendered in green and violet on the top side of the molecule, while the attached synthetase domains are colored in pale green and pale violet in the bottom part of the picture respectively. This is the only structure where the nitrilase (glutaminase) domains do not form the conserved dimer interface that is observed in other nitrilase superfamily members.

The formation of the oligomeric structures is obligatory for enzymatic activity in amidases, and this is clearly demonstrated by the crystal structures of the RAPc8 amidase and the *Mycobacterium tuberculosis* glutamine-dependent NAD^+ synthetase (MTB $NADSyn^{gln}$) as examples. In the RAPc8 amidase hexameric structure (Fig. 1.8), the interactions at the conserved dimer interfaces (A-B, C-D, E-F in figure 1.8 (B)) are important for the formation of the active site (Fig. 1.8 (C)), as the interacting helices at the interface are linked to a short 3_{10}-helix that supports the catalytic cysteine (Kimani et al. 2007). This arrangement is unique in all dimeric units of the nitrilase superfamily amidases and it explains why monomeric subunits lack enzymatic activity (Thuku et al. 2009). The interface that is perpendicular to A-B and A-F (Fig. 1.8 (B)) is important for docking a flexible loop that bears the second catalytic glutamate residue into the active site, and may explain why the branch 2 aliphatic amidases need to be hexameric to be active.

The homooctameric quaternary structure of the MTB $NADsyn^{gln}$ is arranged in such a way that, two stacked rings of four glutaminase domain dimers form a central ring (cube), which is attached to four dimers of the synthetase domain at the four corners resulting in a square-like structure (Fig. 1.7 (II)-(C) and Fig. 1.9 (A)). In this arrangement, each glutaminase domain contacts the synthetase domains of two adjacent subunits (LaRonde-LeBlanc et al. 2009), and as a consequence, access to the active sites of the glutaminase domains is restricted. A glutamine tunnel that is formed by residues of glutaminase and synthetase domains from different subunits exists (Fig. 1.9 (B) and (C)) and it shuffles L-glutamine from the surface of the molecule to the glutaminase active site (LaRonde-LeBlanc et al. 2009). This structural arrangement may explain why the oligomeric structure is important for activity.

Unlike in other multi-domain glutamine amide transfer (GAT) enzymes where ammonia is shuffled between glutaminase and synthetase domains within the same polypeptide (Huang et al. 2001;Nakamura et al. 2006;Raushel et al. 2003), a narrow inter-subunit ammonia tunnel exists in the MTB $NADsyn^{gln}$ structure (Fig. 1.9 (B) and (C)) that connects glutaminase active site to synthetase active site from different subunits (LaRonde-LeBlanc et al. 2009). This implies that the ammonia produced by a glutaminase domain in one subunit is used by a synthetase domain in a different subunit. These observations emphasize the importance of the quaternary structure for the activity of $NADsyn^{gln}$ enzymes.

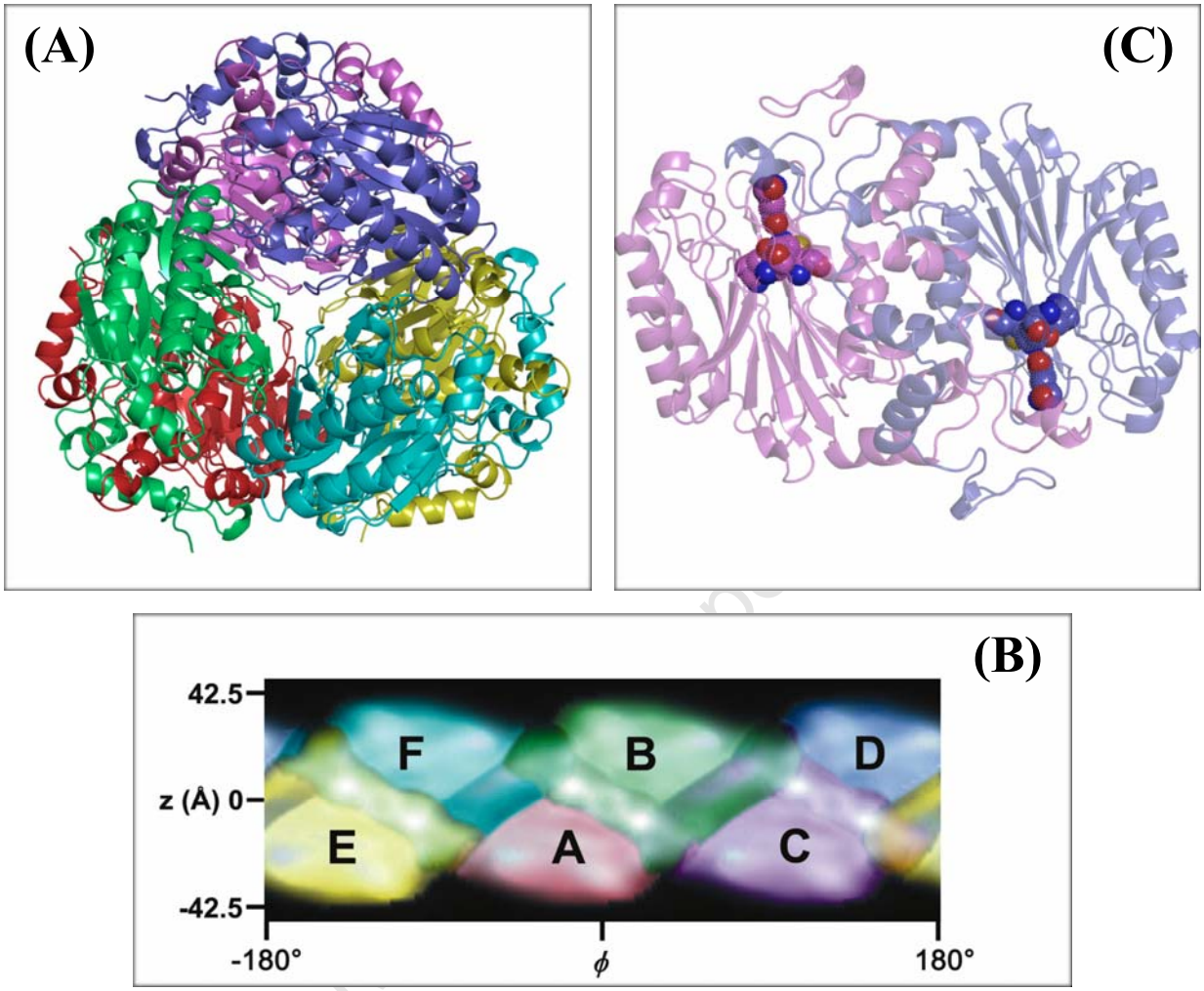


Fig. 1.8: The interfaces in the homohexameric structure of the RAPc8 amidase (PDB id, 2plq) and their relationship with the active site pockets of the enzyme. **(A)** The cartoon representation of the RAPc8 amidase hexamer as viewed down the 3-fold axis. Each monomer is colored differently. **(B)** The cylindrical projection of the hexamer, with the cylinder axis and the three fold axis aligned. The six subunits are labeled A-F and the coloring corresponds to the colors of the subunits in the cartoon rendition in panel (A). The conserved dimer interfaces link subunits A-B, C-D and E-F. **(C)** A cartoon rendition of one of the dimers (C-D) from the hexameric structure. The spheres are atoms of the residues making up the catalytic tetrad and indicate the location of the active sites in each monomer with respect to the conserved dimer interface. The formation of the dimer interface is important for the formation of the active pocket and the activity of the enzyme.

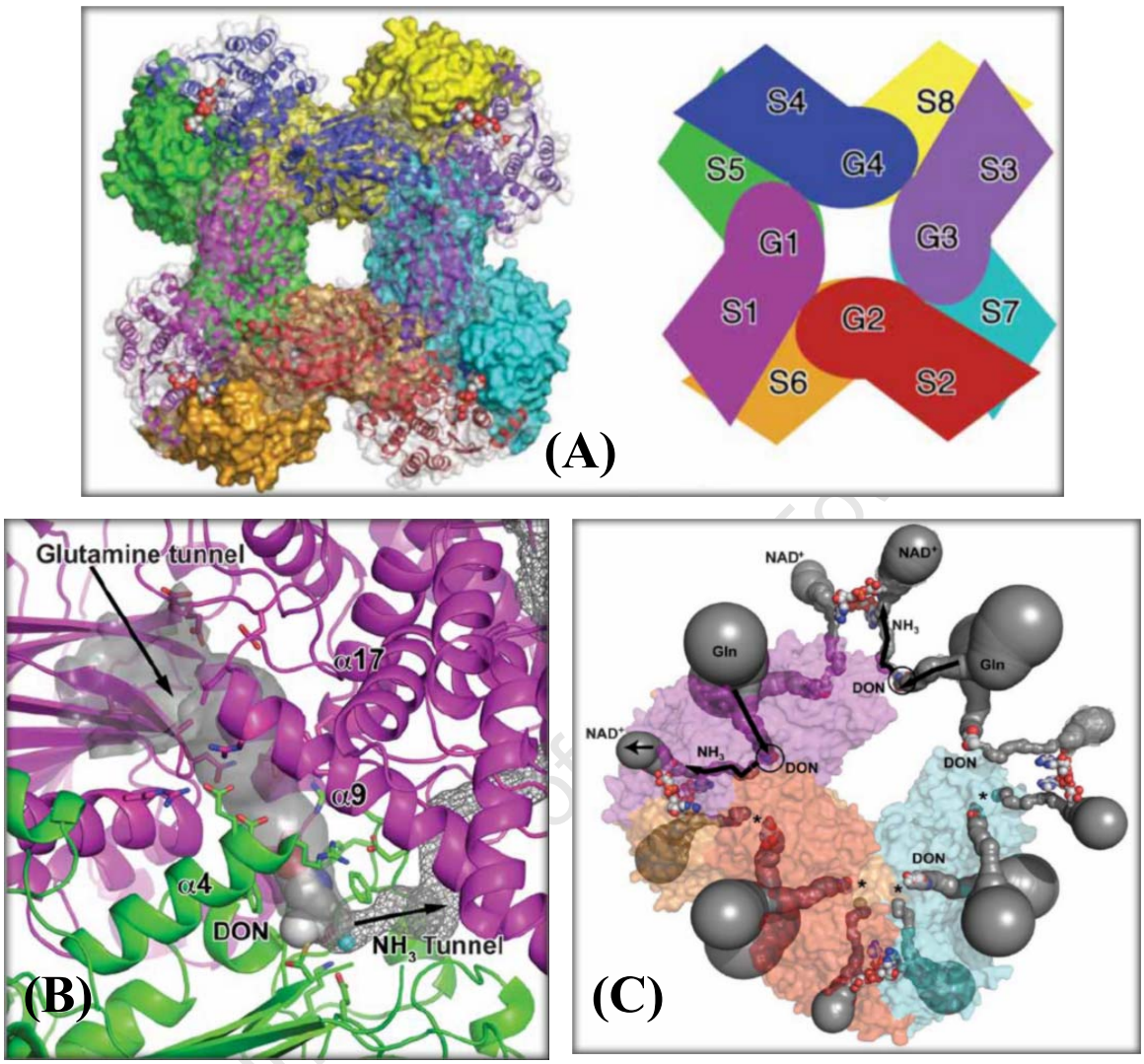


Fig. 1.9: Oligomeric structure and inter-subunit glutaminase and ammonia tunnels in the MTB NADsyn^{gln} (PDB id, 3dla). **(A)** A surface and cartoon representation of the homooctameric structure of MTB NADsyn^{gln}, showing the complex arrangement of domains and subunits in the active complex. Each subunit is colored differently. The ring-like structure in the centre is comprised of four glutaminase domain dimers, while four synthetase domain dimers are placed on the corners of what eventually looks like a square. **(B)** A cartoon representation of parts of two subunits showing the intersubunit glutaminase (grey surface) and ammonia (grey mesh) tunnels. The two subunits are coloured green (showing glutaminase active site with the bound inhibitor, 6-diazo-5-oxo-L-norleucine (DON)) and magenta. **(C)** Surface representation of the 8 inter-subunit tunnels in the homo-octameric complex. Four of the subunits are shown as transparent surface, while the other four are not shown. All figures were taken from LaRonde-LeBlanc et al. 2009.

1.2 Structural basis for substrate specificity in amidases

The emergence of structural data in the past decade has complemented biochemical data in the understanding of enzymatic activity and the basis for substrate specificity in some nitrilase superfamily amidases. A few crystal structures of amidases from branches 2 and 5 (aliphatic amidases and carbamylases) with non-covalently bound substrate molecules are now available, and they have revealed how the two classes of substrates (short aliphatic amides and carbamyl derivatives) interact with the respective enzymes and also provided the basis for substrate preference in these enzymes. Various amidase crystal structures and models with covalent adducts at the catalytic cysteine have also greatly contributed to the pool of knowledge that is now available on the activity of amidases. The following section describes activity and substrate specificity and selectivity in amidases as supported by structural data.

1.2.1 Activity and substrate specificity in branch 2 amidases

1.2.1.1 Enzymatic activity

Prokaryotic aliphatic amidases (EC 3.5.1.4) are highly homologous enzymes belonging to branch 2 of the nitrilase superfamily, and they constitute the most extensively studied amidase group in terms of substrate specificity and selectivity *in vitro*. These enzymes catalyze the hydrolysis of short-chain aliphatic amides to their corresponding acids and ammonia. In addition, aliphatic amidases are often able to transfer the acyl moiety of an amide to hydroxylamine to form hydroxamates (Clarke 1970;Fournand et al. 1998b;Maestracci et al. 1986;Thiery et al. 1986), as well as to hydrazine resulting in the formation of hydrazides (Fournand et al. 1997). While acyl transfer is a common reaction in most aliphatic amidases *in vitro*, amide hydrolysis would be the dominant reaction *in vivo*. All characterized group 2 amidases are active as homohexameric complexes in solution.

1.2.1.2 *In vitro* substrate specificity

Biochemical characterization and activity analyses of the aliphatic amidases from *Pseudomonas aeroginosa* (Clarke 1970), *Brevibacterium* sp. R312 (Maestracci et al. 1984;(Soubrier et al. 1992), *Rhodococcus* sp. R312 (Fournand et al. 1997), *Helicobacter pylori* AmiE (Skouloubris et al. 1997), *Bacillus stearothermophilus* BR388 (Cheong and Oriel 2000) and *Geobacillus pallidus* RAPc8 sp. (Makhongela et al. 2007) revealed a similar substrate

specificity profile in the branch 2 amidases. Acetamide (ACE), acrylamide (ACR) and propionamide (PMD) are the most rapidly and efficiently hydrolyzed substrates, with PMD being the best substrate in all cases. Aliphatic amidases are not capable of hydrolyzing aromatic or branched substrates, and they have little or no activity on very short- (less than 2 carbon atoms) and long-chain (greater than 4 carbon atoms) linear amides (Makhongela et al. 2007;Nel et al. 2011). The substrate specificity similarities are not surprising given the observed sequence homologies in branch 2 amidases, and they are probably an indication of a high conservation of the substrate binding pocket in these enzymes. Studies of the acyl transferase activity also revealed ACE, PMD and ACR to be the best acyl donors to hydroxylamine; however, ACE is the most efficient acyl donor in all characterized amidases (Clarke 1970;Fournand et al. 1997;Makhongela et al. 2007). The difference in substrate preference between the two amidase activities is rather surprising as the mechanism for amide hydrolysis and acyl transfer from is proposed to be the same (Fournand et al. 1997;Maestracci et al. 1986).

In addition to amiE aliphatic amidase (Skouloubris et al. 1997), *Helicobacter pylori* has a second amidase (AmiF) that is believed to be a product of gene duplication (Skouloubris et al. 2001). Unlike AmiE that has activity on short-chain aliphatic amides (propionamide, acetamide and acrylamide) but not on formamide (FMD), AmiF is only capable of hydrolyzing FMD and was thus reported as the first formamidase, belonging to a sub-branch of aliphatic amidases (EC 3.5.1.49) (Skouloubris et al. 2001). The two *H. pylori* amidases (AmiE and AmiF) amidases are believed to be paralogous proteins that have evolved to achieve enzymatic specialization after ancestral gene duplication, and their production is thought to be regulated in order to maintain intracellular nitrogen balance in *H. pylori* along with urease and arginase (Skouloubris et al. 2001). AmiE and AmiF are both homohexameric in solution, and despite their known overall sequence similarities, the marked differences in substrate specificity profiles are indicative of possible structural differences in their substrate binding pockets.

1.2.1.3 The basis for *H. pylori* formamidase strict preference for formamide

In 2007, Hung and colleagues reported the crystal structure of the catalytic cysteine mutant (C166S) of *H. pylori* formamidase AmiF in complex with the substrate, formamide (FMD). Structural analysis revealed a very small substrate binding pocket, with the bulky side chains of Trp137 and Tyr192 forming a tight wall around the substrate (Hung et al. 2007) (Fig. 1.10 (A) and (B)). The limited space in the active site pocket is responsible for discriminating against

amide substrates with greater than one carbon atom and explains the strict specificity of amiF formamidase for FMD only.

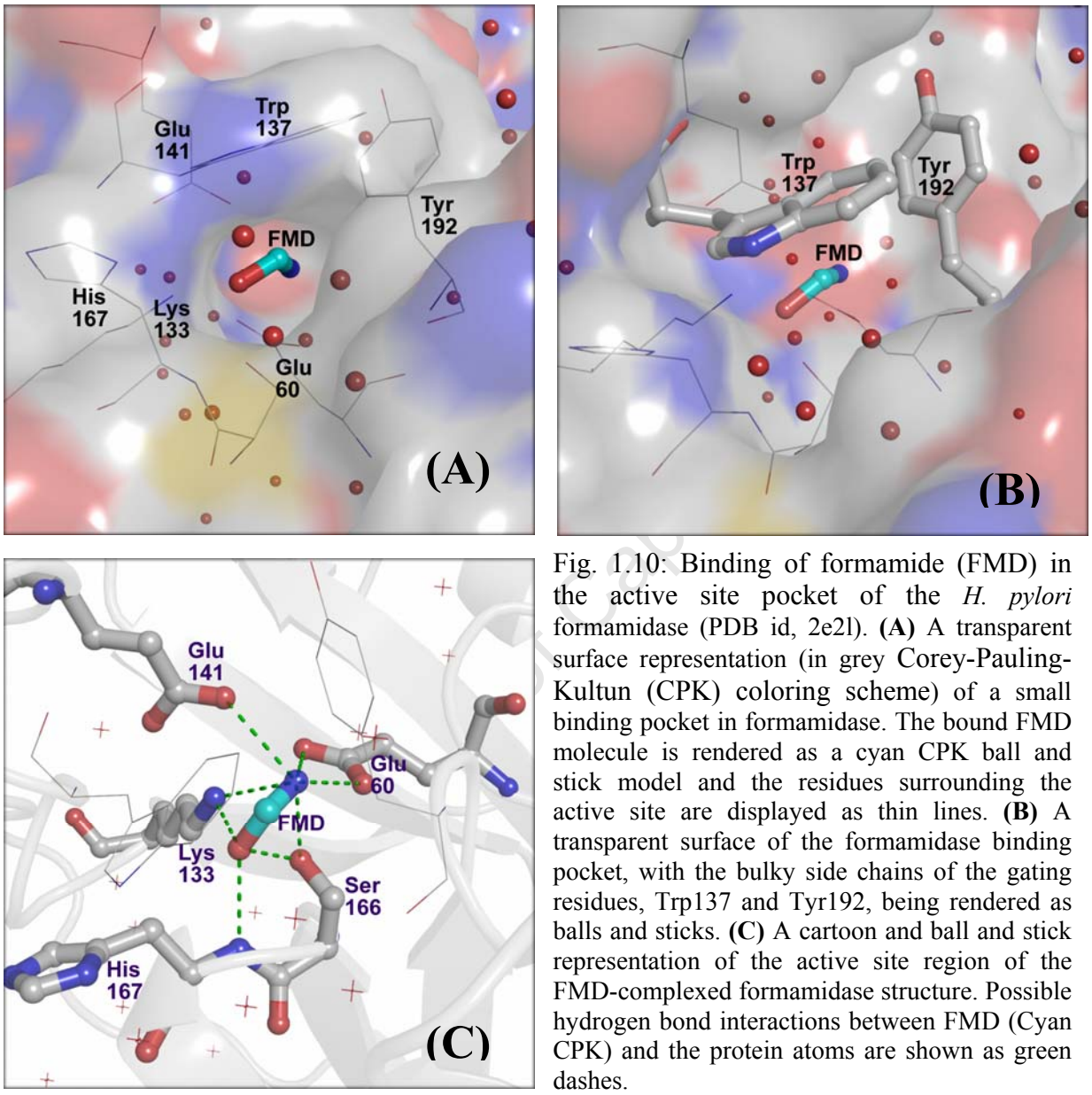


Fig. 1.10: Binding of formamide (FMD) in the active site pocket of the *H. pylori* formamidase (PDB id, 2e2l). **(A)** A transparent surface representation (in grey Corey-Pauling-Kultun (CPK) coloring scheme) of a small binding pocket in formamidase. The bound FMD molecule is rendered as a cyan CPK ball and stick model and the residues surrounding the active site are displayed as thin lines. **(B)** A transparent surface of the formamidase binding pocket, with the bulky side chains of the gating residues, Trp137 and Tyr192, being rendered as balls and sticks. **(C)** A cartoon and ball and stick representation of the active site region of the FMD-complexed formamidase structure. Possible hydrogen bond interactions between FMD (Cyan CPK) and the protein atoms are shown as green dashes.

In the enzyme-substrate complex, catalytic residues (Ser166 changed from Cys166, Glu60, Lys133 and Glu141) make key contacts with the substrate. The carbonyl oxygen atom of FMD is positioned by interactions with the side chain hydroxyl group of Ser166, side chain amino group of Lys133 and the backbone NH group of His167, while the amide amino group interacts with OE_1 atom of Glu60, OE_2 atom of Glu141, OG atom of Ser166 and NZ atom of

Lys133 (Fig 1.10 (C)). The observed polar contacts between the hydroxyl group of Ser166 and both the carbonyl oxygen and the amide amino group of FMD would not be present in the wild-type (WT) enzyme having the nucleophilic cysteine (Cys166) in this position. The positioning or orientation of FMD in the C166S mutant structure is likely to be biased.

1.2.1.4 The basis for short aliphatic amides preference by aliphatic amidases

The determination of the crystal structures of aliphatic amidases from *Geobacillus pallidus* RAPc8 (RAPc8 amidase; PDB id, 2plq) (Kimani et al. 2007) and *Pseudomonas aeruginosa* (PDB id, 2uxy) (Andrade et al. 2007) in 2007 was a major contribution towards understanding the strict specificity of these enzymes for short-chain (with 2 to 4 carbons) linear amides. These two amidases share over 90% sequence identity and are representative of a highly homologous group of enzymes that show catalytic preference for PMD, ACR and ACE for both hydrolysis and acyl group transfer to hydroxylamine activities *in vitro* (Clarke 1970;Makhongela et al. 2007).

Kimani and colleagues (2007) reported the binding cleft of RAPc8 amidase to be buried and relatively shallow, with a volume of 53.59 $Å^3$. Similar characteristics were described by Andrade and colleagues (2007) for the *P. aeruginosa* amidase binding pocket. The relatively small-sized substrate binding pocket supports the specificity of aliphatic amidases for short-chain aliphatic amides. The bulky side chain of Trp138 (Fig. 1.11 (A) and (B)) in both structures covers the entrance to the binding pocket almost completely, leading both groups to conclude that, access of the substrates to the active site is dependent on conformational shift of the loop where this gating residue is located. A model of the D-lactamide acyl-Cys166 thioester intermediate in the pocket of RAPc8 amidase places the carbonyl of the acyl intermediate in a small pocket near Lys134, with the bulk of the substrate acyl moiety extending up towards Trp138. A clash with Trp138 would limit the length of linear substrates to four carbons and would certainly not allow aromatic or branched substrates (Kimani et al. 2007).

The importance of Trp138 side chain as a determinant for substrate specificity in aliphatic amidases was confirmed by Karmali and colleagues (2001), who observed that, a substitution of this residue with either glycine or serine (W138G/S) in the *P. aeruginosa* amidase greatly reduced the affinity of the mutants to ACE and ACR and on the other hand allowed binding and hydrolysis of much bulkier/aromatic substrates - phenylacetamide and *p*-

nitrophenylacetamide (Karmali et al. 2001), that would otherwise not fit in the active site pocket of the WT enzyme. In addition to Trp138 side chain, Tyr192 and Met193 side chains also contribute to the size constraints that are observed in the catalytic pockets of these enzymes.

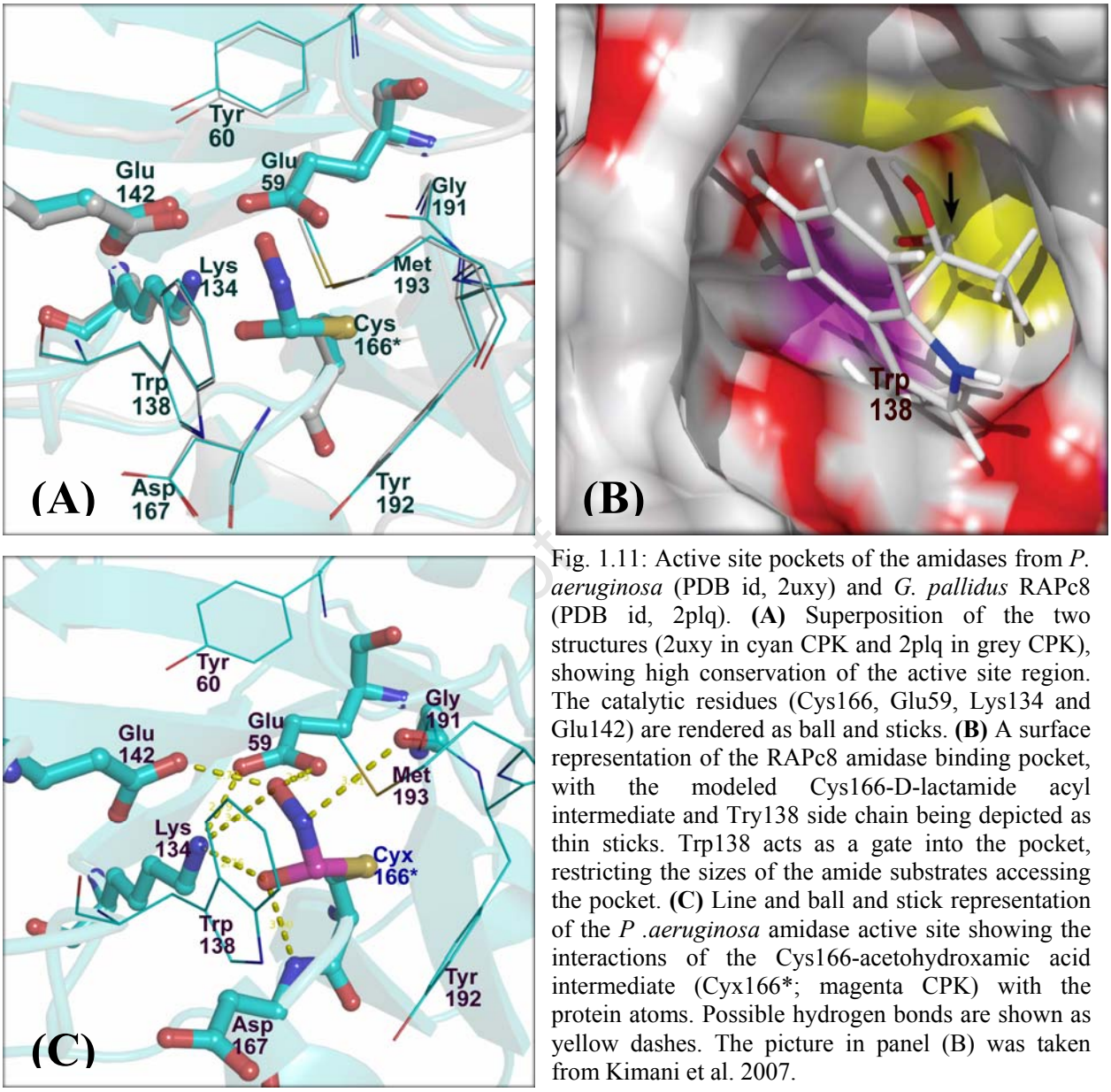


Fig. 1.11: Active site pockets of the amidases from *P. aeruginosa* (PDB id, 2uxy) and *G. pallidus* RAPc8 (PDB id, 2plq). **(A)** Superposition of the two structures (2uxy in cyan CPK and 2plq in grey CPK), showing high conservation of the active site region. The catalytic residues (Cys166, Glu59, Lys134 and Glu142) are rendered as ball and sticks. **(B)** A surface representation of the RAPc8 amidase binding pocket, with the modeled Cys166-D-lactamide acyl intermediate and Try138 side chain being depicted as thin sticks. Trp138 acts as a gate into the pocket, restricting the sizes of the amide substrates accessing the pocket. **(C)** Line and ball and stick representation of the *P .aeruginosa* amidase active site showing the interactions of the Cys166-acetohydroxamic acid intermediate (Cyx166*; magenta CPK) with the protein atoms. Possible hydrogen bonds are shown as yellow dashes. The picture in panel (B) was taken from Kimani et al. 2007.

1.2.1.5 The basis for substrate preference differences between amide hydrolysis and acyl transfer reactions

The structure of *P. aeruginosa* amidase had an unexpected acyl transfer tetrahedral reaction intermediate of acetohydroxamic acid covalently bound to the nucleophilic cysteine (Fig. 1.11 (C)). This reaction was thought to have occurred during the purification procedure as the amidase was being eluted off the affinity column (Epoxy-activated sepharose 6B-acetamide) using acetamide (acyl donor) and hydroxylamine (acyl acceptor) (Andrade et al. 2007). The acetohydroxamic acid-Cys166 tetrahedral intermediate adduct makes hydrogen bonds with the side chains of Glu59, Lys134 and Glu142, the main chain NH group of Asp167 and oxygen atom of Gly191 (Fig. 1.11 (C)). The involvement of the catalytic residues (Glu59, Lys134 and Glu142) in coordinating the covalently bound substrate intermediate is again observed in this complex.

The *P. aeruginosa* amidase tetrahedral reaction intermediate structure not only confirms the occurrence of the reaction intermediates during catalysis in the nitrilase superfamily amidases, but it also rationalizes the observed differences in substrate preference between amide hydrolysis and acyl transfer reactions in aliphatic amidases. Both PMD and ACE are good substrates for hydrolysis and acyl transfer reactions, but PMD is the best hydrolytic substrate while ACE is the best acyl donor to hydroxylamine (Clarke 1970;Fournand et al. 1997;Maestracci et al. 1986). Andrade and co-workers (2007) observed that, the methyl group of the acetohydroxamic acid intermediate is only 3.39 Å away from the CE_1 atom of Tyr192 at the bottom of the binding pocket. The acyl group of PMD, which has one extra carbon, would therefore induce larger conformational strain on the nucleophilic cysteine in order for propionehydroxamic intermediate to fit in the active site pocket (Andrade et al. 2007). Based on this argument, it is apparent why aliphatic amidases prefer ACE to PMD as acyl donors to hydroxylamine.

1.2.1.6 The basis for selectivity of D-enantiomers by RAPc8 amidase

The RAPc8 amidase is able to hydrolyze D-lactamide, while no activity is observed with the L-lactamide enantiomer (Makhongela et al. 2007). The modeling experiments of the reaction intermediates in the RAPc8 structure revealed the substrate binding cavity of this enzyme to be highly asymmetric. The hydroxyl group of the acyl-enzyme intermediate formed from L-lactamide would clash with the carboxyl of Glu142 whereas no clash occurred in the case of D-

lactamide (Fig. 1.11 (B)) (Kimani et al. 2007). The geometry of the active site therefore provides a satisfactory explanation for the observed enantioselectivity of *G. pallidus* RAPc8 amidase for D-enantiomers.

1.2.1.7 Alteration of substrate specificity in the *P. aeruginosa* amidase

The *P. aeruginosa* amidase enzyme has attracted a lot of attention because of the variations in its substrate specificity that can be generated through point mutations. This exploration was pioneered by Clarke and colleagues in the 1970s using inducible *P. aeruginosa* aliphatic amidase strains. In their experiments, various *P. aeruginosa* bacterial strains were subjected to selective pressure by being cultured in media containing various amides (usually those that are not substrates for the aliphatic amide) as the sole carbon source for growth. Those that managed to form colonies had their amidase AmiE gene isolated and checked for possible mutations, and any structural changes on the resultant amidase mutant were then analyzed. Through this approach, substrate specificity of this amidase has been extended through a variety of acquired point mutations to include longer aliphatic amides such as butyramide (Brown et al. 1969;Brown and Clarke 1970) and valeramide (Brown et al. 1969), as well as aromatic amides including acetanilide (*N*-phenylacetamide) (Brown and Clarke 1972) and phenylacetamide (Betz and Clarke 1972).

In the AI3 mutant strain of the bacteria, the T103I amidase mutation was found to be responsible for the amidase activity towards acetanilide (Brown and Clarke 1972). While this mutation is close to but not in the active site pocket, substitution of threonine with isoleucine may have introduced conformational changes that affected the active site pocket, allowing bulkier aromatic substrates to bind. Although urea and hydroxyurea are active site inhibitors of the *P. aeruginosa* wild-type amidase (Gregoriou and Brown 1979;Tata et al. 1994), additional mutations (R188C/H/L, Q190E and W144C) to the amidase gene of mutant AI3 strain resulted in its decreased affinity for urea (Tata et al. 1994). The mutated residues are in the neighborhood of some of the residues forming the active site pocket (Glu142 and Gly191, Fig. 1.11 (C)), indicating a possibility of conformational changes that alter affinity for urea and related compounds.

As already mentioned in the previous section, Trp138 is important as a determinant for substrate specificity in *P. aeruginosa* and *G. pallidus* RAPc8 amidases. In their more recent study on Thr103 and Trp138 mutants of *P. aeruginosa* amidase, Karmali and colleagues (2001)

had some interesting findings: A W138G mutant enzyme had activity on acetamide, acrylamide, phenylacetamide and *p*-nitrophenylacetamide; a double mutant amidase of W138G and T103I was only active on *p*-nitrophenylacetamide and phenylacetamide; while a T103I enzyme acted on *p*-nitroacetanilide and acetamide (Karmali et al. 2001). The derived mutants were found to be less thermostable than the WT enzyme. Increased structural instability was also observed in the double mutant (W138G and T103I) due to conformational changes that were detected by differences in monoclonal antibodies reactions between this mutant and the WT enzyme.

1.2.2 Activity and substrate specificity in branches 5 and 6 carbamylases

1.2.2.1 Activity and substrate specificity in branch 6 carbamylases

Branch 6 bacterial N-carbamyl-D-amino acid amidohydrolases (DCase; carbamylases) catalyze the hydrolysis of N-carbamyl-D-amino acids to the corresponding D-amino acids *in vitro* (Nakai et al. 2000). Although the biological function of bacterial DCases is still not known, they are of biotechnological importance as D-amino acids are useful intermediates in the preparation of β-lactam antibiotics (Nakai et al. 2000).

Biochemical characterization of the carbamylase from *Blastobacter* sp. revealed a broad substrate specificity profile, with the enzyme being able to hydrolyze various corresponding N-carbamyl derivatives of D-amino acids, such as leucine, isoleucine, valine, methionine, phenylalanine, tryptophan, alanine, threonine, serine, phenylglycine and norleucine; substrates with long or bulkier hydrophobic side chains were more preferred (Ogawa et al. 1993;Ogawa et al. 1994). However, the *Blastobacter* sp. DCase did not have any activity on N-carbamyl-L-amino acids, N-methyl derivatives such as N-carbamylsarcosine, N-formyl-amino acids, N-acetyl-amino acids or substrates lacking the carboxyl moiety such as ethylurea (Ogawa et al. 1993;Ogawa et al. 1994). A number of crystal structures of *Agrobacterium* sp. DCases are available (Chen et al. 2003;Hashimoto et al. 2003;Nakai et al. 2000), all of which have provided details of the structural basis for substrate specificity in this branch of amidases.

The bases for substrate specificity and selectivity in bacterial DCases

The binding pocket of the *Agrobacterium* sp. KNK712 WT DCase structure (PDB id, 1erz; (Nakai et al. 2000)) has a relatively large volume, with a number of sub-pockets as shown in figure 1.12 (A). This is consistent with the broad substrate specificity observed in bacterial DCases, as different sizes of D-amino acid side chains would need to be accommodated. Hashimoto and colleagues (2003) determined crystal structures of the C171A/V236A mutant of DCase from *Agrobacterium* sp. KNK712, in complex with N-carbamyl-D-methionine (PDB id, 1uf5), N-carbamyl-D-valine (PDB id, 1uf7) and N-carbamyl-D-phenylalanine (PDB id, 1uf8). In all these structures, two sets of active site residues bind to the carboxyl and carbamyl moieties of the substrates respectively, while the amino acid side chain moieties dock into the same region of the binding pocket as shown in figures 1.12 (B), (C) and (D). The carboxyl moiety of the substrates is positioned by a number of interactions involving Arg174, Arg175, Thr197 and Asn172, while the carbamyl moiety makes contacts with Glu46, Asn196, Lys126, Asn172 and Glu145 (Fig. 1.12 (B) and (C)). The two arginine residues (Arg174 and 175) play the role of substrate recognition in DCases though their positive charges, which provide a favorable field for attracting the carboxyl moiety of the substrate molecules to the binding pocket (Nakai et al. 2000). A substitution of the two Arg residues (Arg175 and Arg176) with alanine and lysine in the *Agrobacterium radiobacter* DCase resulted in mutants that lacked enzymatic activity (Chen et al. 2003). In addition to substrate recognition, the two arginines also play the role of positioning and orienting the substrates relative to the thiol group of the catalytic cysteine (Cys171) for nucleophilic attack (Nakai et al. 2000).

Asn172 is proposed to be the determinant of D-stereoselectivity in DCases as unfavorable van der Waals contacts would occur between its side chain and N-carbamyl-L-amino acid substrates (Nakai et al. 2000). The region of the pocket accommodating different amino acid side chain moieties of the substrates has a large volume (Fig. 1.12 (A) and (D)), which supports specificity for a wide variety of N-carbamyl D-amino acid derivatives by DCases and also explains the observed preference for substrates with long or bulkier hydrophobic side chains. From all the available substrate-complexed DCase crystal structures, it is clear that bacterial DCases only recognize the α-carboxyl and N-carbamyl moieties of the substrates. This provides a satisfactory explanation for the observed lack of activity on N-formyl-amino acids, N-acetyl-amino acids, N-methyl derivatives (such as N-carbamylsarcosine) and on substrates lacking the carboxyl moiety (such as ethylurea).

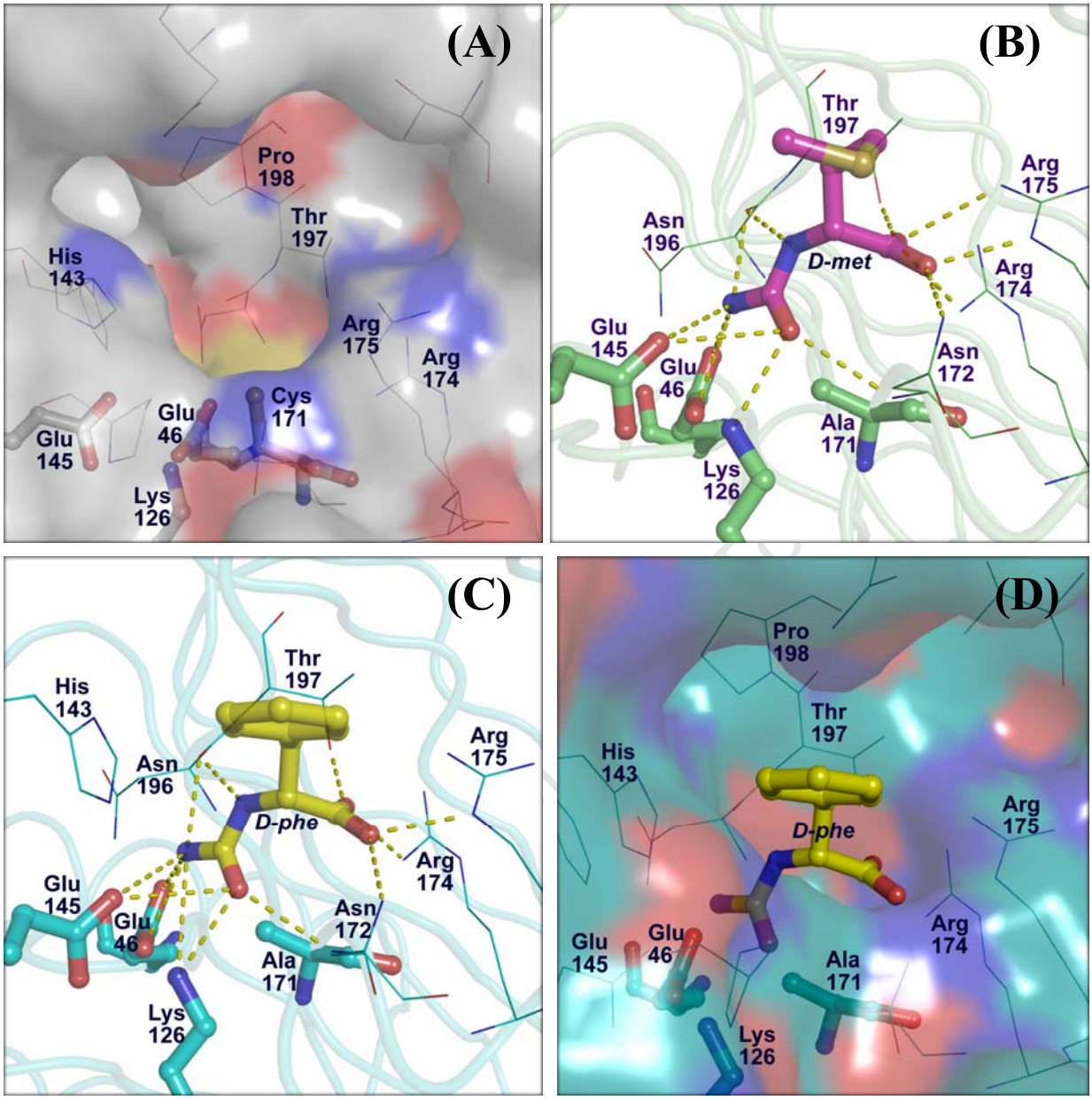


Fig. 1.12: Details of substrate binding in the DCase enzyme from *Agrobacterium* sp. KNK712. **(A)** A transparent surface representation of the active site pocket of the wild-type (WT) DCase enzyme (PDB id, 1erz). The pocket is large and deep with a number of sub-pockets. **(B)** A ball and stick representation of the active site environment of the C171A/V236A mutant of DCase (PDB id, 1uf5) with a bound N-carbamyl-D-methionine (D-met) substrate (in magenta CPK). Possible hydrogen bond interactions between the D-met carboxyl and carbamyl moieties and the protein atoms are shown as yellow dashes. **(C)** A ball and stick representation of the active site pocket of the C171A/V236A mutant of DCase (PDB id, 1uf8) with a bound N-carbamyl-D-phenylalanine (D-phe) substrate (yellow CPK). D-phe binds to the active site in the same manner as the D-met in panel B, and a similar hydrogen bonding network (yellow dashes) is observed. **(D)** A transparent surface representation of the active site pocket of the same structure in panel (C) (PDB id, 1uf8), with the D-phe substrate (yellow CPK) being rendered as a ball and stick model. The region of the pocket occupied by the side chain moieties is large, explaining the observed ability of DCases to bind a broad range of substrates.

1.2.2.2 Activity and substrate specificity in branch 5 β-alanine synthases

(A) Physiological function and activity

The crystal structure of β-alanine synthetase (βAS; EC 3.5.1.6) from *Drosophila melanogaster* (Lundgren et al. 2008) became available recently, providing the first structural insights into a member of the β-ureidopropionase branch of the nitrilase superfamily. βASs function in the reductive pyrimidine bases catabolic pathway in eukaryotes (Piskur et al. 2007) where they catalyze the hydrolysis of N-carbamyl-β-alanine and N-carbamyl-β-aminoisobutyrate to β-alanine and β-aminoisobutyrate respectively (Wasternack 1980). βAS amidases are therefore responsible for ensuring supply of β-alanines, which play different physiological roles in eukaryotes including neurotransmission in mammals (Kvalnes-Krick and Traut 1993;Traut and Jones 1996).

(B) Substrate specificity in β-alanine synthetases

Unlike bacterial DCases that exhibit broad substrate specificity, βASs are fairly specific with the only known physiological substrates being N-carbamyl-β-alanine and N-carbamyl-β-aminoisobutyrate. Differences in the active site pockets are therefore expected between the bacterial DCases and the eukaryotic βASs. A superposition of the βAS structure from *D. melanogaster* (PDB id, 2vhh) and the structure of C171A/V236A DCase mutant complexed with N-carbamyl-D-valine substrate (PDB id, 1uf7) (Fig. 1.13) highlights some key differences between the active site pockets in branch 5 and 6 amidases. The catalytic triad residues (Cys234, Glu120 and K197) in *D. melanogaster* βAS are in the same position and coordination as the corresponding residues in the DCase structure (Cys171, Glu46 and Lys126). Most of the residues involved in the binding of the carboxyl moiety of the DCase substrates (N-carbamyl-D-amino acids) are fairly conserved in the *D. melanogaster* βAS structure, implying that, substrate binding to βASs is likely to be achieved in a manner that is similar to that observed in DCases (Lundgren et al. 2008). The conservation in this region is supportive of its involvement in substrate recognition and positioning, as suggested by Nakai and coworkers (2000). Some differences include the side chain of Arg237 in *D. melanogaster* βAS, which is in a slightly different conformation relative to the corresponding residue (Arg174) in DCase, and His238 in *D. melanogaster* βAS, which functionally replaces Arg175 in DCase (Fig. 1.13 (A)).

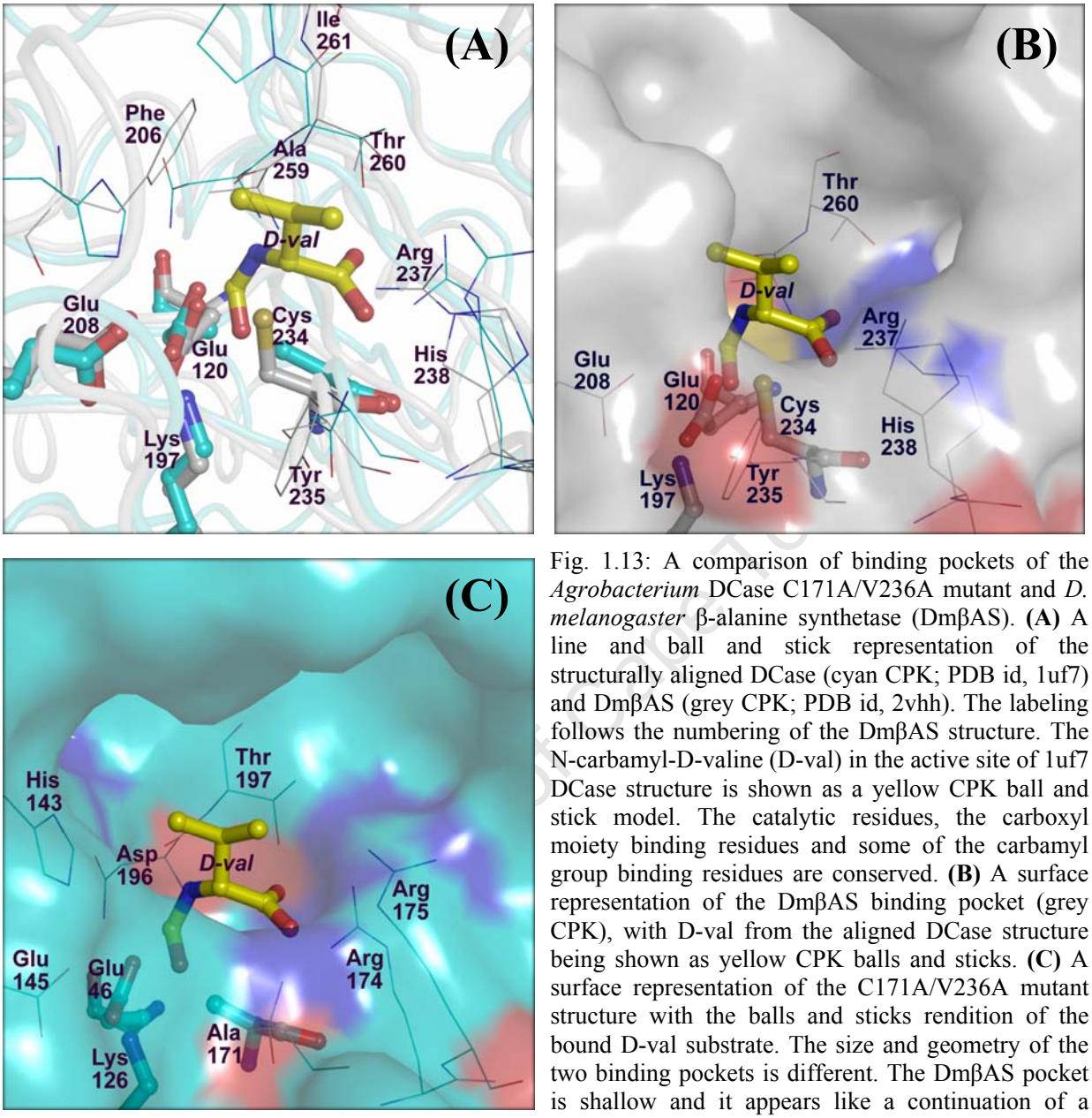


Fig. 1.13: A comparison of binding pockets of the *Agrobacterium* DCase C171A/V236A mutant and *D. melanogaster* β-alanine synthetase (DmβAS). **(A)** A line and ball and stick representation of the structurally aligned DCase (cyan CPK; PDB id, 1uf7) and DmβAS (grey CPK; PDB id, 2vhh). The labeling follows the numbering of the DmβAS structure. The N-carbamyl-D-valine (D-val) in the active site of 1uf7 DCase structure is shown as a yellow CPK ball and stick model. The catalytic residues, the carboxyl moiety binding residues and some of the carbamyl group binding residues are conserved. **(B)** A surface representation of the DmβAS binding pocket (grey CPK), with D-val from the aligned DCase structure being shown as yellow CPK balls and sticks. **(C)** A surface representation of the C171A/V236A mutant structure with the balls and sticks rendition of the bound D-val substrate. The size and geometry of the two binding pockets is different. The DmβAS pocket is shallow and it appears like a continuation of a tunnel that leads to the surface of the molecule.

Although some of the residues involved in the binding of the carbamyl moiety in the bacterial DCase (His143, Asn172 and Asn196) are not conserved in the *D. melanogaster* βAS, the amino group of the carbamyl moieties of the N-carbamyl-β-alanine and N-carbamyl-β-aminoisobutyrate substrates is expected to be located between the carboxylates of the conserved glutamate side chains (Glu120 and Glu208) in *D. melanogaster* βAS, ensuring that the substrates are positioned in the active site pocket in a similar manner to the bacterial

DCases substrates. Asn172 and Asn196 in DCase are replaced by Tyr235 and Ala259 respectively in the *D. melanogaster* βAS. The observed differences may be important for accommodating the additional methylene bridge in the β-amino acids functional groups.

Unlike the N-carbamylated D-amino acid substrates of DCases, βAS substrates (N-carbamyl-β-alanine and N-carbamyl-β-aminoisobutyrate) do not possess an amino acid side chain larger than a methyl group. The differences in the substrates sizes may explain the observed differences in the size and the geometry of the active site regions that accommodate the amino acid side chain moiety in the two enzymes. The active site pocket of the βAS is narrower and smaller in the side chain moiety-binding region as seen in figure1.13 (B) below.

The bacterial DCases are active as homotetrameric complexes (Nakai et al. 2000), while the βAS from *D. melanogaster* forms various oligomeric complexes in solution, but only the homooctameric oligomer with a helical turn-like assembly was observed in the crystal structure (Fig. 1.7) (Lundgren et al. 2008). In both cases, a conserved dimer interface exists between each two monomers in the biological complex. This interface is essential for the activity in all nitrilase superfamily enzymes of which the structure is known as the involved interactions are needed for positioning the catalytic cysteine in the active site. Although bacterial DCases are highly thermostable, the basis for this heat stability has not been elucidated.

1.3 Catalysis in the nitrilase superfamily amidases

1.3.1 Exploration of amidases as thiol enzymes

Since 1950s, researchers have been interested in amidase-catalyzed reactions and various activities involving the cysteine thiol group have been described ever since. Grossowics and colleagues (1950) reported the enzyme-catalyzed acyl transfer activity involving replacement of amide groups of asparagine and glutamine by hydroxylamine, resulting in the formation of hydroxamic acid. In 1970, Clarke reported various reactions catalyzed by the aliphatic amidase from *Pseudomonas aeruginosa* to include (1), amide and ester hydrolysis to carboxylic acids, and (2), transfer of acyl groups from amides, acids and esters to hydroxylamine resulting in the formation of hydroxamates. In all these reactions, Clarke proposed a possible involvement of a thiol group and suggested a mechanism of action involving the formation of an acyl-enzyme intermediate that then undergoes a nucleophilic attack by a water molecule (hydrolytic reaction) or hydroxylamine (acyl transfer) to release the reaction products (Clarke, 1970). Later on in 1978, Jallageas and coworkers were able to demonstrate that the aliphatic amidase from *Brevibacterium* sp. R312 was a thiol enzyme, having at least one -SH group. To-date, a large number of amidases from various organisms has been characterized, and a few crystal structures with or without bound substrates or reactions intermediates are now available.

1.3.2 The proposed Cys, Glu and Lys (CEK) catalytic triad

During their characterization of the nitrilases from *Rhodococcus rhodochrous* J1 and *Rhodococcus rhodochrous* K22, Kobayashi and coworkers (1992a and 1992b) observed enzymatic inactivation that was caused by the loss of a conserved sulfhydryl group after the enzymes were treated with thiol reagents (2-nitrobenzoic acid). A mutation of the conserved cysteines (Cys165 and Cys170) to serine or alanine in the *R. rhodochrous* J1 and *R. rhodochrous* K22 nitrilases respectively resulted in inactive mutants (Kobayashi et al. 1992a;Kobayashi et al. 1992b), confirming their essentiality in activity.

From evolutionary and homology database searches, Bork and Koonin (1994) pulled out a number of similar sequences belonging to what they called the carbon-nitrogen hydrolases

family (presently known as the nitrilase superfamily). Multiple alignments of these sequences identified at least two highly conserved motifs, one of which contained the catalytic cysteine, and the other one having an invariant glutamic acid residue (Bork and Koonin 1994). Due to high conservation of this Glu residue, and the hydrophobic region around it, the authors speculated it could also play a role in catalysis. A similar homology databases search, multiple sequence alignment and phylogenetic study by Novo and colleagues (1995) identified Cys166 as the catalytic nucleophile for the aliphatic amidase from *P. aeruginosa*. Site-directed mutagenesis of this cysteine (Cys166) to alanine or serine resulted in inactive amidase mutants (Novo et al. 1995). A sequence region with an invariant glutamic acid residue and another with a highly conserved lysine were also present in the multiple sequence alignment.

Grifantini and colleagues (1996) also identified Cys172 as the nucleophile in the N-carbamyl-D-amino acid amidohydrolase (DCase) from *Agrobacterium radiobacter* using chemical derivatization studies and site-directed mutagenesis. Similar to the other cases described above, C172A mutant lacked enzymatic activity (Grifantini et al. 1996).

When the crystal structure of DCase from *Agrobacterium sp.* KNK712 (Nakai et al. 2000) was reported (the first crystal structure of a nitrilase superfamily member), details of the active site pocket and catalysis in the nitrilase superfamily began to unfold. A region of buried volume was identified as the binding pocket in DCase, and Cys171 was found to be surrounded by among other residues, Glu46 and Lys126 (Fig. 1.14 (A)). These Glu and Lys residues were also found to be conserved in other nitrilases (Bork and Koonin 1994;Novo et al. 1995). Based on the geometry of the active site residues and molecular docking of N-carbamyl-D-phenylalanine substrate in the active site, Nakai and colleagues (2000) confirmed CEK to be the catalytic triad in the nitrilase superfamily enzymes and proposed a catalytic mechanism for N-carbamyl-D-amino acids hydrolysis (shown later in figure 1.19).

At about the same time as when the DCase structure was reported, the crystal structure of *C. elegans* (worm) NitFhit Rossetta Stone protein (Pace et al. 2000) became available. NitFhit is a Rosetta Stone protein having an N-terminal nitrilase-related Nit domain with unknown biological function, fused to a C-terminal Fhit domain that belongs to the histidine triad superfamily of nucleotide-binding proteins (Brenner et al. 1999). In what was proposed to be the substrate binding pocket of the Nit domain, Cys169 was located in a solvent-accessible area in close proximity to Glu54 and Lys127 (Fig. 1.14 (B)). Based on the observed structure of the

putative active site, Pace and colleagues also proposed CEK to be the catalytic triad in the nitrilase enzymes, with the glutamic acid residue being the general base for the thiol group (Pace et al. 2000).

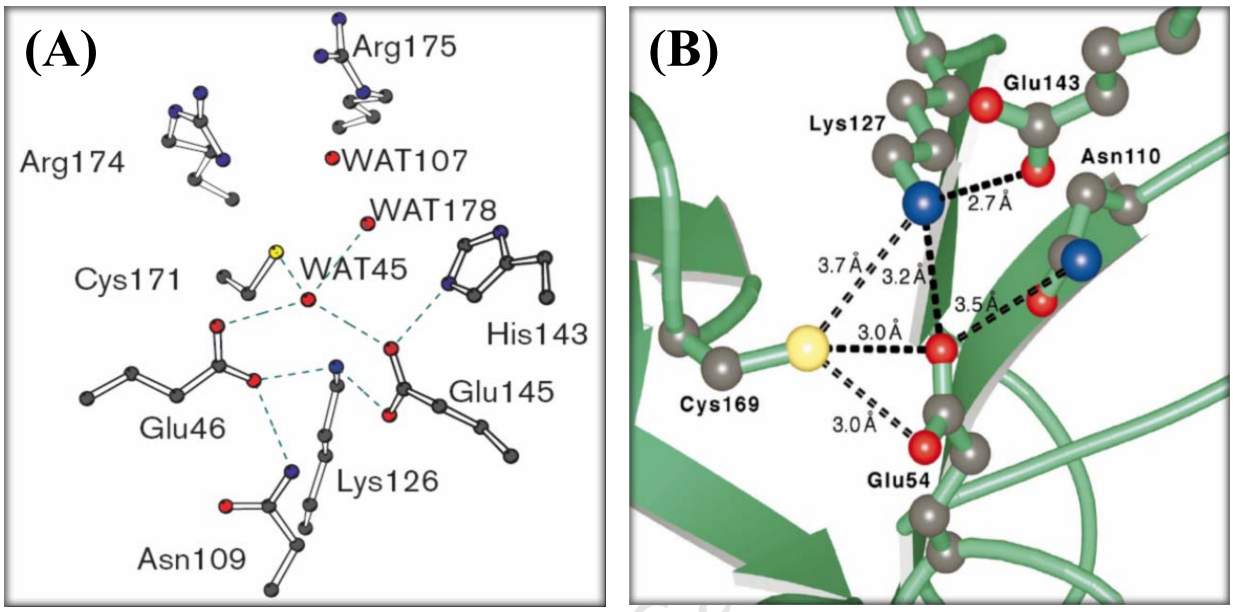


Fig. 1.14: Active site region and the proposed catalytic triad in the structures of *Agrobacterium* sp. DCase (PDB id, 1erz) and Nit domain of worm NitFhit protein (PDB id, 1ems). **(A)** Residues and polar contacts network that maintain the configuration of the active site pocket in DCase. Cys171, Glu46 and Lys126 form the proposed catalytic triad. **(B)** The Cys169, Glu54, Lys127 catalytic triad in the binding pocket of Nit domain of the worm NitFhit protein, shown as balls and sticks. These residues are conserved in all the other members of the nitrilase superfamily. The pictures were taken from Nakai et al. 2000 - (A) and Pace et al. 2000 - (B) respectively.

The availability of the first two nitrilase superfamily structures in 2000 paved the way for further investigations on catalysis in sequences whose structures were not known yet. A homology model of the *P. aeruginosa* amidase generated based on the Nit domain of the worm NitFhit structure revealed the putative catalytic triad of this amidase to be comprised of Cys166, Glu59 and Lys134 (Novo et al. 2002). Site-directed mutagenesis of Glu59 and Lys134 resulted not only in inactive mutants but also in changes in the quaternary structure (Novo et al. 2002). E59Q mutant was unstable, with the quaternary structure existing in equilibrium between trimers and hexamers, while E59N mutation caused a complete disintegration of the quaternary structure and loss of activity. Based on these observations, this glutamate residue was proposed to be essential for catalysis as well as in maintaining the quaternary structure of amidases (Novo et al. 2002). A E59V mutation of the same amidase had previously been found

to break the quaternary structure, resulting in the dissociation of hexamers into inactive dimers (Karmali et al. 2000). L134R mutant of *P. aeruginosa* amidase was both inactive and extremely unstable, while a L134N mutation produced a stable enzyme that lacked enzymatic activity (Novo et al. 2002). Woods and coworkers (1975) had proposed involvement of a lysine residue in catalysis during their kinetics studies on the *P. aeruginosa* amidase.

Mutation of all the three putative catalytic triad residues (Cys176, Glu52 and Lys121) to alanine in the NADsyn^{gln} from MTB resulted in the complete loss of glutaminase domain activity in all three cases (Bellinzoni et al. 2005), confirming the involvement of the three residues in catalysis. These findings also helped explain the glutamine-dependence activity observed in eukaryotic and some prokaryotic NAD^+ synthetases.

All crystal structures of the nitrilase superfamily enzymes available to-date show the invariant Cys, Glu, Lys catalytic triad. The orientation and coordination of these residues is conserved in all members (Fig. 1.15 (A)). In addition to the catalytic triad, a second glutamic acid residue is present in the active sites of all nitrilase structures and most known sequences (Kimani et al. 2007;Soriano-Maldonado et al. 2011). The configuration of this residue is conserved in all member structures except for the worm NitFhit protein structure (PDB id, 1ems (Pace et al. 2000)) (Fig. 1.15 (A)). In both the Nit domain of the worm NitFhit protein and the mouse nitrilase 2 enzyme (Barglow et al. 2008), the extra (second) glutamate residue is located at one end of a flexible beta hairpin structure (Fig. 1.15 (B)) that is replaced by a short loop in all the other nitrilase structures. While this beta hairpin is proposed to be involved in substrate binding (Barglow et al. 2008), it may also play a role in docking the side chain of the extra Glu residue into the active site. The orientation and position of the second glutamic acid in these two structures may therefore be dependent on whether the beta hairpin loop is in an open or a closed conformation, which may explain why Glu143 (extra Glu) in the worm NitFhit structure is in a different configuration. Indeed, the beta hairpin loop is in an open conformation in the worm NitFhit structure as compared to the more closed conformation in the mouse nitrilase 2 structure (Fig. 1.15 (B)).

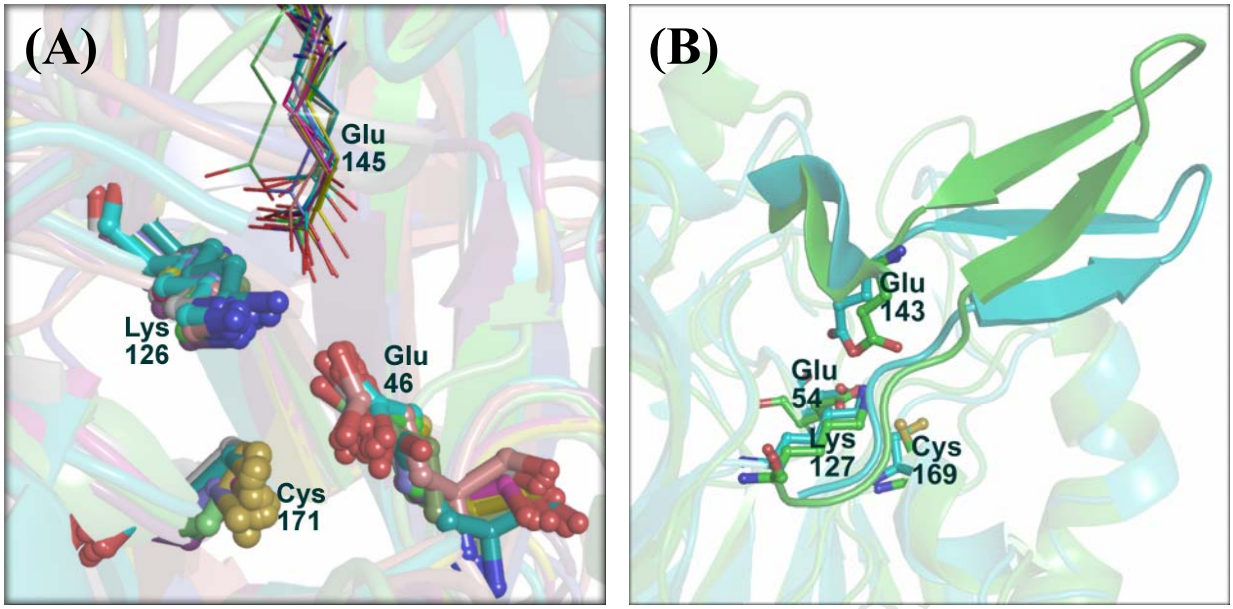


Fig. 1.15: The catalytic residues in the 15 unique structures of the nitrilase superfamily enzymes. **(A)** Superposition of 15 unique nitrilase superfamily structures (PDB ids: 1erz (Nakai et al. 2000), 1j31 (Sakai et al. 2004), 2plq (Kimani et al. 2007), 2uxy (Andrade et al. 2007), 3dla (LaRonde-LeBlanc et al. 2009), 2vhh (Lundgren et al. 2008), 3ilv (Palani et al. 2009), 2e2l (Hung et al. 2007), 1ems (Pace et al. 2000), 1f89 (Kumaran et al. 2003), 1fo6 (Wang et al. 2001), 2e11 (Chin et al. 2007), 3hkx (Nel et al. 2011), 3ivz (Raczynska et al. 2010) and 2w1v (Barglow et al. 2008)). The residues forming the generally accepted catalytic triad are shown as balls and sticks, while the conserved extra Glu residue is rendered as lines. The residues are labeled following the numbering of the *Agrobacterium* sp. DCase structure (PDB id, 1erz). The extra Glu residue in the worm NitFhit structure (in green lines) is in a different conformation. **(B)** An alignment of the Nit domain from the worm NitFhit structure (green CPK; PDB id, 1ems) with the mouse nitrilase 2 structure (cyan CPK; PDB id, 2w1v) showing the beta hairpin loop that is only found in these two structures. The catalytic residues are shown as balls and sticks, with the labeling following the worm NitFhit structure. This loop structure is in an open conformation in 1ems (green) as compared to a more closed conformation in 2w1v (cyan).

1.3.3 Evidence for the formation of reaction intermediates

The occurrence of an acyl-enzyme intermediate was first demonstrated by Maestracci and colleagues (1986), who reported the kinetics and the mechanism of the hydrolysis and acyl transfer reactions that are catalyzed by the aliphatic amidase from *Brevibacterium* sp. R312. Using acetamide as a substrate and hydroxylamine as a co-substrate in addition to water in the reaction medium, the acyl transfer reaction mechanism was shown to be of the 'ping pong bi bi' type, with a similar mechanism being extrapolated for amide hydrolysis reaction (Maestracci et al. 1986). In the described mechanism, the amide (acyl donor) first reacts with the enzyme resulting in an acyl-enzyme complex with the release of ammonia as the first

product. Subsequent reaction of the intermediate with either water (hydrolysis reaction) or hydroxylamine (acyl transfer reaction) results in the formation of the second product (carboxylic acid or hydroxamic acid), the release of which regenerates an active enzyme (Maestracci et al. 1986). A schematic diagram of the proposed 'ping pong bi bi' reaction pathway is shown in figure 1.16.

The 'ping pong bi bi' catalytic process was also observed by Fournand and colleagues (1998b), who performed a detailed acyl transfer kinetics study on the amidase from *Rhodococcus* sp. R312. This study confirmed that the acyl-enzyme intermediates do indeed occur in the amidase-catalyzed reactions, and identified the breakdown (deacylation) of the acyl-enzyme complex to be the rate limiting step of the reaction pathway. Since water is a less potent nucleophilic agent than hydroxylamine, the hydrolysis reaction is always slower than the acyl transfer reaction (Fournand et al. 1998a;Maestracci et al. 1986).

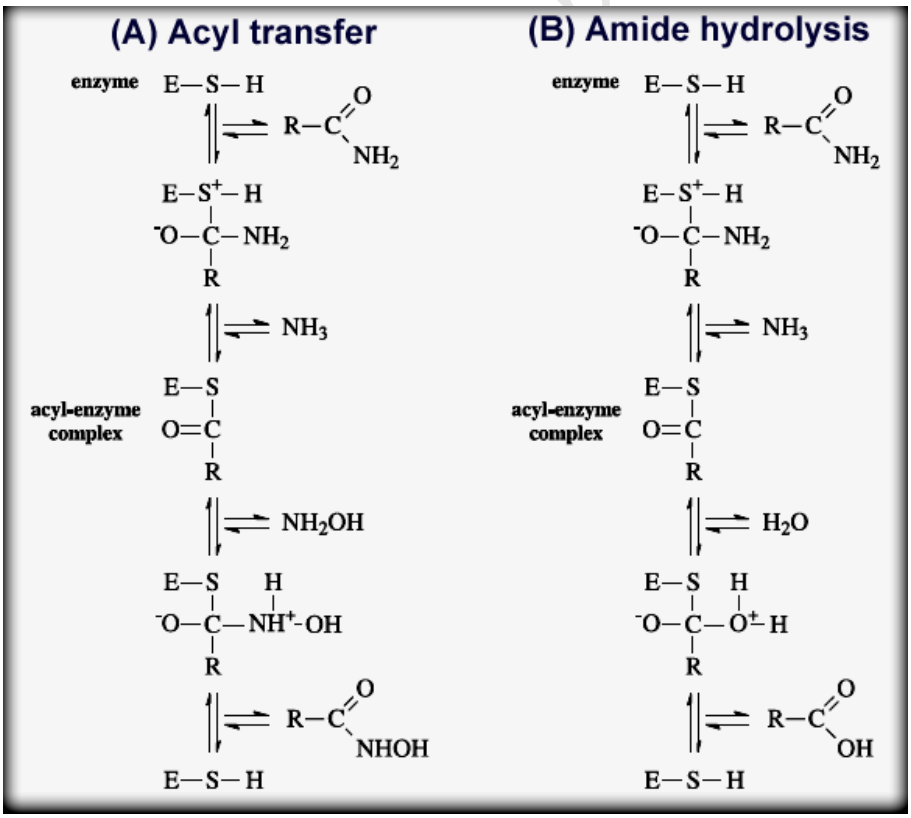


Fig. 1.16: The proposed 'ping pong bi bi' catalytic pathways of (a) an acyl transfer reaction from an amide to hydroxylamine and (b) an amide hydrolysis reaction catalyzed by the aliphatic amidase from *Brevibacterium* sp. R312 (Maestracci et al. 1986). Both reactions are proposed to proceed via an acyl-enzyme intermediate. The picture was taken from Fournand and Arnaud 2001.

LaRonde-LeBlanc and colleagues (2009) crystallized the MTB NADsyn^{gln} in the presence of a glutaminase domain inhibitor, DON (6-diazo-5-oxo-L-norleucine). DON is a glutamine analogue that reacts with the nucleophilic cysteine following the reaction scheme in figure 1.17 below, and resulting in an irreversible covalent intermediate (Fig. 1.18 (A)) that resembles the γ-glutamylthioester intermediate (Fig. 1.4) from the natural substrate (L-glutamine) (LaRonde-LeBlanc et al. 2009).

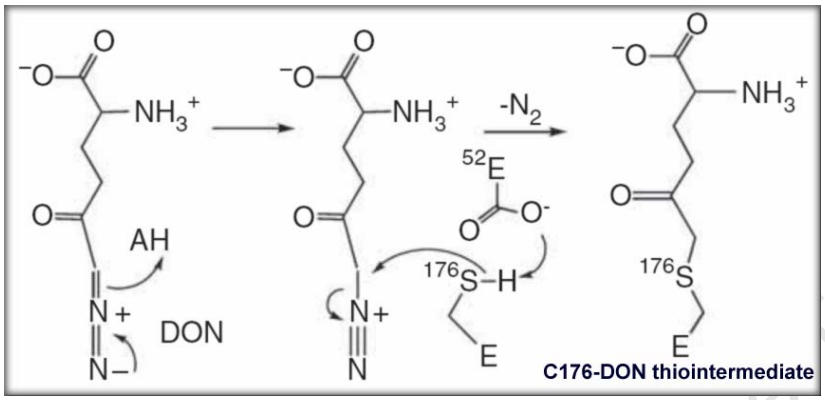


Fig. 1.17: Illustration of a reaction mechanism that leads to the irreversible covalent modification of the nucleophilic cysteine (Cys176) by DON in the glutaminase domain of the MTB NADsyn^{gln} structure. The figure was taken from LaRonde-LeBlanc et al. 2009.

The crystal structure of the aliphatic amidase from *P. aeruginosa* (PDB id, 2uxy) had an acyl transfer reaction second tetrahedral intermediate of acetohydroxamic acid covalently bound to the nucleophilic cysteine (Andrade et al. 2007) (Fig. 1.18 (B)). Although this reaction was not intended, it occurred when the amidase was being eluted from the affinity column (Epoxy-activated sepharose 6B-acetamide) using acetamide (acyl donor) and hydroxylamine (acyl acceptor) during purification. The two substrates reacted according to the mechanism shown in Fig. 1.16 panel (B), but the reaction stopped just before the final product, acetohydroxamic acid could be released, resulting in the second tetrahedral intermediate adduct being trapped in the active site. The visualization of this intermediate is satisfactory evidence, that nitrilase superfamily amidases are thiol enzymes and that various reaction intermediates do occur in the course of amidase-catalyzed reactions.

The crystal structure of XC1258 Nit nitrilase from *Xanthomonas campestris* (PDB id, 2e11; (Chin et al. 2007)) had a dimethylarsenate tetrahedral covalent adduct at the catalytic cysteine trapped in the active site (Fig. 1.18 (C)). The catalytic cysteine (Cys143) reacted with the cacodylate (dimethylarsenic acid) in the crystallization buffer, resulting in a stable adduct that is analogous to the natural tetrahedral transition state that is proposed for the hydrolysis of the

acyl-enzyme intermediate in the reaction pathways of nitrilase and amidase enzymes (Chin et al. 2007).

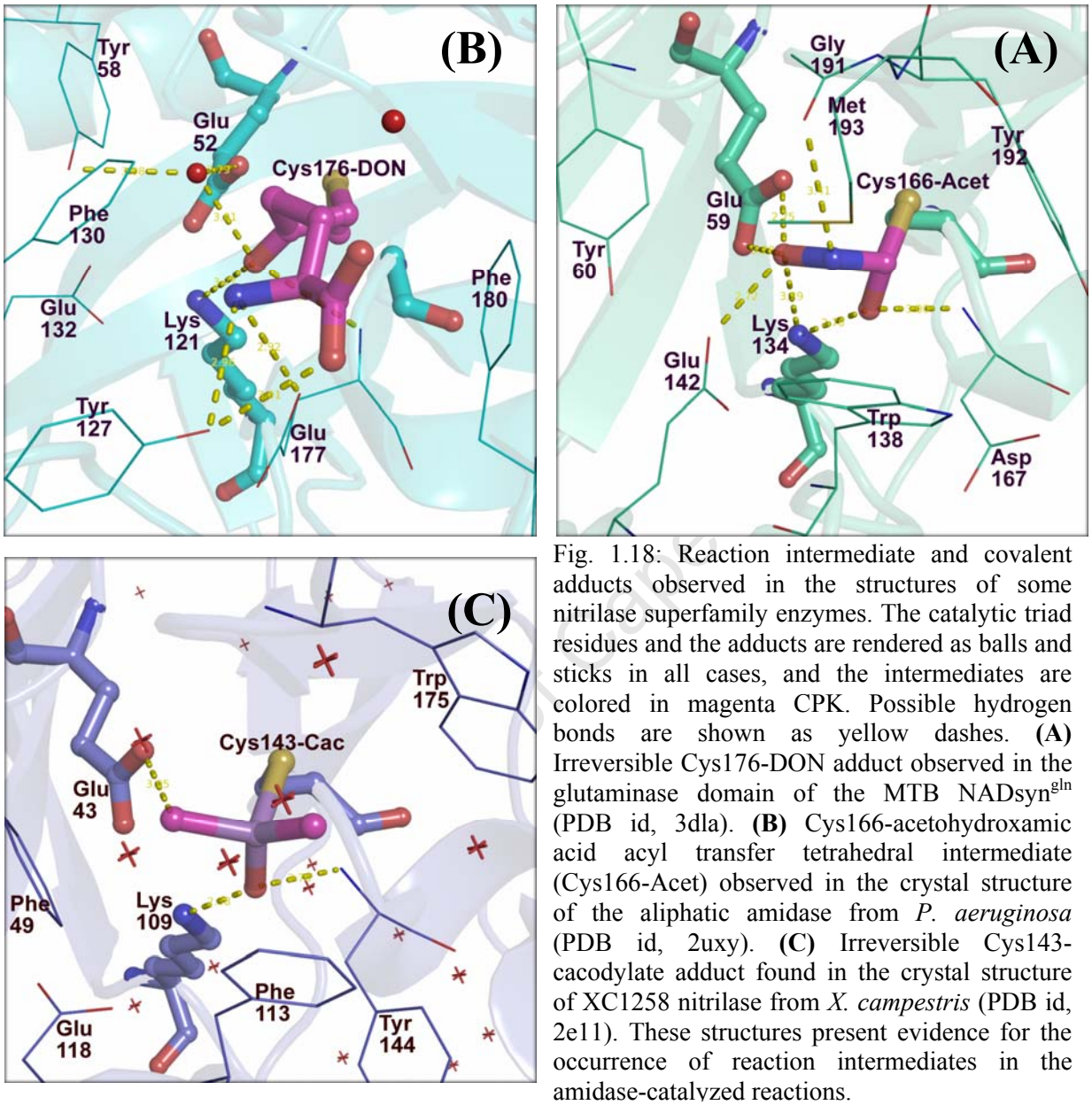


Fig. 1.18: Reaction intermediate and covalent adducts observed in the structures of some nitrilase superfamily enzymes. The catalytic triad residues and the adducts are rendered as balls and sticks in all cases, and the intermediates are colored in magenta CPK. Possible hydrogen bonds are shown as yellow dashes. **(A)** Irreversible Cys176-DON adduct observed in the glutaminase domain of the MTB NADsyn^{gln} (PDB id, 3dla). **(B)** Cys166-acetohydroxamic acid acyl transfer tetrahedral intermediate (Cys166-Acet) observed in the crystal structure of the aliphatic amidase from *P. aeruginosa* (PDB id, 2uxy). **(C)** Irreversible Cys143-cacodylate adduct found in the crystal structure of XC1258 nitrilase from *X. campestris* (PDB id, 2e11). These structures present evidence for the occurrence of reaction intermediates in the amidase-catalyzed reactions.

In their recent biochemical studies, Buddelmeijer and Young (2010) reported that the apolipoprotein N-acyltransferase from *Escherichia coli* mainly exists as a thioester acyl-enzyme intermediate in the extracytoplasm. This enzyme is involved in post-translational modification of lipoproteins where it condenses an acyl group from sn-1-glycerolphospholipid onto the free α-amino group of the N-terminal cysteine of apolipoproteins. Although the extracytoplasmic thioester form of this reverse amidase was proposed to be in a state of rapid

synthesis and breakdown (Buddelmeijer and Young 2010), it does confirm existence of such intermediates in the amidase-catalyzed reactions.

Stevenson and colleagues (1990) demonstrated the formation of enzyme-substrate reaction intermediates during their mass spectrometric characterization of the nitrilase from *Rhodococcus* ATCC 39484 pre-incubated with various nitrile substrates. The outcome of this study not only confirmed this nitrilase to be a thiol enzyme, but the observed mass shifts indicated specific covalent modifications during the reaction process (Stevenson et al. 1990). It was also established that, the breakdown of the covalent acyl-enzyme intermediate was the rate limiting step and served as a distinguishing factor between good and poor substrates. This is in agreement with similar findings by Fournand and coworkers during their elucidation of the catalytic mechanism of the aliphatic amidase from *Rhodococcus* sp. R312 (Fournand et al. 1998a).

1.3.4 The proposed catalytic mechanism

On determining the crystal structure of the native *Agrobacterium* sp. DCase, Nakai and coworkers (2000) proposed a catalytic mechanism, where the residues of the putative catalytic triad (Cys, Glu, Lys) play key roles in driving substrate hydrolysis in this enzyme. The carboxyl group of the glutamic acid (Glu46) was proposed to be a general base catalyst that abstracts a proton from the thiol group of catalytic cysteine, hence enhancing its nucleophilicity. The activated cysteine (Cys171) then acts as a nucleophile that attacks the amide carbon of the substrate, resulting in the formation of a tetrahedral intermediate. The lysine residue (Lys126) is thought to function at this point, stabilizing the oxyanion of the tetrahedral intermediate through its positively charged amino group. The tetrahedral complex then collapses, resulting in the formation of an acyl-enzyme intermediate and the release of ammonia. The carboxyl group of the glutamate residue (Glu46) acts as a general base catalyst again, activating a water molecule that in turn attacks the acyl-enzyme complex, resulting in a second tetrahedral intermediate that is stabilized by the positively charged amino group of the catalytic lysine (Lys126) side chain. Rearrangements involving bonds breakage and formation results in the release of the acid product and recovery of the enzyme (Nakai et al. 2000). The proposed hydrolysis reaction mechanism is shown in figure 1.19.

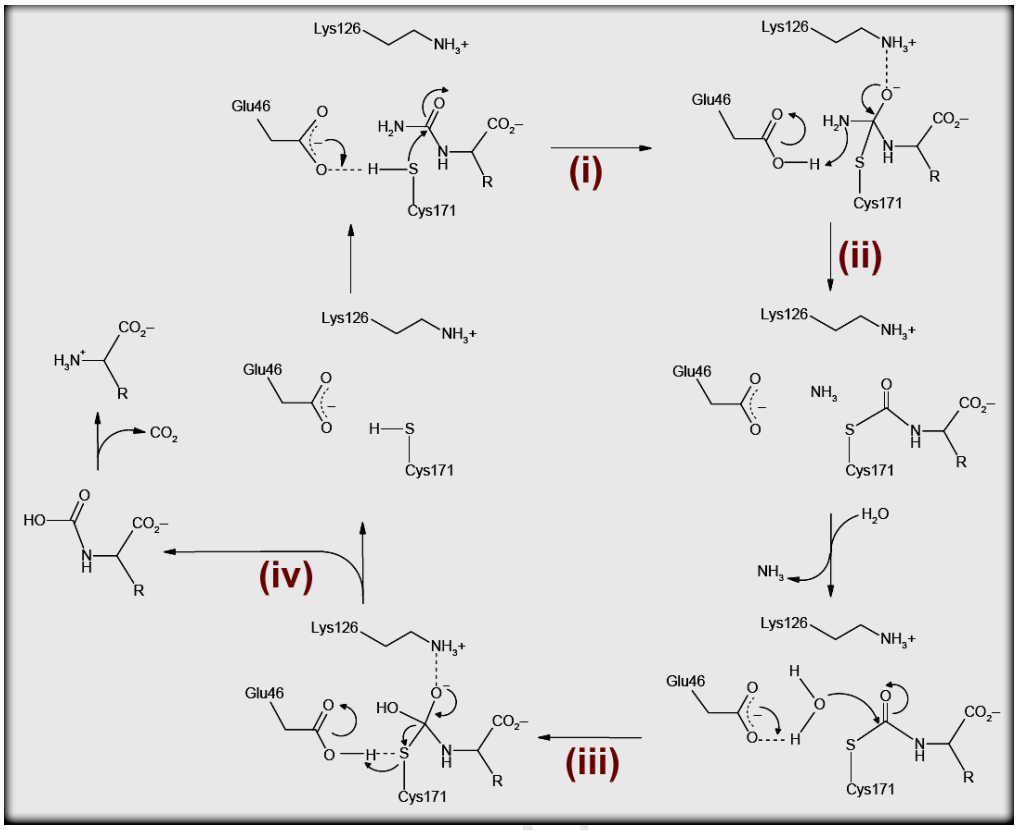


Fig. 1.19: The proposed catalytic mechanism for the DCase enzyme from *Agrobacterium* sp. Glu46 acts as a general-base catalyst that enhances the nucleophilicity of Cys171. Activated Cys171 then acts a nucleophile attacking the amide carbon of the substrate and resulting in the formation of a tetrahedral intermediate (i). Lys126 stabilizes the tetrahedral intermediate, which then collapses to form an acyl-enzyme intermediate (ii).General base activation of a water molecule leads to the attack of the acyl-enzyme complex (iii), which results in a second tetrahedral intermediate. The intermediate then collapses leading to the release of the acid product and regeneration of an active enzyme (iv). The figure was taken from Nakai et al. 2000.

The catalytic mechanism proposed by Nakai and others (2000) is generally accepted for all nitrilase superfamily enzymes. Hung and colleagues (2007) proposed a similar mechanism for the formamidase from *Helicobacter pylori* after its native and substrate-bound crystal structures were determined. The acyl transfer reaction (observed mainly in branch 2 amidases) is also proposed to follow the same mechanism, with the only difference being the deacylation step where hydroxylamine or any other acyl acceptor would carry out the nucleophilic attack of the acyl-enzyme complex instead of a water molecule (Maestracci et al. 1986). An acyl transfer reaction mechanism (Fig. 1.20) was recently proposed by Buddelmeijer and young (2010) for the apolipoprotein N-acyltransferase from *Escherichia coli*. This branch 9 enzyme catalyzes amide condensation reactions by transferring an acyl group from *sn*-1-glycerolphospholipid to

the free α-amino group of the N-terminal cysteine of apolipoproteins, resulting in triacylated lipoproteins (Buddelmeijer and Young 2010).

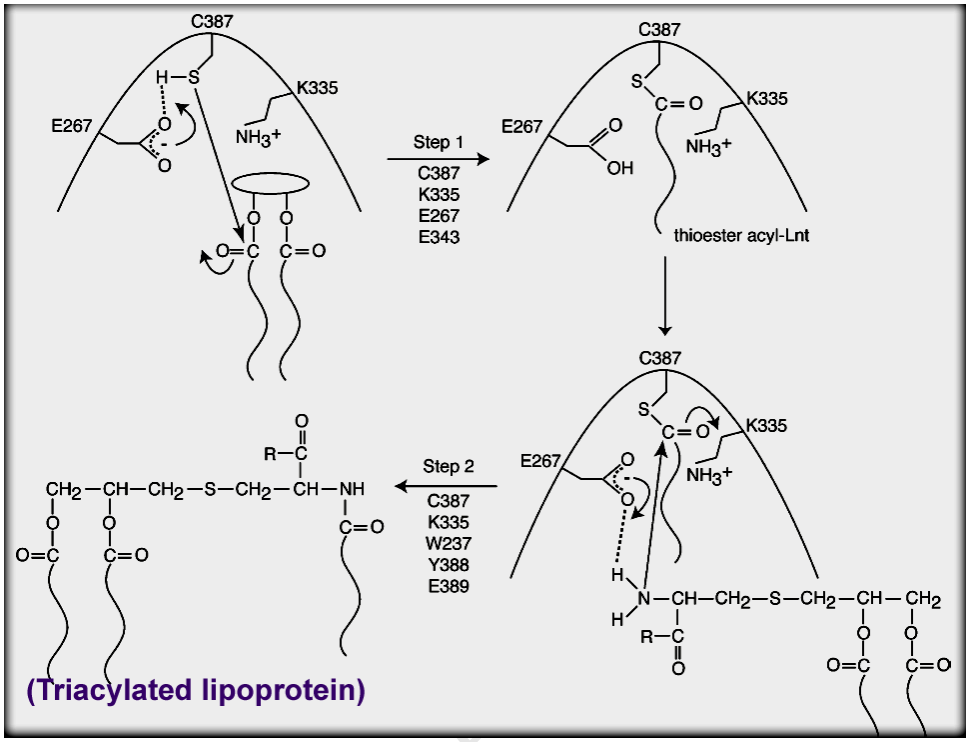


Fig. 1.20: The proposed catalytic mechanism for the amide condensation reaction catalyzed by the *E. coli* apolipoprotein N-acyltransferase reverse amidase. Activation of the sulfhydryl group of Cys387 by Glu267 results in the attack of the 1-carbonyl group of phospholipid substrates leading to the formation of a tetrahedral intermediate that is stabilized by K335. An acyl-enzyme intermediate is subsequently formed, with the release of lysophospholipid by-product. Activation of the apolipoprotein by Glu267 is then followed by the attackat the acyl-enzyme complex, resulting in the acyl group being transferred to the N-terminal cysteine of the apolipoprotein. This yields a triacylated mature lipoprotein. The figure was taken from Buddelmeijer and Young 2010.

Insights into the configuration of the tetrahedral transition state that results from the attack of the acyl-enzyme intermediate by a water molecule (or other acyl group acceptors) can be obtained from the Cys-dimethylarsenate adduct observed in the XC1258 nitrilase crystal structure (PDB id, 2e11; (Chin et al. 2007)). Analysis of this artificial adduct showed that, the volume that is occupied by one of methyl groups in the cacodylate complex (Fig. 1.21) is large enough to be occupied by a nucleophilic water molecule that would attack the acyl-enzyme intermediate during the deacylation phase of the reaction. Chin and colleagues (2007) argued that a water molecule in that position would be geometrically favoured to carry out the nucleophilic attack.

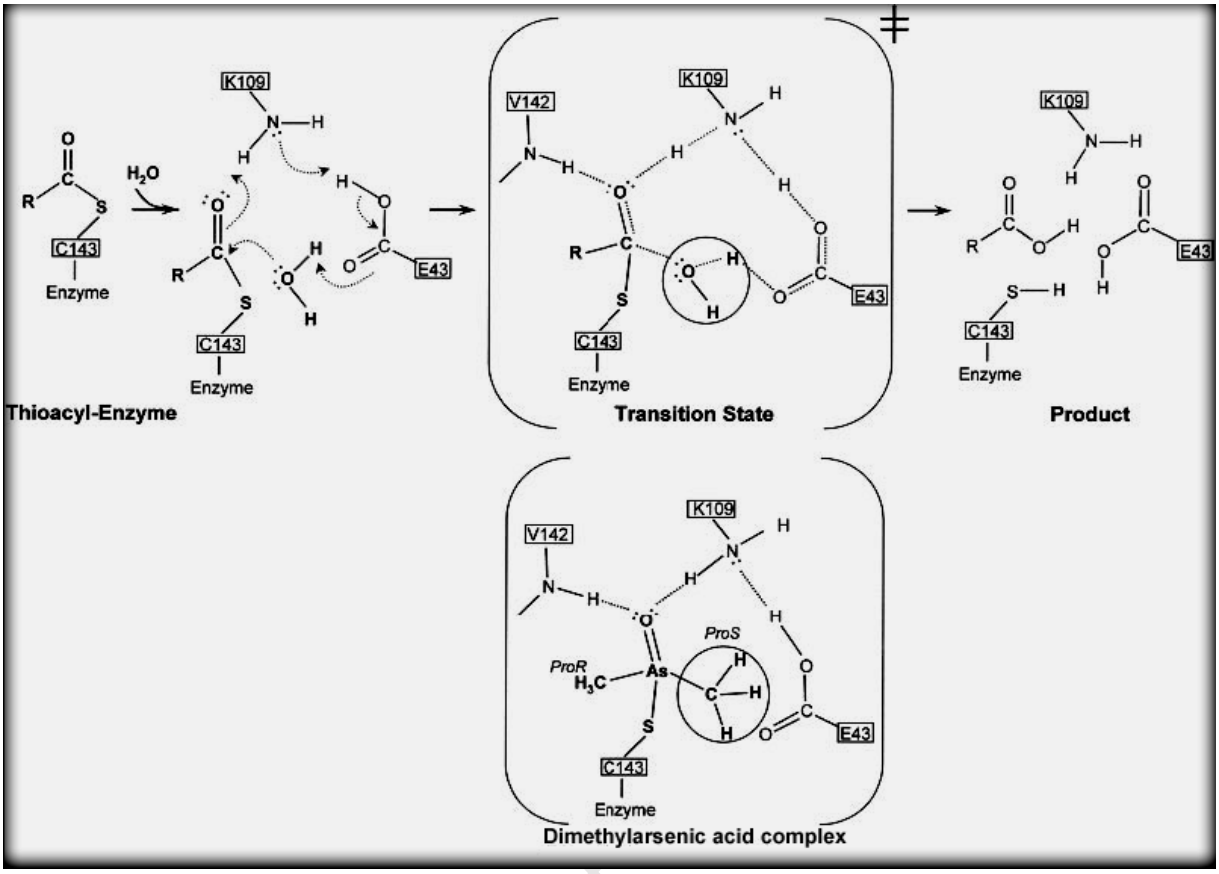


Fig. 1.21: A Cys-dimethylarsenate complex that is analogous to the tetrahedral transition state of the hydrolysis of the acyl-enzyme intermediate after an attack by a water molecule. The position occupied by one of the methyl groups (ProS) of the complex can accommodate a nucleophilic water molecule during amide and nitrile hydrolysis reactions. The picture was taken from Chin et al. 2007.

While the proposed mechanisms for amide hydrolysis and acyl transfer are generally accepted, there are still unclarified questions regarding (1), the source of the proton that gets added to the amide NH_2 group of the first enzyme-substrate tetrahedral intermediate resulting in the release of ammonia by-product; (2), the role of the extra ('second') glutamic acid residue in catalysis; and (3), the exact function of the catalytic lysine residue. Nakai and colleagues (2000) suggested the proton was likely to come from the carboxyl group of the 'first' glutamate residue (general base catalyst) or it may result from electrostatic interactions between the substrate and the extra ('second') glutamic acid residue. LaRonde-LeBlanc and colleagues (2009) supported the proposition of the proton coming from the general base catalyst ('first') glutamate residue.

Raczynska and colleagues (2010) recently published four structures of a putative branch 1 aliphatic nitrilase from *Pyrococcus abyssi,* having various ions and ligands in the active site pocket. Based on the hydrogen bonding network in the binding pocket, positioning of various residues and a docked fumaronitrile substrate, the authors proposed a concerted acid-base catalytic mechanism for nitrile-hydrolyzing enzymes. The 'first' glutamate residue (Glu42 in the *P. abyssi* nitrilase) is proposed to be the catalytic base that activates the nucleophilic cysteine, while the lysine residue (Lys113) is thought to play the role of a catalytic acid, donating a proton to the terminal nitrogen of the nitrile substrates (Raczynska et al. 2010). Protonation of the nitrile substrates is a requirement for the formation of the thioimidate intermediate (Stevenson et al. 1990) in the nitrilase-catalyzed reactions (see section 1.3.4.1 below). Raczynska and coworkers proposes the 'second' glutamic acid residue (Glu120) to be involved in enhancing the pKa of Lys113, ensuring that it is positively charged and ready to donate a proton to the nitrile substrate. However, given the high pKa of the catalytic lysine amino group (predicted to be above 15 is some cases) due to its location between the carboxylates of the two glutamates, it might be incapable of donating a proton to the nitrile substrate molecules. Since the active site pockets of nitrilases and amidases are usually hydrated, water is likely to be the best candidate for hydrating the nitrile substrates into thioimidate during catalysis.

The carboxyl group of the extra ('second') glutamate residue is involved in polar contacts with the amide amino group of various substrates in the amide-complexed co-crystal structures of amidases available to-date (Chen et al. 2003;Hashimoto et al. 2003;Hung et al. 2007). Hung and co-workers (2007) suggested the second glutamate could be important for maintaining the side chain geometry of the CEK catalytic triad, as well as for facilitating the docking of the substrate molecules in the active site pocket, while Kimani and colleagues (2007) suggested that it could be involved in catalysis as a second general base catalyst in the deacylation phase of the reaction (Fig. 1.22). From the model of the D-lactamide acyl-enzyme intermediate of the *G. pallidus* RAPc8 amidase structure, the 'first' glutamic acid (Glu59) was found to be occluded from the solvent environment, leaving the 'second' glutamate (Glu142) well poised to carry out the activation of a water molecule in the second nucleophilic attack (Kimani et al. 2007).

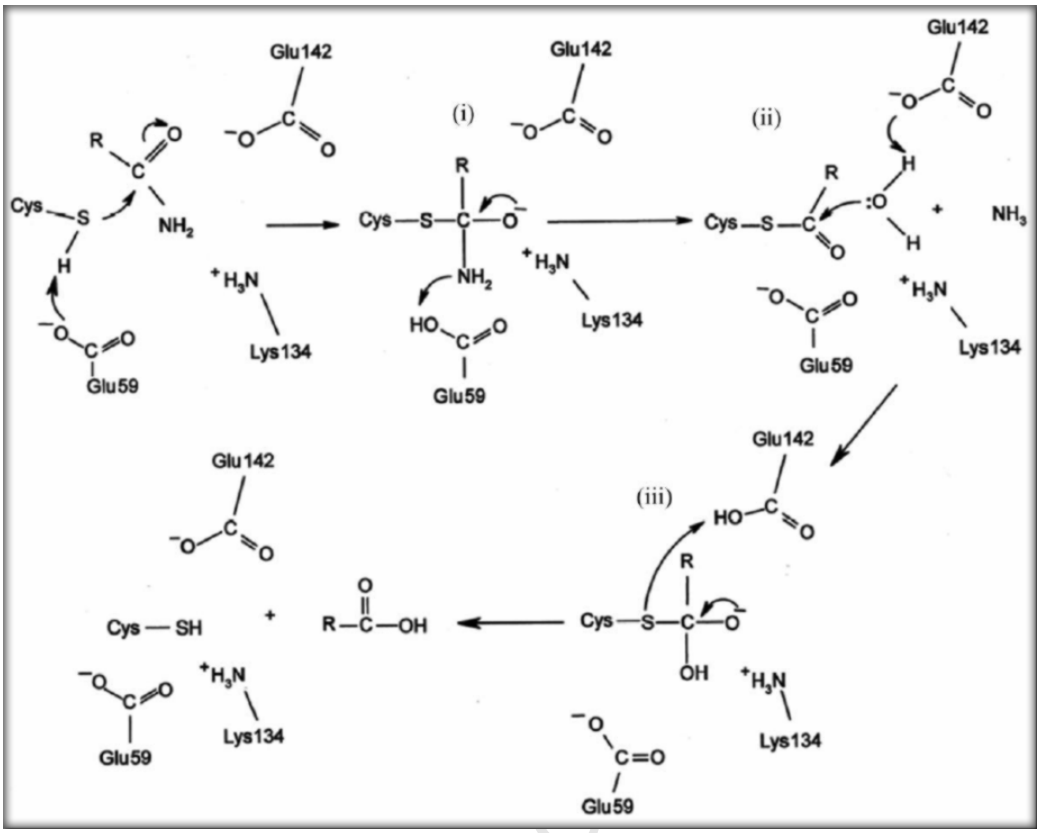


Fig. 1.22: The proposed catalytic mechanism for the aliphatic amidase from *G. pallidus* RAPc8, with the role of the 'second' glutamate (Glu142) predicted. The 'first' glutamate (Glu59) activates Cys166 and Lys134 stabilizes the oxyanion of the tetrahedral intermediate. Upon the formation of acyl-enzyme intermediate, Glu142 is proposed to play the role of a general base catalyst, activating a water molecule that initiates the deacylation process. Figure was taken from Kimani et al. 2007.

Buddelmeijer and Young (2010) observed that, the 'second' active site glutamate (Glu343) in the apolipoprotein N-acyltransferase from *E. coli* was required for the acyl-enzyme complex formation as well as for acyl group transfer to the lipoproteins. The authors proposed that this residue (Glu343) could play the role of activating the nucleophilic cysteine (Cys387) together with the 'first' Glu residue (Glu267) or it might be involved in the positioning of the phospholipid substrate in the active site, or in stabilizing the interaction between the enzyme and the phospholipid substrates (Buddelmeijer and Young 2010). Further work will need to be done to verify all these speculations.

A recent biochemical and mutational study of the formamidase from *Bacillus cereus* CECT 5050T revealed that the 'second' glutamate mutant (E140D) enzyme lacked detectable enzymatic activity even though it was still able to bind formamide (FMD) substrates (Soriano-

Maldonado et al. 2011). These observations prove that the 'second' glutamate residue has a key role in substrate hydrolysis (Soriano-Maldonado et al. 2011), although its exact function in the catalytic process is still not clear.

1.3.4.1 The proposed catalytic pathway for branch 1 nitrilases

A lot of research has been carried out to investigate the catalytic mechanism of branch 1 nitrilase enzymes that catalyze the conversion of nitriles to their corresponding acids and ammonia, or occasionally to their corresponding amides with the release of ammonia. A catalytically-essential thiol residue was shown to be present in these nitrilases (Harper 1977a;Harper 1977b;Kobayashi et al. 1992b;Kobayashi et al. 1993;Kobayashi et al. 1989) and the nitrile hydrolysis reaction was proposed to proceed through thioimidate, tetrahedral and acyl-enzyme intermediates (Bork and Koonin 1994;Harper 1977b;Kobayashi et al. 1998;Mahadevan and Thimann 1964;O'Reilly and Turner 2003;Stevenson et al. 1990) (Fig. 1.23). A common hydrolytic reaction pathway is shared by branch 1 nitrilases and amidases, with the only difference being the protonation step that leads to the formation of thioimidate intermediate in the nitrile-hydrolysis reactions (Fig. 1.23).

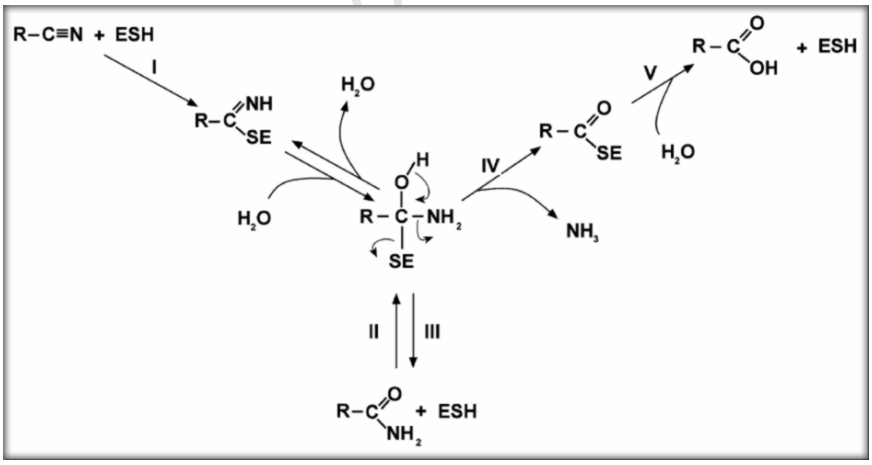


Fig. 1.23: The proposed catalytic mechanism for nitrilases and cyanide hydratases from branch 1. The catalytic cysteine thiol group attacks the cyano-carbon of nitrile molecules resulting in the protonation of the nitrogen to form the thioimidate (I). An addition of a water molecule and further protonation of the NH group leads to the tetrahedral intermediate, which then collapses to an acyl-enzyme intermediate with the release of ammonia (IV). A water molecule then reacts with the thioester intermediate (V) leading to the formation of the acid product and regeneration of an active enzyme. In the cyanide hydratases and other nitrilases that form amide products, an amide is released on rearrangement of the tetrahedral intermediate (III). Substrate protonation step presents the key difference between amidase and nitrilase catalyzed reactions. The picture was taken from O'Reilly and Turner 2003.

1.4 Activity and substrate specificity of the *Nesterenkonia* sp. amidase (NitN)

1.4.1 Activity

A unique nitrilase-related amidase (NitN) was characterized recently from a psychrotrophic *Nesterenkonia* sp. bacterium isolated from the Antarctic desert soil (Nel et al. 2011). Although the physiological function of NitN is not known, screening of the purified enzyme for activity against a variety of amide, nitrile and carbamyl substrates revealed that NitN had amide hydrolysis activity only. NitN was therefore classified as an aliphatic amidase, following its *in vitro* capability in hydrolyzing a small number of short-chain aliphatic amides: propionamide (PMD) > fluoroacetamide (FAE) > butyramide (BMD) > acetamide (ACE) in decreasing affinity (Nel et al. 2011). However, unlike other branch 2 aliphatic amidases that are known to rapidly hydrolyze short aliphatic amides (mainly PMD, acrylamide (ACR) and ACE) (Asano et al. 1982;Makhongela et al. 2007;Skouloubris et al. 2001), the catalytic rates and affinities of NitN for these substrates are quite low, with the K_M values in the millimolar (mM) range; 38.6 mM and 137.7 mM for PMD and ACE respectively. In addition, NitN does not exhibit any acyl transferase activity that is observed in other bacterial aliphatic amidases (Nel et al. 2011), and also lacks hydrolytic activity on ACR. It is worth noting that NitN displays the highest activity at 30 ˚C *in vitro*, which is approximately 10 ˚C above the optimal growth temperature of the parent bacterium, the psychrophilic *Nesterenkonia* strain AN1 (~ 21 ˚C) Nel et al. 2011.

The crystal structure of the native NitN (PDB id, 3hkx) was reported recently (Nel et al. 2011). NitN is active as a homodimer and its overall structural fold is similar to that of other nitrilase superfamily enzymes. Analysis of pockets and cavities in both the monomeric and dimeric forms of NitN using the CASTp server (Binkowski et al. 2003), revealed a relatively large substrate binding pocket (Fig. 1.24 (A)), with a solvent accessible volume of 61.84 $Å^3$ and a molecular surface volume of 363 $Å^3$. NitN binding pocket is much larger compared to the pockets of the aliphatic amidases with known structures (Andrade et al. 2007;Hung et al. 2007;Kimani et al. 2007). In the monomeric structure, the catalytic triad residues, Glu61, Lys131 and Cys165, are involved in a hydrogen bonding network (Fig. 1.24 (B)) that maintains the configuration of the pocket and provides an electronic environment that is required for catalysis.

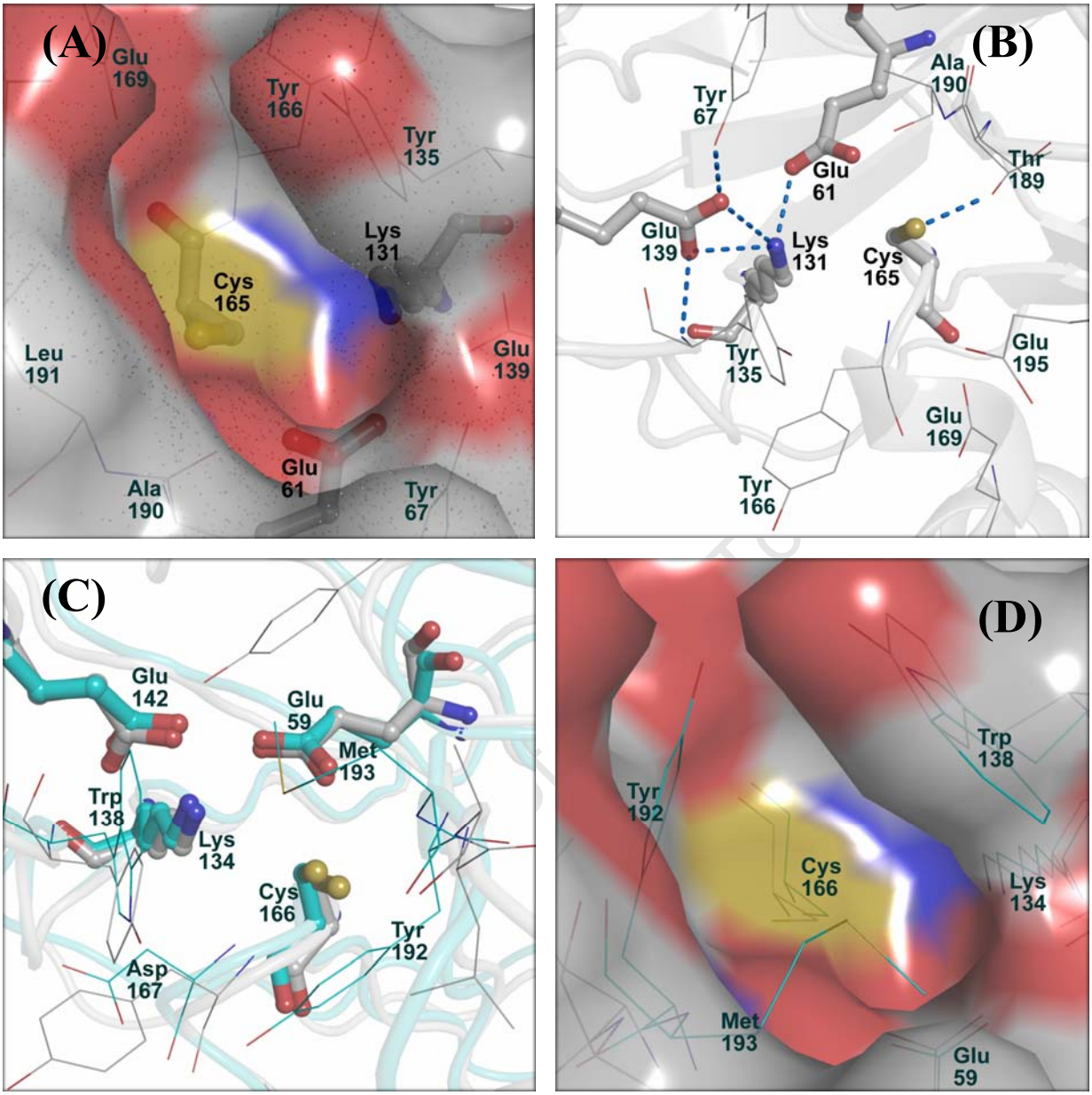


Fig. 1.24: Comparisons of the active site pockets of the NitN and *G. pallidus* amidase structures. **(A)** A transparent surface representation of the NitN pocket (PDB id, 3hkx). The dots represent atoms that were predicted to be lining the binding pocket by CASTp. The binding pocket is relatively large and deep. **(B)** A faint cartoon, line, ball and stick rendition of the NitN binding pocket. The potential hydrogen bonding contacts within the pocket are shown as blue dashes. **(C)** Superposition of the active site regions of NitN (grey CPK) and *G. pallidus* amidase (cyan CPK; PDB id, 2plq). The residues are labeled following the *G. pallidus* amidase structure numbering. The side chains of Trp138, Tyr192 and Met193 in *G. pallidus* amidase are involved in constraining the pocket size in aliphatic amidases. **(D)** A transparent surface of the NitN binding pocket (grey CPK), with the active site residues of the aligned *G. pallidus* amidase structure being shown as cyan CPK lines. The active site residues are labeled following the numbering of G. *pallidus* amidase. The side chains of Trp138, Tyr192 and Met193 in the *G. pallidus* amidase structure can be seen protruding into the NitN pocket surface, an indication of pocket size differences between the two enzymes.

1.4.2 The basis for NitN specificity on aliphatic amidases or lack of it

A comparison of NitN active site pocket with that of the aliphatic amidase from *G. pallidus* RAPc8 (Fig. 1.24 (C) and (D)) reveals some interesting differences. The geometric and size constraints that confer preference for short aliphatic amides in aliphatic amidases are not present in the NitN amidase. Trp138 residue that is located at the entrance to the *G. pallidus* amidase binding pocket is an important determinant of substrate specificity as its bulky side chain docks into the pocket, restricting the length of linear amides that can be accommodated to those with less than 4 carbon atoms (Kimani et al. 2007). Trp138 is replaced by a less bulky Tyr135 side chain in NitN, making the pocket more spacious and accessible.Tyr192 and Met193, whose side chains form a tight wall in the binding pockets of the aliphatic amidases from *G. pallidus* and *P. aeruginosa,* are replaced by Leu191 and Ala192 respectively in NitN amidase. The shorter side chains ensure more room in the NitN binding pocket, and they eliminate most of the geometric and size constraints that are observed in the branch aliphatic amidases. The basis for the observed specificity on short aliphatic amides by NitN can therefore not be immediately deduced from the WT NitN crystal structure. Co-crystal structures of NitN in complex with various aliphatic amide substrates are going to be needed in order to understand the interaction of NitN with these substrates and possibly explain the observed activity profile in this enzyme. Based on the active site pocket size and structure, it is possible that the biological substrates for NitN are not closely related to short aliphatic amide substrates identified *in vitro* (Nel et al. 2011).

1.5 Motivation and study objectives

While a highly conserved Cys, Glu, Glu and Lys (CEEK) catalytic tetrad is shared by all nitrilase superfamily amidases, a wide substrate spectrum is observed between and within different amidase families. The available structural data have not only revealed the monomeric structural fold, the quaternary complexes and the active site structure of enzymes from various amidase branches, but they have also provided useful insights into the basis for substrate specificity and preferences observed in some of the amidase families. The available amidase-substrate co-crystal structures have revealed that, the geometry and size of the active pocket are the key determining factors of the chirality and the size of the substrate molecules that individual members can accommodate. In addition, the amino acid composition of the binding pocket plays an important role in substrate recognition and in binding various moieties of the amide substrate molecules. The active site structure has allowed the proposition of a catalytic mechanism for amidase-catalyzed reactions and crystal structures of some amidases with covalently bound reaction intermediates and adducts have provided evidence that the reaction pathway of these enzymes proceed via various reaction intermediates. Despite their diverse physiological functions, some amidases are tools for drug synthesis; others are potential biocatalysts in the fine chemical industry, while those from pathogenic prokaryotes are potential drug targets. Understanding the biology of these enzymes is critical for the design of more potent industrial biocatalysts and in drug-targeting efforts.

1.5.1 Existing gaps and study motivation

The newly characterized aliphatic amidase from *Nesterenkonia* sp. (NitN) exhibits hydrolytic activity on short aliphatic amides (PMD, FAE, BMD and ACE) *in vitro* (Nel et al. 2011) even though its substrate binding pocket is much larger and less constrained compared to the active site pockets of branch 2 aliphatic amidases (Fig. 1.24 (C) and (D)) that have a similar substrate specificity profile. The basis for NitN specificity on these amides is therefore not understood, as it is difficult to make any conclusions based on the size and the geometry of the WT NitN active site pocket. While branch 2 aliphatic amidases are known to rapidly hydrolyze short aliphatic amide substrates *in vitro* (Makhongela et al. 2007;Skouloubris et al. 1997;Soubrier et al. 1992), the catalytic rates and affinities of NitN on these amide substrates are much lower (Nel et al. 2011). The reason for the reduced catalytic rates in NitN is not clear. Crystal

structures of NitN in complex with various short aliphatic amide substrates will be required in order to visualize the interactions of these substrates with the NitN pocket and hopefully explain the observed peculiarities. Even though the *in vivo* substrates of NitN are not known, more efficiently hydrolyzed (better) substrate molecules will need to be identified if NitN is going to be developed as a biocatalyst in the fine chemical industry.

The crystal structure of the catalytic cysteine mutant (C166S) of the formamidase from *Helicobacter pylori* in complex with formamide was reported four years ago (Hung et al. 2007), and it is the only substrate-complexed co-crystal structure of the branch 2 amidases that is available to-date. While the small size of the binding pocket was enough to explain the strict preference of this amidase for formamide, it is not clear how the hydrogen bonds between the carbonyl oxygen and the amino group of the formamide molecules, and the side chain hydroxyl group of Ser166 (Fig. 1.10 (C); see also (Hung et al. 2007) for details of the interactions) would affect their positioning and orientation within the binding pocket. Such interactions would not occur in the active enzyme, having the catalytic cysteine in position 166. The visualization of non-covalently bound short aliphatic amides in the NitN pocket, having the catalytic cysteine mutated to an alanine instead of a serine, would not only reveal the unbiased amidase-amide interactions, but would also provide insights into how substrates with more than one carbon atoms are positioned within the binding pocket.

While there is consensus on the general reaction pathway and the core steps in the proposed catalytic mechanism of the nitrilase superfamily amidases, details of each step and the chemistry behind them are still unclear. In addition to the generally accepted Cys, Glu and Lys (CEK) catalytic triad, a second active site glutamate residue is present in all nitrilase superfamily structures (Kimani et al. 2007) and most sequences (Soriano-Maldonado et al. 2011), but its catalytic role remains unclear. Due to the conservation of its location and coordination, there are speculations that it could be involved in catalysis, probably as a general base catalyst in the deacylation phase of the reaction (Kimani et al. 2007), essentially suggesting that the catalytic entity could be comprised of a CEEK tetrad rather than a triad. Hung and colleagues (2007) suggested that the second glutamate residue could be important for maintaining the configuration of the CEK triad or for facilitating docking of the substrates in the active site. All These proposals remain to be clarified. Soriano-Maldonado and colleagues (2011) have recently demonstrated that the second glutamate is essential for substrate hydrolysis in the formamidase from *Bacillus cereus* CECT 5050T, but its exact

function in the catalytic process remains unclear. The catalytic functions of the two active site glutamates remain to be elucidated. The wild-type (WT) NitN protein readily forms crystals and is therefore an excellent model for high resolution structural studies of the catalytic mechanism and substrate specificity in this family of enzymes.

One surprising observation of the NitN activity profile was its lack of enzymatic activity on ACR, even though it was able to hydrolyze PMD and FAE (Nel et al. 2011), short aliphatic amides that are similar in size to ACR. ACR is one of the best substrates for branch 2 aliphatic amidases (Makhongela et al. 2007), which makes branch 2 amidases attractive biocatalysts for bioremediation of ACR- and acrylonitrile-containing waste water from industries. The exact reason why NitN is unable to hydrolyze ACR is not known.

1.5.2 Study objectives

This study was aimed at addressing three main objectives:

- **To investigate the structural determinants of NitN specificity on short aliphatic amide substrates by analyzing binding and interactions of these molecules with the NitN binding pocket** - this is in an attempt to visualize the positioning of these molecules within the binding pocket and possibly understand activity and specificity of NitN for these molecules. The specific objectives include mutation of the catalytic cysteine (Cys165) in NitN to serine and alanine, purification of the mutant enzymes, binding and co-crystallization of the mutants with the short aliphatic amides, collection of X-ray diffraction data, and solving and refining co-crystal structures and their analyses.

- **To probe the catalytic role of the two active site glutamate residues (Glu61 and Glu139) using NitN as a model enzyme.** The specific objectives include, mutating Glu61 and Glu139 in NitN separately, purification of the mutant enzymes, incubating the mutants with the short aliphatic amides and performing mass spectrometric analysis to check for trapped reaction intermediates or adducts in the active site pocket, crystallization of the mutant proteins pre-incubated with the substrates, collection of X-ray diffraction data, and solution and refinement of structures and their analyses.

- **To monitor the activity, interactions and reactivity of the WT NitN and the Glu61 and Glu139 NitN mutants with ACR -** in order understand why NitN lacks hydrolytic activity on ACR. The specific objectives include enzymatic activity assays with ACR, mass spectrometric analyses of the WT and the two glutamates NitN mutants pre-incubated with ACR, crystallization of the NitN proteins pre-reacted with ACR, collection of X-ray diffraction data, and determination and refinement of structures and their analyses.

1.5.3 Study significance and impact

Despite its relatively poor activity on short aliphatic amides, NitN has a great potential as a biocatalyst in the fine chemical industry. Understanding the basis for its specificity on these substrates will pave way into identification and design of better substrates and its development for industrial use.

There are a large number of medically-relevant amidases. The outcome of this study will inform the design of compounds directed at inhibiting these, as well as informing the use of amidases for the synthesis of fine chemicals and pharmaceuticals.

Chapter 2

Materials and Methods

2.1 Preparation of *Nesterenkonia* sp. amidase (NitN) mutants

All mutants NitN (E139Q/L/A/D, C165A/S, E61Q/L, K131H/Q, E61Q-C165A and E61L-C165S) were made by site-directed mutagenesis using non-overlapping primers, with the forward primer bearing the desired mutation (Table 2.1). All primers were synthesized by Advanced DNA Technology (ADT), and they all contained a 5' phosphorylation modification. The wild-type (WT) structural gene of NitN was contained in a high copy number pET28a(+) expression plasmid (Novagen; with kanamycin resistance gene), pETNitN^{wt}, between *Nde* I and *Xho* I cloning sites, and downstream of an N-terminal hexaHistidine-tag. The purified pETNitN^{wt} plasmid containing the WT NitN gene was stored at -20 ºC.

Polymerase Chain Reaction (PCR) was performed to amplify the entire pETNitN^{wt} plasmid (and in the processing generating mutated genes) using the Phusion™ High-Fidelity PCR kit (New England Biolabs) under the conditions described by the manufacturer. 50 µl reactions were prepared containing 200 µm dinucleotide triphosphates (dNTPs), 0.5 µm of each primer, 10 ρg of template plasmid DNA, 0.02 U/µl Phusion™ Hot Start DNA polymerase and buffer. The thermocycling conditions were as follows: 1 cycle of initial denaturation at 95 ^{o}C for 30 sec followed by 25 cycles of denaturation at 98 ^{o}C for 10 sec, annealing at 60-70 ^{o}C for 30 sec and extension at 72 ^{o}C for 3 min and 40 sec and 1 cycle of final extension at 72 ^{o}C for 10 min before a final hold step at 4 ºC. The presence of a 6.4 kb amplified plasmid band was confirmed by gel electrophoresis using 0.8% agarose gels and HyperladderI (Bioline) DNA molecular weight markers.

The amplified expression plasmids bearing the mutations (pETNitN^{mut}) were ligated overnight at room temperature using T4 DNA ligase (Fermentas Life Sciences), after which they were transformed into chemically competent *Escherichia coli* (*E. coli)* DH5α cloning cells (Invitrogen) following standard transformation methods (Hanahan et al. 1991). The *E. coli* DH5α cells harbouring the plasmids were grown overnight with shaking at 37 ºC in 5 ml Luria Broth (LB) (Davis et al. 1980) cultures containing 25 µg/ml kanamycin (Kan) for plasmid selection. The plasmids were extracted and purified from overnight cultures using Biospin® plasmid extraction and purification kit following manufacturer's instructions. Incorporation of the desired mutations was confirmed by sequencing using T7 promoter and T7 terminator universal primers. In cases where T7 promoter and T7 terminator primed sequences did not

overlap well in the region of the mutation, sequencing was also performed using a designed internal forward primer (5'-GCCGAGCTCGCCGATGAA-3').

Table 2.1 Primers for introducing various point mutations in the NitN gene. Mutated codons (underlined) were introduced in the forward primers. For multiple mutations in a codon, the primers were designed in such a way that the reverse primer was shared. Prefixes Rev- and Forw- refer to reverse and forward primers respectively.

Mutation	Primers	Primer sequence
E139Q/A/L/D	Rev-E139	5'-CTCGGGTCCGTAGAGCTGGA-3'
	Forw-E139Q	5'-CAGAAGGCCGCCTTCGTCC-3'
	Forw-E139A	5'-GCCAAGGCCGCCTTCGTCCC-3'
	Forw-E139L	5'-CTGAAGGCCGCCTTCGTCCC-3'
	Forw-E139D	5'-GACAAGGCCGCCTTCGTCCC-3'
C165A/S	Rev-C165	5'-GACCAGCAGGCTCAGCT-3'
	Forw-C165A	5'-GCCTACGACGTGGAGTTCCC-3'
	Forw-C165S	5'-AGCTACGACGTGGAGTTCCC-3'
E61Q/L†	Rev-E61	5'-CGGGGTGAGCAGCAGCTGG-3'
	Forw-E61Q	5'-CAGCTCTTCGGCTTCGGCTACGTC-3'
	Forw-E61L	5'-CTGCTCTTCGGCTTCGGCTACGTC-3'
K131H/Q/	Rev-K131	5'-CTGGCAAGCACCTCTCCGTGTTC-3'
	Forw-K131H	5'-CTACCAGCACGTCCAGCTCTACGG-3'
	Forw-K131Q	5'-CTACCAGCAGGTCCAGCTCTACGG-3'
E61Q-C165A†	Rev-E61	5'-CGGGGTGAGCAGCAGCTGG-3'
	Forw-E61Q	5'-CAGCTCTTCGGCTTCGGCTACGTC-3'
	Rev-C165	5'-GACCAGCAGGCTCAGCT-3'
	Forw-C165A	5'-GCCTACGACGTGGAGTTCCC-3'
E61L-C165S†	Rev-E61	5'-CGGGGTGAGCAGCAGCTGG-3'
	Forw-E61L	5'-CTGCTCTTCGGCTTCGGCTACGTC-3'
	Rev-C165	5'-GACCAGCAGGCTCAGCT-3'
	Forw-C165S	5'-GCCTACGACGTGGAGTTCCC-3'

†The same primers that were used for the E61 and C165 single mutations were also used to prepare E61L-C165S and E61Q-C165A double mutants.

2.2 Expression of the mutant proteins

Chemically competent *E. coli* BL21 (DE3) strain cells (Invitrogen) were transformed with the pET28a(+)-derived expression plasmids, pETNitN^{mut}, carrying the structural gene for NitN with the desired mutation. The initial expression conditions were similar to those described by Nel and others (2011) for the WT NitN enzyme with some modifications, since most mutant proteins were largely expressed as insoluble aggregates in *E. coli*. In the optimized protocol, the *E. coli* BL21 (DE3) cells bearing the mutant DNA plasmids (pETNitN^{mut}) were grown overnight with aeration at 37 ºC in 5 ml LB precultures containing 25 µg/ml Kan, following inoculation with single colonies previously plated out on Luria Agar (LA) (Davis et al. 1980) also containing 25 µg/ml Kan. The overnight cultures were used to inoculate 0.5 litre volumes of 2X YT (2X Yeast extract and tryptone) rich medium supplemented with 0.5-1 % glucose and 25 µg/ml Kan. The cultures were incubated at 37 ºC with shaking until they reached mid-log phase (OD_{600nm} of 0.5). They were then transferred to a refrigerated incubator between 16 ºC and 20 ºC and allowed to stand for 20 minutes before induction with 0.1 mM Isopropyl-Beta-D-Thiogalactoside (IPTG). Protein expression was carried out at 16-20 ºC for 24 hours with aeration. Cells were harvested by centrifugation for 15 minutes at 5000 revolutions per minute (rpm) at 4 ºC and immediately stored at -20 ºC until further processing. 1 ml aliquot cultures were removed from each culture before and after induction, harvested, and resuspended in SDS-PAGE (sodium dodecyl sulfate - polyacrylamide gel electrophoresis) sample application buffer (SAB; 50 mM Tris-HCl pH 6.8, 2% sodium dodecyl sulfate (SDS), 10% glycerol, 1% β-mercaptoethanol, 12.5 mM EDTA and 0.02% bromophenol blue) before being frozen away at -20 ºC. The aliquoted samples were used to confirm protein expression using SDS-PAGE.

2.2.1 Co-expression of the E139Q/L mutants with the chaperones

E139Q and E139L NitN mutants were co-expressed with a number of chaperones in an attempt to improve soluble expression (Thomas et al. 1997). The chaperones were purchased from TaKaRa BIO INC. (Japan) encoded in various plasmids that carried a chrolamphenicol (Cm) resistance gene.

Co-transformation of *E. coli* BL21 (DE3) cells with the chaperones and the mutants expression plasmids

The chaperone plasmids, pG-KJE8 (*dnaK-dnaJ-grpE-groES-groEL*) (Nishihara et al. 1998;Nishihara et al. 2000) and pG-Tf2 (*groES-groEL-tig*) (Nishihara et al. 2000), were transformed separately into chemically competent BL21 (DE3) *E. coli* cells and the transformants (BL21-pG-KJE8 and BL21-pG-Tf2) selected on LA containing 35 µg/ml Cm. The *E. coli* BL21 (DE3) cells containing the two chaperone plasmids (BL21-pG-KJE8 and BL21-pG-Tf2) were made chemically competent using rubidium chloride method (Hanahan et al. 1991) and immediately transformed with the expression plasmids carrying the mutant gene of interest (pETNitN^{E139Q} and pETNitN^{E139L}). The 'double' transformants were selected with LA medium containing 25 µg/ml Kan and 35 µg/ml Cm.

Co-expression of the E139Q/L mutant genes with the chaperones

To co-express the mutant genes with the chaperones, the overnight precultures of co-transformants were inoculated into 2X YT medium containing 35 µg/ml Cm and 25 µg/ml Kan for plasmid selection, as well as 0.5 mg/ml L-arabinose (Sigma-Aldrich) and/or 1 ng/ml tetracycline (Sigma-Aldrich) for induction of chaperone expression. Both L-arabinose and tetracycline were used with pG-KJE8, while tetracycline only was used with pG-Tf2. The cultures were incubated at 37 ºC with aeration until they reached mid-log phase (OD_{600nm} of 0.5). Induction, expression, harvesting of cell and storage were carried out as described in section 2.2.0 above.

2.2.2 Expression of the E139Q/L mutants using TaKaRa cold shock pCold™ system

Cloning of E139L/Q mutant genes into pCold-I™ and pCold-TF™ plasmids

Expression vectors harbouring genes for E139L/Q mutants (pETNitN^{E139Q} and pETNitN^{E139L}) were extracted and purified as described previously (section 2.1). The plasmids were simultaneously digested with *Nde* I and *Xho* I (Fermentas Life Sciences) at 37 ºC for 2 hrs following manufacturer's instructions. The reactions were stopped by heating at 65 ºC for 10 min after which the digests were run on 0.8% agarose gel electrophoresis. The excised fragments of 1116 bp were cut out from the gels, extracted and purified using Biospin® DNA Gel extraction kit following manufacturer's instructions. The recipient pCold-I™ and pCold-TF™ expression vectors were also double-digested with the same restriction enzymes (*Nde* I

and *Xho* I) and de-phosphorylated immediately by incubating with fast alkaline phosphatase (FastAP, Fermentas Life Sciences) for 30 minutes at 37 ºC. The reactions were stopped by heating at 75 ºC for 10 minutes. The insert DNA fragments were ligated between *Nde* I and *Xho* I sites of the digested pCold-I^{TM} and pCold-TF^{TM} vectors using T4 DNA ligase (Fermentas Life Sciences), resulting in the following plasmids: pColdI-E139Q, pColdTF-E139Q, pColdI-E139L and pColdTF-E139L.

Expression of the E139Q/L mutants in pCold^{TM} plasmids

The pCold^{TM} vectors (pColdI-E139Q, pColdTF-E139Q pColdI-E139L and pColdTF-E139L) harbouring the NitN mutant genes were transformed into chemically competent BL21 (DE3) *E. coli* cells and transformants selected using LA containing 100 µg/ml ampicillin (Amp). Preparation of overnight pre-cultures, inoculation and growth of expression starter cultures were all done as described previously (section 2.2.0). On reaching mid-log phase (OD_{600nm} of 0.5), the cultures were transferred to 15 ºC and allowed to stand without shaking for 30 min. Expression was induced by addition of 0.1 mM IPTG and continued for the next 24 hrs at 15 ºC with aeration. Cells were harvested as described in section 2.2.0 above.

2.3 Cell lysis and protein purification

Cells were removed from the freezer and allowed to thaw on ice for 15 minutes before being resuspended in 1/50th the original culture volume in lysis buffer (20 mM Tris-HCl, pH 7.9, 60 mM imidazole, 500 mM NaCl and a protease inhibitor cocktail from Roche). Cells were then disrupted by sonication for 6 minutes (15 second cycles with 15 seconds cooling periods) using a Misonix sonicator fitted with a flat tip, and using ice/ethanol slurry to cool the sample. The supernatant (cell free extract (CFE)) was then clarified by centrifugation at 18 000 rpm for 20 minutes.

The mutant proteins were purified in a two-step protocol involving immobilized nickel affinity chromatography (AF) and size exclusion chromatography (SEC). All chromatography matrices were from GE Healthcare (formerly Pharmacia/Amersham) and standard reagents and chemicals were sourced from Sigma-Aldrich unless otherwise stated.

After clarification, the CFE was diluted 1:2 with the affinity column equilibration/binding buffer (20 mM Tris-HCl, pH 7.9, 60 mM imidazole, 500 mM NaCl) and filtered through a 0.45 µm filter. A 5 ml HisTrap™ HP affinity column (specific for the hexaHistidine tag) was equilibrated with 10 column volumes of the binding/equilibration buffer, before the supernatant was loaded onto the column using the same low imidazole buffer at a flow rate of 2 ml/min. After loading, the column was washed with 10 column volumes of binding buffer at 2 ml/min flow rate before bound proteins were eluted with a 90 ml linear gradient of increasing imidazole concentration (60 mM to 300 mM) at 3 ml/min flow rate. The gradient was prepared by mixing varying percentages of elution buffer (20 mM Tris-HCl, pH 7.9, 300 mM imidazole, 500 mM NaCl and occasionally 10% glycerol) with the binding in a two-chamber gradient mixer. The whole purification procedure was carried out on a Gilson 321 mixing pump, the elution of protein was monitored by a Waters 484 spectrophotometer at OD_{280nm} and fractions were collected automatically by a Gilson 204 fraction collector at 0.5 min/tube. The chromatographic procedure was controlled via Gilson Unipoint software (Version 3.4). Fractions with significant protein absorbance at OD_{280nm} were analyzed by SDS-PAGE; those containing the N-terminal hexaHistidine-tagged NitN fusion protein with a monomeric molecular weight (MW) of 30 kDa were pooled and concentrated in preparation for a gel filtration polishing step.

Preparative SEC was carried on a 24 ml HR Superdex 200 (S200) column pre-equilibrated with gel filtration (GF) buffer (50 mM Na_2HPO_4/NaH_2PO_4, pH 7.5, 150 mM NaCl, 2 mM DTT and occasionally 10% glycerol). The column was run at ambient temperature on a Gilson system equipped with a 305 pump and operating at 0.5 ml/min flow rate. Protein elution was monitored by the inline Gilson UV/VIS 151 spectrophotometer and fractions were collected using a Gilson 203b unit. The chromatographic procedure was controlled via Gilson Unipoint software (Version 3.4) and collected fractions were analyzed by SDS-PAGE. The Protein standards (Biorad) used for molecular mass calibration on SEC profiles were thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), and vitamin B12 (1.35 kDa).Tobacco mosaic virus (TMV) and 3% (v/v) acetone were included in the calibration standards to determine the void volume (V_0) and the total volume (V_t) of the column, respectively. The mass of the native protein was calculated from its retention time. A calibration standard curve that was used throughout the purification is shown in Appendix I. Fractions containing purified dimeric NitN protein were pooled and concentrated to desired concentrations for various biochemical experiments. Any protein material that was not used

immediately was buffer-exchanged into a GF buffer containing 15% glycerol, concentrated to over 10 mg/ml, and flash-frozen in liquid nitrogen in cryo-vials for storage at -80 ºC.

Concentration and buffer exchanges of large volumes (up to a volume of 5 ml) were performed in an Amicon (Bioseparations) stirred cell ultrafiltration system fitted with a Millipore 10 kDa molecular weight cut-off (MWCO) membrane and operating at the maximum permitted pressure at 4 °C. Smaller volumes were processed in 10 kDa MWCO Micro (3.5 ml) and Nanosep (500 µl) Centricon centrifugal concentrators from Pall Corporation.

2.4 Biochemical and Biophysical analyses of the mutants

2.4.1 Protein quantification

Protein concentrations were determined using the Bradford assay (Bradford 1976), with bovine serum albumin (BSA) as a standard. Bradford reagent was purchased from Bio-Rad and the manufacturers instructions were followed for the standard assay as designed for micro-titre plate applications. The results of the assays were read with a Titertek Multiscan plus MKII plate reader fitted with a 595 nm filter, and the data were processed in the Genesis Lite software. A BSA standard curve was calculated each time protein concentrations were determined. In some cases, a Nanodrop® 2000 machine was used to determined protein concentrations from 2 µl aliquots using $Protein_{280}$ module, with a theoretical extinction coefficient of 24500 and a 30 kDa monomeric MW.

2.4.2 SDS-PAGE of purified fractions

Discontinuous SDS-PAGE was carried out according to (Laemmli 1970) protocol using a Mini-PROTEAN 3 cell from Bio-Rad. Acrylamide solution (40%) from Bio-Rad was used to prepare 12% separating and 5% stacking gels. Protein samples were premixed with sample application buffer (SAB) and boiled for 4 mins in water bath before loading. Pre-stained molecular weight markers (Fermentas Life Sciences) were included in all electrophoretic runs. Gels were stained overnight at room temperature with 0.25% Coomassie Brilliant Blue R250 in 45% methanol and 10% acetic acid and destained for 3 hrs in a destain solution containing 30% methanol and 10% acetic acid. Protein purity and relative solubility were estimated manually by observing the gels under a light box.

2.4.3 Amide/nitrile hydrolysis activity assay

Enzyme activity was determined by the release of ammonia upon hydrolysis of amides or nitriles using a modification of indophenol-blue method by (Chaney and Marbach 1962). In this method, the generated ammonia reacts with hypochrite under alkaline conditions to form a monochloramine, which then reacts with phenol to form indophenol-blue in the presence of sodium nitroprusside (Na-nitroprusside) as a catalyst (Aminot et al. 1997).

Preparation of substrate samples

Amide substrates (propionamide (PMD), fluoroacetamide (FAE), butyramide (BMD), acetamide (ACE) and acrylamide (ACR)), were prepared as 1 M stock solutions by dissolving their powders in water or the reaction buffer (50 mM NaH_2PO_4/Na_2HPO_4, pH 7.5 and 150 mM NaCl). Adipamide was prepared as a 0.25M stock suspension in water. Further dilutions of the amides were made in reaction buffer where necessary. Nitrile substrates (propionitrile (PNT), fluoroacetonitrile (FNT), butyronitrile (BNT) and acetonitrile (ANT)) were diluted from their original stocks to 1 M stock solutions using 100% methanol, with further dilutions being made in water or reaction buffer where necessary.

Indophenol-blue ammonia assay

A standard reaction mixture (1 ml) contained the reaction buffer, 100 mM substrate and 1.25 µg of the enzyme. The reaction mixture was incubated at room temperature (~ 25 ˚C) for 1 hour, with 150 µl of aliquots being taken after every 10 minutes. After the withdrawal of the aliquots, the reaction was immediately stopped by the addition of 6 µl of 0.11 M phenolic alcohol with thorough mixing. The indophenol-blue colour development was initiated by the addition of 6 µl of 0.017 M Na-nitroprusside, followed by 15 µl of the oxidizing solution, containing 1 part sodium hypochlorite (NaOCl) and 4 parts alkaline complexing reagent (0.25 M NaOH and 0.775 M Trisodium citrate). This was followed by 1 hour incubation in the dark at room temperature to allow colour development. The indophenol-blue formed was measured spectrophotometrically in a micro-titre plate (in 100 µl volumes) using Titertek Multiscan plus MKII plate reader fitted with a 620 nm filter. The data (OD readings) were captured in the Genesis Lite software and processed in Microsoft Excel. Negative control reactions were carried out in the absence of the enzyme while the WT NitN reactions were performed as positive controls. An ammonia standard curve (an example in Appendix II) that was constructed using 0.04 - 0.4 mM ammonium chloride, was used to calculate the amount of

ammonia released from each reaction aliquot. All assays were performed in triplicate and an ammonia standard curve was calculated alongside the assay reactions at all times.

2.4.4 Mass spectrometry

The two active site glutamate mutants (E61Q/L and E139Q) were reacted at 1 mg/ml (0.0333 mM) with 100 - 500 molar excess (3.33 - 16.65 mM) of various amide/nitrile substrates at room temperature (25 ºC) for two hours. The reactions were stopped by transferring the reaction mixtures to 4 ºC, where they were stored until further processing. Reactions with the WT NitN enzyme were performed along side mutant enzyme reactions as a control at all times. Just before MS experiments, the samples were buffer-exchanged into the reaction buffer using 10 kDa MWCO Nanosep (500 µl) Centricon spin columns (Pall Corporation) to remove any unreacted substrates and to ensure proper buffering conditions for trypsin reactions. Tryptic digestions, sample clean-ups and MALDI-TOF (Matrix Assisted Laser Desorption Ionization - Time of Flight) mass spectrometric (MS) analyses were carried out by Dr. Maré Vlok at the Centre for Proteomic & Genomic Research (Cape Town, South Africa), using established standard protocols (Chapman 1996;Caprioli et al. 1996). Mass spectra were collected on an ABI 4800 MALDI TOF/TOF machine. Data processing was carried out in the GPS EXPLORER software from ABI and mass peaks of interest were identified and analyzed manually.

2.5 Protein crystallization

Crystallization of all mutants was performed in 96-well vapour diffusion sitting drop plates (Greiner Bio-One) using equal volumes of protein and precipitant (1 µl each) in the 2 µl crystallization drops over 100 ml of reservoir buffer at ambient temperature.

2.5.1 Crystallization of unbound (apo) NitN mutant proteins

The mutant proteins, from storage at -80 ºC, were thawed rapidly and diluted to 8, 6, 5 and 4 mg/ml using GF buffer containing 10% glycerol, and clarified by centrifugation at 13 000 rpm in a desktop microcentrifuge prior to crystallization. Initial crystallization was attempted using conditions identified by Nel at al. (2011) for the WT NitN enzyme. Where no crystallization

was observed, full screening was carried out using Crystal Screen HR2-110 and PEG/Ion HR2-126 Screen (Hampton Research) solutions to identify new conditions. The identified conditions were optimized by varying the precipitant concentration or the pH of the buffering substances where necessary.

2.5.2 Crystallization of the mutants/WT with the substrates

Stock solutions of substrates were prepared as described in section 2.4.3 above. The mutant proteins were rapidly thawed from -80 ºC storage, buffer-exchanged five times into the reaction buffer and re-concentrated using 10 kDa MWCO Nanosep concentrator columns (Pall Corporation). The protein at 6, 5 and 4 mg/ml concentration were reacted with 100 - 500 molar excess of the amide/nitrile substrates for 2 hours at room temperature (25 ºC), clarified by centrifugation at 13 000 rpm and stored at 4 ºC until crystallization. Crystallization of the NitN mutants protein pre-incubated with the substrates was carried out using vapour diffusion sitting drop method in a 2 µl crystallization containing equal volumes of the reacted protein and the precipitant solution. Initial crystallization experiments were performed using conditions identified for the unbound mutant protein, with screening for new conditions only being performed if no crystallization was observed or if the crystal quality needed to be improved. After substrates binding, the protein was kept on ice throughout the crystallization set-ups. The plates were incubated at room temperature for between a few days to several weeks, with regular check-ups on the progress in crystal formation and growth.

All assessments of crystal growth and manipulations were performed on a Leica MZ6 stereo microscope fitted with a CLS 150X cold light source. Photographs of crystals were captured with a Zeiss Axiocam camera attached to a Nikon Diaphot inverted fluorescence microscope fitted with 10x, 20x and 100x objective lenses.

2.6 X-ray diffraction data collection

Cryo-preservation

Suitable crystals were looped from their drops using LithoLoops™ (Molecular Dimensions) fitted in Molecular Dimensions Magnetic Cryo-Caps (bases) or Hampton HT Crystal Caps. The crystals were then dipped briefly in drops of cryo-protectant on a slide before being flash-frozen in liquid nitrogen contained in cryo-vials (Molecular Dimensions). Cryo-protectant solutions contained crystallization reservoir buffers supplemented with 20% glycerol and 50 mM substrates for mutant-substrate co-crystals.

Data collection

Data collection was performed at both PROXIMA I beamline at the Synchrotron Soleil, in Paris, France and BM14 beamline at the European Synchrotron Radiation Facility (ESRF), in Grenoble, France. At BM14, the data were collected at 100 K from single crystals with a single X-rays wavelength of 0.95373 Å, over 180 to 270 frames with 1º oscillations and three spindle passes per image. The crystals were mounted using a SC3 robotic sample changer and the diffraction images were recorded on a MAR CCD detector with an area of 225 x 225 mm. Data collection procedures were controlled by a MARCCD software platform.

At PROXIMA I beamline, data were collected from single crystals at 100 K using 0.918 Å single wavelength X-rays. Complete data sets were obtained from 180 to 270 frames with 1º oscillations interval and 4 passes per oscillation. The diffraction images were recorded on an ADSC Quantum 315r CCD detector with an area of 315 x 315 mm. The crystals were mounted manually and the data collection was controlled by an ADSC software platform.

2.7 Structure determination, refinement and model validation

Data processing and reduction

Integration, scaling and merging of most data sets were carried out using *XD*S (Kabsch 2010), *HKL2000* package (Otwinowski and Minor 1997) and *d*TREK* (Pflugrath 1999). In some cases, integration was performed with *iMOSFLM* (Battye et al. 2011) while Laue group and space group assignments were evaluated in *POINTLESS* (Evans 2006) within the Collaborative

Computational Project number 4 (CCP4) suite (CCP4 1994). Scaling, reduction, and merging of diffraction data from *iMOSFLM* was carried out using *SCALA* (Evans 2006) within the CCP4 suite (CCP4 1994).

Initial data analysis

Unit cell content analyses were carried out in the *MATTHEWS_COEF* (CCP4 1994) program within the CCP4 suite, and in cases where a crystal had a different space group from the $C222_1$ of the WT NitN structure, self-rotation functions were calculated in MOLREP (Vagin and Teplyakov 1997;CCP4 1994). In addition to data quality statistics obtained from the merging and scaling step of data processing, the *Xtriage* program within the *PHENIX* (Python-Based Hierarchical Environment for Integrated Xtallography) suite (Adams et al. 2010) was used to check for crystal twinning, translational pseudo-symmetry, possible outlier reflections (Read 1999) and ice rings-related problems in the merged reflections datasets.

Phase determination, model building and refinement

Phasing of all datasets was carried out by molecular replacement (MR) using *PHASER* (McCoy et al. 2007;CCP4 1994) and occasionally in *MOLREP* within the CCP4 suite (Vagin and Teplyakov 1997;CCP4 1994). The crystal structure of NitN WT protein (PDB id, 3hkx; (Nel et al. 2011)) at 1.66 Å resolution was used as an initial search model. The coordinates file was prepared by removing all non-protein atoms prior MR searches, and the correctness of the MR solutions was determined from the cross-correlation statistics, which indicate how well the oriented and positioned model in the unit cell matches the experimental data.

The solutions from MR searches were subjected to rigid body refinement in *REFMAC5* within CCP4 suite (Murshudov et al. 1997;CCP4 1994) over 10 maximum likelihood cycles. The MR solutions in the $P2_12_12_1$ space group having 6 subunits in the asymmetric unit (ASU) ($P2_12_12_1$) had each protomer being treated as a rigid body in this refinement. *COOT* (Emsley and Cowtan 2004) was used for checking the geometry for the initial model and for subsequent model rebuilding which involved adding residues on the terminal ends, adjustment of side chains and surface loops, addition of ions, water molecules and other monomers, real-space refinement of poor model segments, as well as identification, insertion and adjustment of the substrate molecules. To minimize potential model bias, sigma-A (σA) weighted composite omit maps were calculated while rebuilding initial models with *PHENIX* (Adams et al. 2010) using prime-

and-switch phasing strategy (Terwilliger 2004). The composite omit maps were used for initial model building before maximum likelihood restrained refinement was performed against the working set of crystallographic data, using *PHENIX*. Subsequent cycles of model rebuilding guided by σA weighted $2F_{obs}$-F_{calc} and F_{obs}-F_{calc} maps in *COOT* alternating with restrained refinement using *REFMAC5* and *PHENIX* were repeated, with the progress of model building being monitored by changes in crystallographic R_{cryst} and R_{free} after each cycle. Structure refinement was considered complete when all the model errors had been addressed, and when R_{free} ceased to decrease. Individual anisotropic atomic displacement parameters for all atoms excluding water molecules were refined in later rounds of refinement for data sets higher than 1.5 Å resolution.

Unmodeled density for substrates (in the co-crystal structures) and other monomers were identified and manually built in *COOT* after the rest of the model had been refined to almost completion. The substrate molecules coordinates were built using *SKETCHER* (CCP4 suite; (CCP4 1994)), while their parameter and topology files for refinement were generated by *LIBCHECK* (CCP4 suite; (CCP4 1994)).

The test set data (5%), put aside for cross-validation was excluded from the refinement process. In cases where the present data was of lower resolution than the starting MR model (PDB id, 3hkx), the test set was selected to consist of the same reflection indices that were used during the refinement of the starting model in order to minimize model bias. This was achieved by during data scaling and merging step, by copying the 'test' column (containing the randomly generated numerical flag used for the selection of test set) from the structure factor file that was used to solve the starting model structure, into the new crystal reflection data files.

Model validation

RAMPAGE program from the CCP4 suite (Lovell et al. 2003;CCP4 1994) was used to analyze the geometry of final structures, particularly the backbone dihedral angles, while *SFCHECK* (CCP4 suite; (CCP4 1994)) was used to analyze the final model against experimental data. *MolProbity* (Davis et al. 2007) was also used for validating the structure geometry, as well as for performing all-atom contacts analysis that among other things provide information on unfavorable atomic clashes thus highlighting local structural inaccuracies (Lovell et al. 2003). Coordinate error estimations and average Wilson temperature factors were taken from the model analysis statistics provided by *REFMAC5*.

2.8 Structure analysis and molecular visualization

The *BAVERAGE* program within CCP4 suite (CCP4 1994) was used to analyze the average B-factor values of each residue in the refined models. Superposition of different structures was done using the *MATCHMAKER* feature in *UCSF CHIMERA* (Meng et al. 2006;Pettersen et al. 2004), and occasionally *ALIGN* (Cohen 1997). Root mean square deviation (rmsd) values of various atom categories in superimposed structures were calculated by *Swiss-PdbViewer* (Guex and Peitsch 1997). The residues responsible for substrate molecules binding in the NitN-substrate co-crystal structures, the associated hydrogen bonds and close contacts were identified by *LIGAND EXPLORER* java applet from the RCSB PDB (Berman et al. 2000). The *FindHBond* feature within the *UCSF CHIMERA* (Pettersen et al. 2004) was also used for analysis of hydrogen bonds within the active site pockets of unbound (apo) and covalent adduct- and reaction intermediate-bearing NitN mutant structures. Analysis of pockets and cavities in the structures was carried out using the CASTp server (Binkowski et al. 2003). H++ server (Gordon et al. 2005) was used to estimate the pKa values of the active site residues in the refined models.

Unless otherwise stated, rendering of the molecular coordinates and electron density maps was carried out in *PyMOL* (DeLano 2004), and the models were colored following the Corey-Pauling-Kultun (CPK) coloring scheme.

2.9 Quantum Mechanical calculations

Quantum mechanical calculations were performed using the hybrid density functional theory method B3LYP as implemented in GAUSSIAN 09 (Frisch et al., 2009). Geometry optimizations were done *in vacuo* using the 6-311+G(2df,p) basis set with a multiplicity of 1. The charge on the complex was -2. All non hydrogen atoms with the exception of those of the active site cysteine with its thioester adduct and the water molecules were fixed in their crystallographically determined locations. The initial model was biased towards the observations made in the various crystal structures, in particular, a water was placed between the two active site glutamates (in the A position) and the carbonyl oxygen was located in the oxyanion hole.

Chapter 3

Substrate binding and specificity in NitN

3.1 Abstract

The aliphatic amidase from *Nesterenkonia* species (NitN) exhibits hydrolytic activity on short aliphatic amides (propionamide, fluoroacetamide, butyramide and acetamide) *in vitro*. The crystal structure of the WT enzyme shows that the active site is considerably larger than these substrates and suggests the possibility of larger substrates. Site-directed mutagenesis of the nucleophilic cysteine (Cys165) and X-ray crystallographic studies of the mutants in the presence of short aliphatic amides were undertaken with the aim of visualizing substrate molecules in the active site pocket. The catalytic cysteine was mutated to an alanine and a serine and both mutants (C165A and C165S) had similar solubility characteristics to the wild-type (WT) NitN enzyme. The crystal structure of C165A revealed that the active site environment was preserved while that of C165S had Ser165 side chain adopting a different conformation from that of Cys165, hence displacing water molecules and altering the active site configuration. The co-crystal structures of the C165A mutant with propionamide (PMD), butyramide (BMD) and fluoroacetamide (FAE) were determined to high resolution. In most structures two amide substrate molecules were observed binding in the NitN pocket back-to-back, indicating that NitN would be capable of binding larger substrates. The molecules at the catalytic site (at the bottom of the pocket) had their amide amino groups positioned by hydrogen bonding with the two active-site glutamates Glu61 and Glu139, while the carbonyl oxygen was located in the 'oxyanion hole' that is formed by the protonated amino group of Lys131 and the peptidic NH group of Tyr166. The C165S mutant co-crystal structures had no amide substrates in the active site, probably due to partial blockage of the active site by the side chain of Ser165. Co-crystallization of C165A mutant with propionitrile (PNT) revealed that the 'oxyanion hole' can also be filled by a nitrile nitrogen atom, which has implications on the mechanism of branch 1 nitrilases. The crystal structure of the catalytic lysine mutant (K131Q) had a possible adipamide (ADM) thioester intermediate at the catalytic cysteine, which provided hints on the possibility of NitN being able to hydrolyze ADM or related compounds. An activity assay using the WT NitN confirmed that ADM is indeed a better substrate for NitN than the known short aliphatic amides, which implies that NitN has the potential of being a biocatalyst for mass production of adipic acid, one of the precursors of nylon:6.6.

3.2 Introduction

The newly characterized nitrilase superfamily protein from *Nesterenkonia* species (NitN) was classified as an aliphatic amidase based on its *in vitro* preference for a number of amide substrates. NitN selectively hydrolyzes linear short aliphatic amides (propionamide, fluoroacetamide, butyramide and acetamide) (Fig. 3.1) *in vitro*, although at much slower rates than those observed in the branch 2 aliphatic amidases (Nel et al. 2011). It has no detectable activity on nitrile or carbamyl compounds and it surprisingly lacks acyl transferase activity (Nel et al. 2011) that is very common with the branch 2 aliphatic amidases that show a similar substrate specifity profile. Despite the observed *in vitro* amidase activity, the physiological function of NitN is not known.

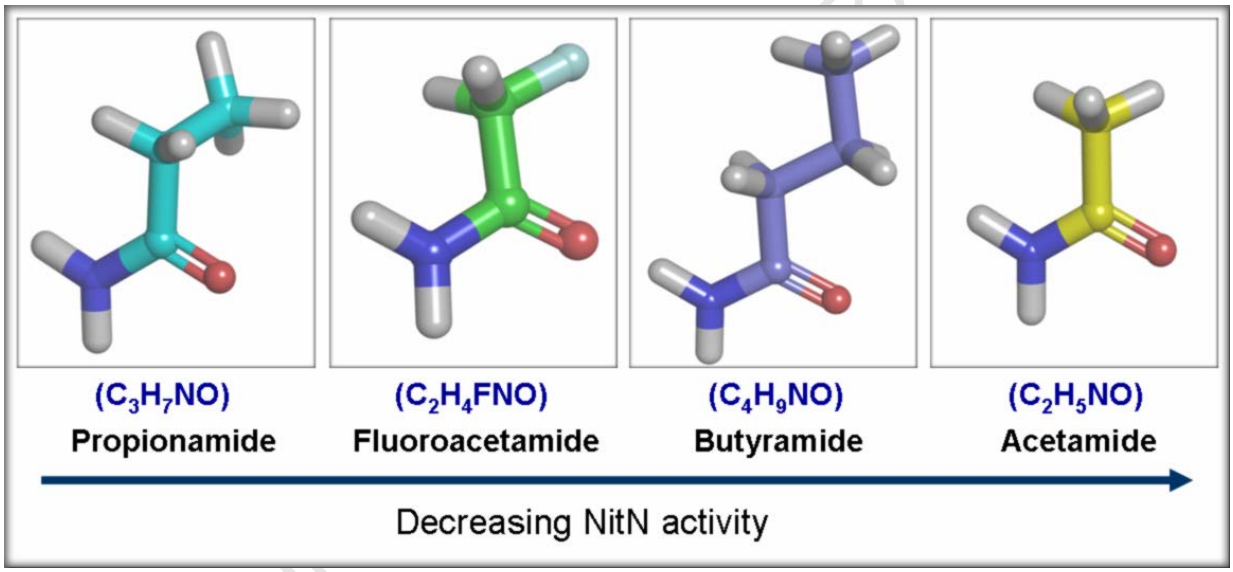


Fig. 3.1: Ball and stick rendition of the NitN short aliphatic amide substrates from the most efficiently hydrolyzed to the least preferred.

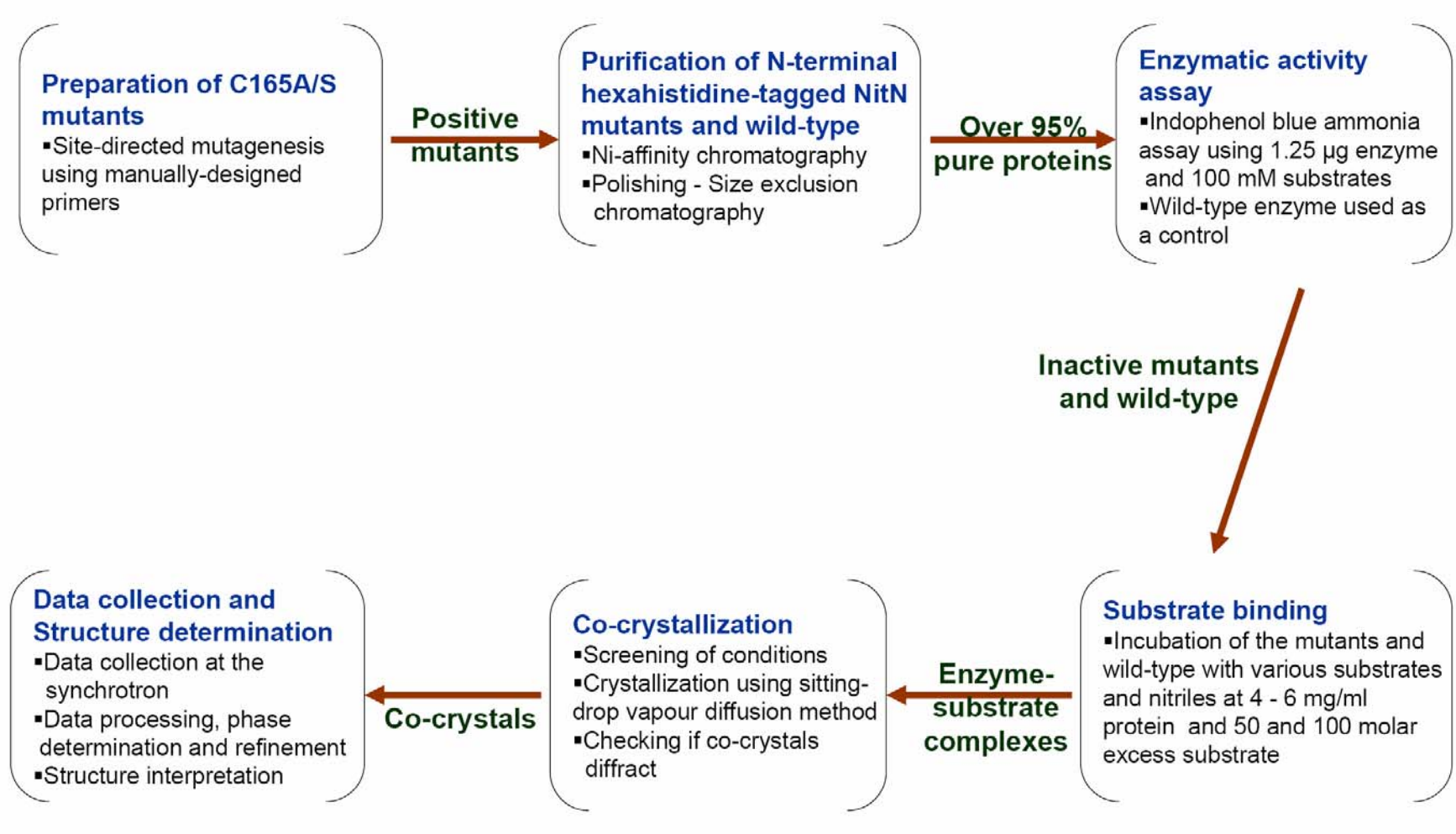
The size and geometry of the substrate binding pocket are important factors in determining substrate specificity in amidases, while the nature of residues lining the active site plays a key role in substrate recognition and binding. The substrate binding pocket of NitN is much larger and less constrained compared to the active sites of other aliphatic amidases (Nel et al. 2011), making it difficult to explain the preference for short aliphatic amides by this amidase. The work reported in this chapter was aimed at visualizing short aliphatic amide molecules in the active site pocket of NitN, in order to understand the structural determinants of NitN specificity on these substrates and to obtain insights into the positioning of aliphatic amides with more

than one carbon atom within the binding pocket. Co-crystallization of NitN with 'active' substrate molecules would require an inactive mutant enzyme. Mutation of the nucleophilic cysteine (Cys165) to an alanine and a serine was considered in this study as this would allow non-covalent binding of the amide substrates. The catalytic cysteine has been a target for mutation in substrate binding structural studies of amidases, and stable cysteine to serine or alanine mutants have successfully been made and crystallized previously (Chen et al. 2003;Hashimoto et al. 2003;Hung et al. 2007).

Preparation of NitN mutant proteins (C165A/S), binding of various amide and nitrile substrates to the mutants and co-crystallization of the complexes, X-ray crystallographic determination of co-crystal structures as well as structural analysis will be presented in this chapter. Structural studies on the binding of nitrile substrates to the wild-type (WT) enzyme as well as identification of better amide substrates for NitN will be also be reported.

3.3 Methods - A summary

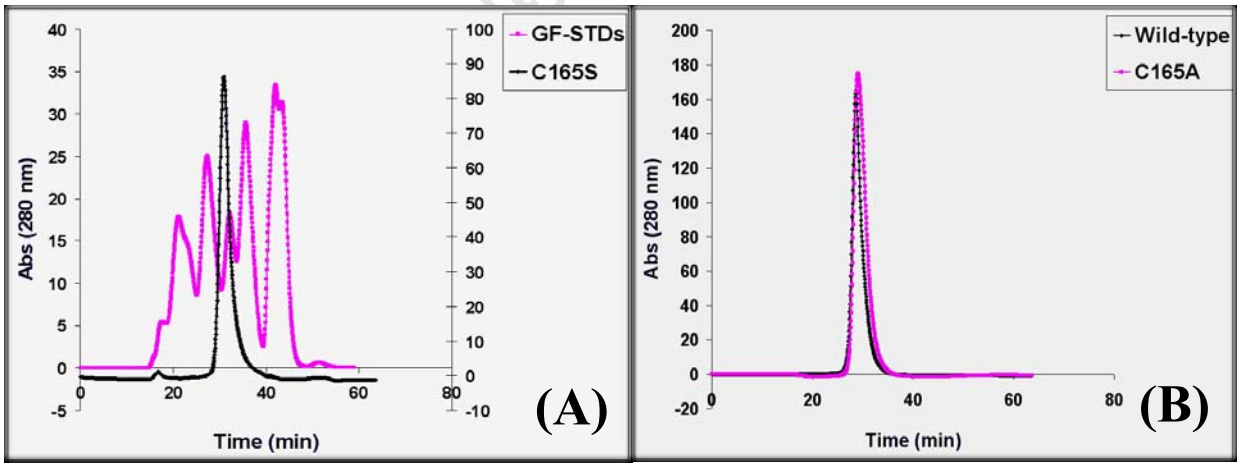
The steps followed in the preparation of C165S/A mutants, purification, crystallization and structure determination of NitN co-crystal structures. The wild-type NitN enzyme was purified following the same procedure and was used as a control in the characterization and crystallization steps. Details of each step are found in the materials and methods section (Chapter 2.0).



3.4 Results and discussion

3.4.1 Expression, purification and characterization of C165A and C165S mutants

The C165S and C165A mutations were incorporated into the WT NitN gene contained in a pET28a vector using site-directed mutagenesis, and the mutations confirmed by sequencing. Soluble expression of the hexaHis-tagged mutants in *E. coli* was enhanced by strategies that slow down protein expression rates, thus giving the cells enough time to fold protein that would otherwise form inclusion bodies. Expression was induced with low concentrations of IPTG (0.1 mM), followed by overnight expression at 20 °C. The mutants were purified to over 95% purity and homogeneity by Ni-chelating affinity chromatography and polished by a size exclusion chromatography step. The purified NitN mutants eluted off Superdex S200 gel filtration column as a single symmetric peak (Fig. 3.2), with an apparent molecular weight (MW) of 45 kDa. SDS-PAGE analyses of the peak fractions confirmed the presence of 30 kDa hexaHis-tagged proteins. The elution behavior of the dimeric mutant in which the apparent MW is 45 kDa instead of the expected 60 kDa proteins was also observed with the WT NitN enzyme. It was concluded that NitN most certainly interacts with the gel matrix, resulting in delayed elution. The WT enzyme was expressed and purified alongside the mutants following the same procedure.



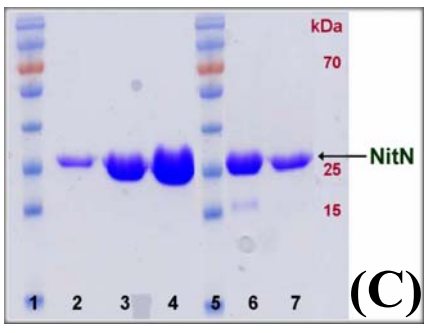


Fig. 3.2: Gel filtration (GF) purification of the WT and the C165A/S mutants of NitN amidase. **(A)** A superposition of the C165S mutant elution profile (black) with that of the calibration protein standards (pink). The C165S mutant elutes as a single symmetric peak of 45 kDa MW, with similar retention time to ovalbumin standard (44 kDa). **(B)** An overlay of the WT (black) and the C165A mutant (pink) GF elution profiles. The two proteins elute as a single peak of the same MW. **(C)** SDS-PAGE gel of the purified WT and C165 mutants. Lanes 1 and 5 are MW markers, lanes 2-3 are from the purified C165A fractions, while 6 and 7 are for the WT fractions. 30 kDa monomeric proteins are observed.

In addition to the observed gel filtration elution profiles (Fig. 3.2), the catalytic cysteine mutants were highly stable (based on their solubility profiles) during all *in vitro* manipulations; characteristics that were identical to the WT enzyme. This indicates preservation of conformational integrity and confirms that the mutations did not disrupt either the secondary or the quaternary structure of the proteins. The mutants did not show any detectable activity with any of the four amide substrates relative to the wild-type enzyme under the same assay conditions. This was expected given the role of the cysteine residue as a nucleophile in catalysis.

3.4.2 Substrate binding and crystallization of protein-substrate complexes

Co-crystallization was adopted as a strategy for crystallographic studies of NitN mutants in complex with various amide substrates. The hexaHis-tagged mutant proteins (4-6 mg/ml) were premixed with the substrates (at 10 molar excess) before setting up of the crystallization drops. The initial co-crystallization conditions tested were based on those used for the crystallization of the wild-type enzyme (Nel et al. 2011). The crystals obtained diffracted well but they did not have any substrates bound as evidenced by a lack of extra density in the active site following phasing by molecular replacement. This was thought to be due to low binding affinities of NitN for the short amide substrates, and also because of the presence of glycerol (10%) in the protein buffer, which was shown to inhibit the wild-type enzyme in biochemical activity assays.

In order to fully saturate the binding sites of the C165S/A mutants, propionamide (PMD) and propionitrile (PNT) were added at 50-100 molar excess while fluoroacetamide (FAE), butyramide (BMD) and acetamide (ACE) were used at 100-500 molar excess of the protein concentration (4-6 mg/ml). The mutants were incubated with the substrates for 2 hours at 25 °C, and then transferred to ice before crystallization. Glycerol was omitted from the protein buffer and this lead to the mutant enzymes failing to crystallize under the same conditions as the WT NitN, which is more stable and crystallizes readily in the presence of 10% glycerol. New crystallization conditions for the substrate-complexed mutant proteins (Table 3.1 (A)) were identified through screening using Crystal Screen HR2-110 and PEG/Ion Screen HR2-126 kits from Hampton research. The WT NitN enzyme was also crystallized at 5mg/ml both in the unbound form (apo) and in the presence of 33.33 mM PNT. Diffraction quality crystals of unbound and substrate-complexed WT and mutant proteins were obtained within 3-5 days of crystallization set-ups. A summary of the crystals that were selected for diffraction data

collection and structure determination is given in table 3.1 (B), and images of some of the crystals within their crystallization drops are presented in figures 3.3 (A) and (B).

Table 3.1 (A): Crystallization reservoir buffers for the WT and the C165A/S NitN co-crystals.

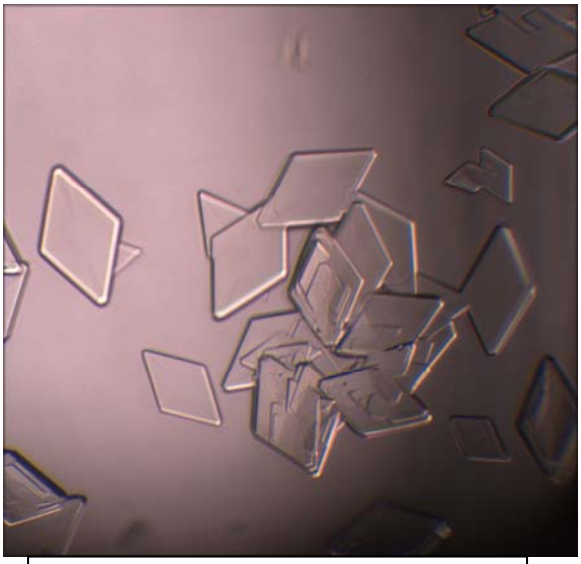
Crystallization buffer[a]	Contents
4	0.1 M Tris-HCl pH 8.5, 2.0 M ammonium sulfate
15	0.1 M sodium cacodylate trihydrate pH 6.5, 0.2 M ammonium sulfate, 30% polyethylene glycol (PEG) 8000
39	0.1 M HEPES sodium pH 7.5, 2% PEG400, 2.0 M ammonium sulfate
38P[b]	0.2 M ammonium tartrate dibasic, 20% PEG3350

[a]The reservoir buffers are named according to their numbering in the Hampton Research Crystal Screen HR2-110 kit.
[b]Buffer 38P is from the Hampton Research PEG/Ion Screen HR2-126.

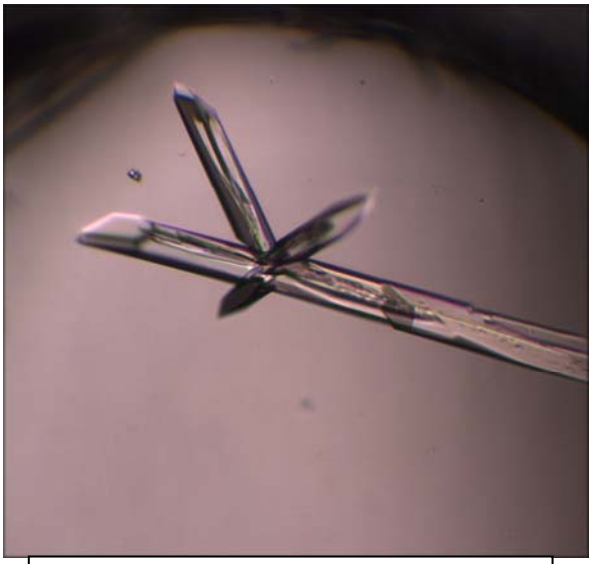
Table 3.1 (B): A summary of the crystals that were used for structure determination

Crystal name	NitN Protein	Substrate	Crystallization buffer
WT-Apo[&]	WT	-	32
WT-PNT	WT	PNT	4
C165A-Apo	C165A	-	15
C165A-PMD	C165A	PMD	39
C165A-BMD	C165A	BMD	39 /15[#]
C165A-FAE	C165A	FAE	38P
C165A-PNT	C165A	PNT	39
C165S-Apo	C165S	-	15
C165S-ACE	C165S	ACE	39
C165S-PMD	C165S	PMD	39
K131Q-ADM*	K131Q	ADM	32

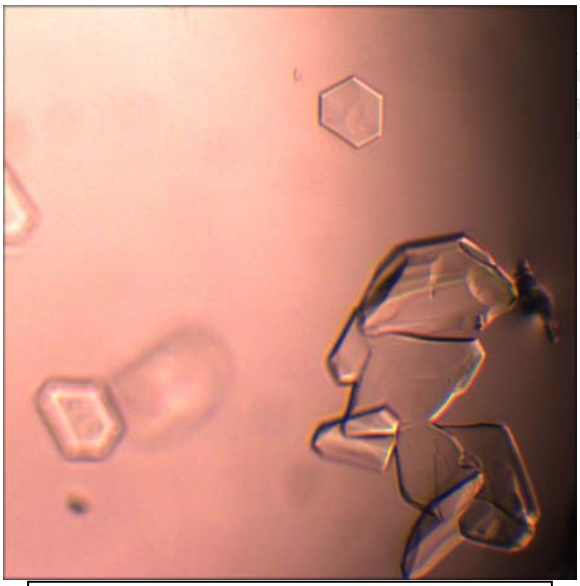
[&]The unbound WT enzyme (WT-Apo) was crystallized in an attempt to obtain a higher resolution NitN native structure.
[#]Two C165A-BMD crystal forms were obtained in different crystallization reservoir buffers.
*K131Q-ADM is a crystal from the catalytic lysine NitN mutant (K131Q) that had an unexpected adipamide (ADM) or adipamic acid reaction intermediate covalently attached to the catalytic cysteine (Cys165). The structure of the K131Q-ADM adducted protein provided hints on identification of a better substrate for NitN as detailed in section 3.3.11, even though the identity of the adduct (ADM or adipamide acid) has not been resolved. Although details of the K131Q mutant protein preparation have not been given in the text, this protein was purified and crystallized following the same protocols as the C165S/A mutants.



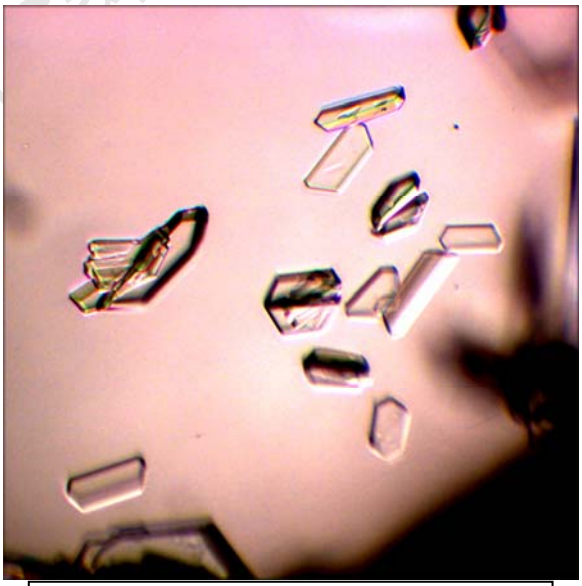
Crystals of the unbound WT NitN (WT-Apo)



Crystals of the C165A in the presence of FAE (C165A-FAE)

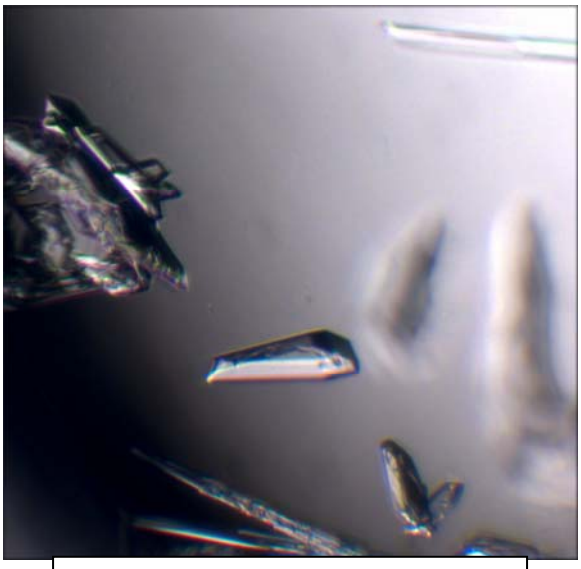


Crystals of the C165S in the presence of PMD (C165S-PMD)

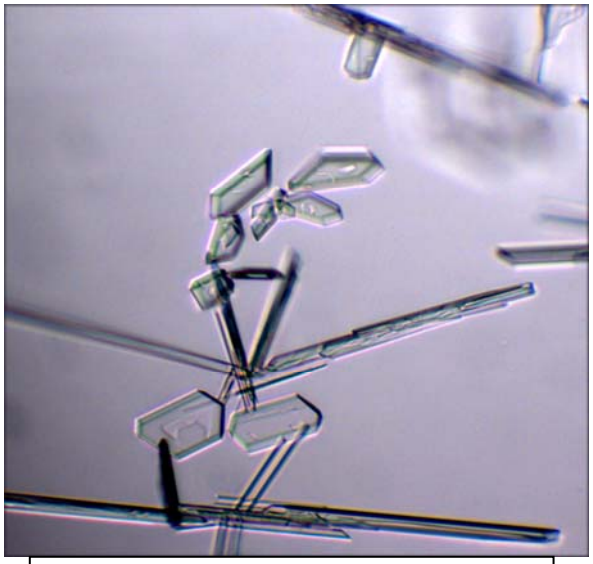


Crystals of the unbound C165A NitN (C165A-Apo)

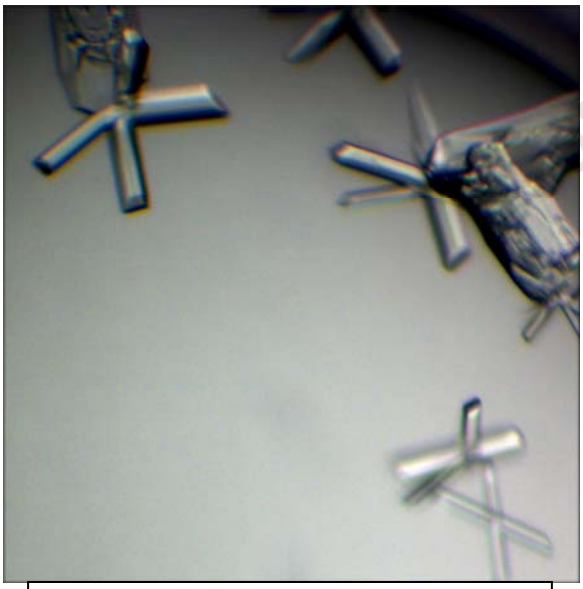
Fig. 3.3 (A): Images of crystals for some of the structures reported in this chapter.



Crystals of the C165A in the presence of Pnt (C165A-Pnt)



Crystals of the C165A in the presence of PMD (C165A-PMD)



Crystals of the unbound C165S NitN (C165S-Apo)

Crystals of the C165A in the presence of BMD (C165A-BMD)

Fig. 3.3 (B): Images of crystals for some of the structures reported in this chapter.

3.4.3 X-ray data collection and processing

Diffraction data were collected at the synchrotron using single crystals of the cysteine mutants and the wild-type enzyme grown in the presence and absence of various amide substrates. Even though all crystals diffracted to much higher optical resolution, the resolution cutoff for structure determination was determined based on the completeness and multiplicity of the data, as well as a signal-to-noise ($I/\sigma(I)$) ratio of about 2.0 and merging R-factor (R_{merge}) values of less than 50% in the highest (outer) resolution shell. These criteria ensured that very weak reflections were omitted from the datasets. All datasets were of high quality as indicated by the data collection statistics presented in tables 3.2 (A) and (B), and were therefore suitable for co-crystal structure determination. The overall and high resolution shell data had completeness levels approaching 100% in all cases, indicating that almost all of theoretically expected unique reflections had been measured; the observed high redundancy values indicate that the unique reflections had been measured multiple times, which is a requirement for good-quality structures; high average ($I/\sigma(I)$) values indicate high quality data with minimal errors ; and low overall R_{merge} values indicate high accuracy in the measurement of identical reflections. Analysis of the merged datasets with the *Xtriage* program (Adams et al. 2010) found no possible outliers among acentric and centric reflections (Read 1999), and no ice rings-related problems were detected in the ice ring-sensitive resolution ranges.The space group and unit cell parameters were the same as those of the WT NitN structure (PDB id, 3hkx) except for the C165A-FAE crystal which had $P2_12_12_1$ space group symmetry with six monomers in the asymmetric unit (ASU). The change in symmetry is thought to be as a consequence of the crystallization buffer (38P; see table 3.1 (A)) rather than of substrate binding.

Table 3.2 (A): Data collection statistics (I)

Data collection statistics						
Dataset	**C165A-Apo**	**C165A-PMD**	**C165A-BMD^{100}**	**C165A-BMD^{500}**	**C165A-FAE**	**C165A-PNT**
Crystallization buffer	15	39	39	15	38P	39
Synchrotron source	BM14[a]	BM14[a]	BM14[a]	Proxima I[b]	Proxima I[b]	BM14[a]
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$P2_12_12_1$	$C222_1$
Unit cell parameters	*a* = 75.120 (Å) *b* = 115.45 (Å) *c* = 65.610 (Å) *α*=*β*=*γ*= 90.0°	*a* = 75.100 (Å) *b* = 115.30 (Å) *c* = 65.350 (Å) *α*=*β*=*γ*= 90.0°	*a* = 75.180 (Å) *b* = 114.90 (Å) *c* = 65.080 (Å) *α*=*β*=*γ*= 90.0(°)	*a* = 75.570 (Å) *b* = 115.32 (Å) *c* = 65.170 (Å) *α*=*β*=*γ*= 90.0°	*a* = 116.42 (Å) *b* = 124.59 (Å) *c* = 131.21 (Å) *α*=*β*=*γ*= 90.0°	*a* = 75.470 (Å) *b* = 115.49 (Å) *c* = 65.190 (Å) *α*=*β*=*γ*= 90.0°
Cell content analysis						
Unit cell volume (Å^3)	569009.063	565894.813	562151.625	567886.375	1903169.000	568217.750
V_m (Å^3/Da)[c]	2.37	2.36	2.34	2.37	2.64	2.37
Molecules in ASU[d]	1	1	1	1	6	1
Solvent content (%)	48.15	47.87	47.52	48.05	53.50	48.08
Resolution range (Å)[e]	26.42 - 1.50 (1.58 - 1.50)	24.65 - 1.55 (1.63 - 1.55)	23.55 - 1.44 (1.52 - 1.44)	37.78 - 1.50 (1.55 - 1.50)	54.93 - 2.30 (2.42 - 2.30)	28.86 - 1.51 (1.59 - 1.51)
No. of observed reflections	432957	273901	367058	250144	890010	331475
No. of unique reflections	45894	41219	51279	45452	85348	44728
Completeness (%)[e]	100.0 (100.0)	99.6 (98.8)	99.9 (99.8)	99.0 (98.7)	100.0 (100.0)	99.6 (99.1)
Redundancy[e]	9.4 (9.4)	6.6 (6.0)	7.2 (7.0)	5.50 (5.46)	10.4 (10.6)	7.4 (7.4)
Mean $(I)/\sigma(I)$[e]	27.0 (10.6)	17.1 (4.4)	15.1 (4.1)	12.0 (2.8)	12.9 (5.0)	18.5 (5.4)
R_{merge} (%)[e, f]	5.20 (20.1)	6.80 (42.1)	6.90 (38.5)	6.10 (42.0)	17.1 (54.8)	6.70 (36.8)

[a]BM14 - Macromolecular beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble.
[b]Proxima I - Macromolecular beamline at the Synchrotron Soleil, Paris.
[c]V_m is the Matthew's coefficient, which allows estimation of the number of protein subunits in the asymmetric unit. $V_m = V/(MW*n)$, where V is the volume of the unit cell, MW is the molecular mass of the molecule and n is the number of asymmetric units in the unit cell.
[d]ASU - asymmetric unit. [e]Values in parentheses are for the highest resolution shell. [f]$R_{merge} = [\Sigma_h\Sigma_j |I_j(h) - \langle I(h)\rangle|] / \Sigma_h\Sigma_j |I_j(h)|$, where $I_j(h)$ and $\langle I(h)\rangle$ are the jth and the mean measurements of the intensity of reflection h, respectively.
C165A-BMD^{100} and C165A-BMD^{500} were datasets for crystals grown in the presence of 100 and 500 molar excess of butyramide respectively.

Table 3.2 (B): Data collection statistics (II)

Data collection statistics Dataset	**C165S-Apo**	**C165S-ACE**	**C165S-PMD**	**WT-Apo**	**WT-PNT**	**K131Q-ADM**
Crystallization buffer	15	39	39	32	4	32
Synchrotron source	BM14[a]	Proxima I[b]	BM14[a]	Proxima I[b]	BM14[a]	BM14[a]
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$C222_1$
Unit cell parameters	*a* = 75.400 Å) *b* = 114.10 Å) *c* = 65.590 Å) *α*=*β*=*γ*= 90.0°	*a* = 75.180 (Å) *b* = 115.64 (Å) *c* = 65.380 (Å) *α*=*β*=*γ*= 90.0°	*a* = 74.920 (Å) *b* = 114.69 (Å) *c* = 65.080 (Å) *α*=*β*=*γ*= 90.0°	*a* = 74.880 (Å) *b* = 115.83 (Å) *c* = 65.590 (Å) *α*=*β*=*γ*= 90.0°	*a* = 75.170 (Å) *b* = 115.65 (Å) *c* = 65.620 (Å) *α*=*β*=*γ*= 90.0°	*a* = 76.500 (Å) *b* = 115.04 (Å) *c* = 64.730 (Å) *α*=*β*=*γ*= 90.0°
Cell content analysis						
Unit cell volume ($Å^3$)	568694.875	568401.625	559160.875	569012.313	570470.688	569660.313
V_m ($Å^3$/Da)[c]	2.37	2.37	2.33	2.37	2.38	2.37
Molecules in ASU[d]	1	1	1	1	1	1
Solvent content (%)	48.12	48.10	47.24	48.15	48.29	48.21
Resolution range (Å)[e]	23.65 - 1.47 (1.55 - 1.47)	34.30 - 1.55 (1.63 - 1.55)	24.58 - 1.46 (1.54 - 1.46)	32.80 - 1.50 (1.55 - 1.50)	28.53 - 1.57 (1.65 - 1.57)	32.37 - 1.90 (2.00 - 1.90)
No. of observed reflections	337984	463567	347713	377930	287224	168708
No. of unique reflections	48675	41728	48952	45898	40234	22924
Completeness (%)[e]	99.9 (99.9)	100.0 (100.0)	99.9 (99.6)	99.8 (100.0)	99.9 (99.8)	100.0 (100.0)
Redundancy[e]	6.9 (6.7)	11.1 (11.2)	7.1 (6.6)	8.23 (8.20)	7.1 (6.9)	7.4 (7.4)
Mean $(I) /\sigma(I)$[e]	17.0 (4.5)	18.7 (5.9)	17.5 (4.5)	14.9 (4.9)	14.4 (4.5)	12.7 (3.3)
R_{merge} (%)[e, f]	6.60 (37.5)	8.90 (40.3)	6.20 (37.9)	5.6 (33.1)	8.50 (38.6)	11.4 (51.6)

[a]BM14 - Macromolecular beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble.
[b]Proxima I - Macromolecular beamline at the Synchrotron Soleil, Paris.
[c]V_m is the Matthew's coefficient, which allows estimation of the number of protein subunits in the asymmetric unit. $V_m = V/(MW*n)$, where V is the volume of the unit cell, MW is the molecular mass of the molecule and n is the number of asymmetric units in the unit cell.
[d]ASU - asymmetric unit. [e]Values in parentheses are for the highest resolution shell.
[f]$R_{merge} = [\Sigma_h\Sigma_j |I_j(h) - \langle I(h)\rangle|] / \Sigma_h\Sigma_j |I_j(h)|$, where $I_j(h)$ and $\langle I(h)\rangle$ are the jth and the mean measurements of the intensity of reflection h, respectively.

3.4.4 Structure solution, model building and refinement

The initial phases for all datasets were determined by molecular replacement (MR) using the atomic coordinates of the wild-type NitN structure (PDB id, 3hkx; (Nel et al. 2011)) as the search model. The MR solutions were clear in all cases as indicated by cross-correlation scores. Inspection of the initial electron density maps indicated the structural differences between the wild-type, substrate-complexed and uncomplexed mutants and also highlighted conformational differences in the solvent accessible residues. Most of the structures were determined to better than 1.8 Å resolution as shown in the model refinement and validation statistics tables (Tables 3.3 (A) and (B). The agreement of the models with the observed data was improved through alternating cycles of manual building and restrained refinement, the progress of which was monitored by changes in crystallographic R-factor (R_{cryst}) and the cross-validation free R-factor (R_{free}) (Brunger 1992) values. The refinement of the structures was considered complete when all the errors highlighted by the σA-weighted F_{obs} - F_{calc} difference electron density maps had been addressed and when R_{free} ceased to decrease. The substrates were built into the electron density present in the difference maps of the active site region after the protein component had been refined to near completion. Although the $P2_12_12_1$ asymmetric unit of the C165A-FAE data had six monomers related by non-crystallographic symmetry (NCS), NCS averaging was not utilized during refinement as each chain was different in the number of resolvable residues and the presence or absence of bound substrate molecules.

The first 20 residues in the protein chain are part of an N-terminal hexaHis-tag and the linker region, and 10 of these were missing in all structures. In addition, no interpretable density was visible in the helix region made up of residues 258 to 268. These residues are also missing in the WT NitN structure. Since there is no indication from mass spectrometry that they could be missing in the protein sequence, the two regions are therefore likely to be disordered in the crystal lattice. The hexaHis-tag is likely to have undergone auto-cleavage during crystallization. The side chains for Glu79, Arg89, Glu110, Gln111, Glu121, Glu138, Glu147, Gln240, Glu244, Glu272 and Gln277 have high B-factors (individual isotropic atomic displacement parameters) in all structures, with some of them having alternative conformers, indicating a high degree of structural variability. These residues are on the surface of the molecule, where they interact with the solvent and may take a variety of conformations. Residues 13-20 of the hexaHis-tag linker region are also disordered in the models where some or all of them are visible. Leu153 was reported to have a poor rotamer in

most of the final models according to *MolProbity* (Davis et al. 2007) validation criterion, while Ser165 in both C165S-Apo and C165S-Prop mutant structures had disallowed backbone dihedral angles. The catalytic cysteine has a highly constrained backbone conformation in all crystal structures of the nitrilase superfamily enzymes; a feature that is thought to be important for maintaining the cysteine side chain for catalysis (Nakai et al. 2000).

All models were of high quality as indicated by the validation statistics in tables 3.3 (A) and (B). The final R-factor values are what would be expected for data of the observed resolution range. R_{cryst} values of less than 20% in all cases indicate a good agreement between the final model and the experimental data, and small differences between R_{cryst} and R_{free} values (less than 5% in all cases) are an indication of correct modeling, with very minimal over-fitting of the model to the data. Figure of merit values of greater than 0.9 in most cases indicate high accuracy of the final models. The Ramachandran plots (Ramachandran et al. 1963) from both *RAMPAGE* (Lovell et al. 2003) and *MolProbity* (Davis et al. 2007) had over 99.5% of all residues falling in the favoured and allowed regions, an indication of correctness in the backbone geometry of all structures. The models had reasonable bond lengths and bond angles, as judged from the root mean square deviation (rmsd) values that are within the acceptable range for well-defined structures of similar resolution. The atomic clashes indicated by the *MolProbity* all-atom clash score involve the hydrogen atoms that were added by the program during the validation process in all cases. The Ramachandran plots for all residues in the final models are presented in Appendix III.

Table 3.3 (A): Model refinement and validation statistics (I)

Dataset	C165A-Apo	C165A-PMD	C165A-BMD[100]	C165A-BMD[500]	C165A-FAE	C165A-PNT
Refinement statistics						
Resolution range	24.71 - 1.50	24.36 - 1.55	23.37 - 1.45	31.60 - 1.50	53.21 - 2.3	28.53 - 1.51
No. of reflections in working set (95%)	41377	37228	45076	40865	806844	40271
No. of reflections in test set (5%)	2316	2135	2533	2288	4213	2281
No. of atoms						
Protein[b]	1966	1989	2021	1971	11753	2013
Ligand/ion	0	24	27	6	25	26
Water	293	291	244	235	1187	252
Missing residues	A1-19 A259-266	A1-12 A17 A258-267	A1-12 A258-267	A1-13 A258-268	A1-12, A259-267 B1-10, B260-267 C1-17, C259-267 D1-10, D259-267 E1-20, E259-267 F1-12, F259-267	A1-12 A258-267
R_{cryst} (%)[c]	13.13	13.45	12.81	14.29	16.39	13.30
R_{free} (%)[d]	16.49	17.41	16.05	18.07	21.60	16.68
Overall figure of merit	0.924	0.909	0.922	0.899	0.879	0.915
Average B-factor (Å^2)	14.12	15.82	17.81	19.15	19.14	16.44
Model validation						
Rms[e] deviations from ideality						
Bond lengths (Å)	0.03	0.03	0.03	0.03	0.02	0.03
Bond angles (°)	2.23	2.28	2.46	2.24	1.89	2.30
Bad rotamers (%)	0.50	0.50	0.50	0.50	1.30	0.50
Ramachandran outliers (%)	0.00	0.00	0.00	0.00	0.10	0.00
All-atom clashscore	3.95	5.37	5.59	3.10	n.d[a]	5.60

[a]n.d - not determined. [b]Number of protein atoms include duplicated atoms in alternate conformations.
[c]$R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$ where F_o and F_c are observed and calculated structure factor amplitudes of reflection h of the working set of reflections, respectively.
[d]R_{free} is equal to R_{cryst} for h belonging to the test set of reflections. [e]Rms - root mean square.

C165A-BMD^{100} and C165A-BMD^{500} were datasets from crystals grown in the presence of 100 and 500 molar excess of BMD respectively.

Table 3.3 (B): Model refinement and validation statistics (II)

Dataset	**C165S-Apo**	**C165S-ACE**	**C165S-PMD**	**WT-Apo**	**WT-PNT**	**K131Q-ADM**
Refinement statistics						
Resolution range	23.22 - 1.47	32.69 - 1.55	24.29 - 1.46	32.80 - 1.50	28.43 - 1.57	32.37 - 1.90
No. of reflections in working set (95%)	43719	37539	43936	41068	36159	20618
No. of reflections in test set (5%)	2509	2084	2462	2305	2056	1145
No. of atoms						
Protein[a]	1998	1996	2015	2009	2011	1980
Ligand/ion	0	31	12	5	15	23
Water	270	273	246	228	290	132
Missing residues	A1-12 A17 A258-267	A1-13 A259-267	A1-12 A258-267	A1-12 A17-18 A260-267	A1-12 A259-267	A1-12 A258-267
R_{cryst} (%)[b]	13.22	13.72	13.89	13.62	14.19	16.88
R_{free} (%)[c]	15.83	17.41	16.88	17.22	16.59	20.73
Overall figure of merit	0.927	0.922	0.922	0.913	0.920	0.876
Average B-factor (Å^2)	15.59	14.15	18.42	13.12	15.58	23.91
Model validation						
Rms[d] deviations from ideality						
Bond lengths (Å)	0.03	0.03	0.03	0.03	0.02	0.03
Bond angles (°)	2.32	2.36	1.93	1.93	1.59	2.21
Bad rotamers (%)	0.50	1.50	0.50	0.00	0.50	1.50
Ramachandran outliers (%)	0.40	0.40	0.40	0.00	0.00	0.00
All-atom clashscore	3.63	7.15	5.39	2.58	5.35	4.38

[a]Number of protein atoms include duplicated atoms in alternate conformations.
[b]$R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$ where F_o and F_c are observed and calculated structure factor amplitudes of reflection h of the working set of reflections, respectively.
[c]R_{free} is equal to R_{cryst} for h belonging to the test set of reflections.
[d]Rms - root mean square.

3.4.5 The unbound C165A-Apo and C165S-Apo structures

3.4.5.1 The overall and the active site structure

The secondary structure organization and the dimeric oligomeric form were preserved in all unbound and substrate-complexed mutant structures. Superposition of the Cα atoms of the unbound cysteine mutant structures (C165A-Apo and C165S-Apo) with the WT-Apo NitN structure (Fig. 3.4 (A) and (C)) showed identical conformation (root mean square deviation (rmsd) of 0.163 Å for both mutants), indicating that the substitution of cysteine with serine or alanine did not perturb the overall structure of the mutants. In the C165A-Apo structure, two active site water molecules (WatA and WatB) are observed; the positions of these water molecules are conserved in the WT-Apo structure although their exact locations are slightly shifted due to the mutation (Fig. 3.4 (B)). The hydrogen bonding networks within the WT-Apo and the C165A-Apo NitN binding pockets are very similar, except for a few polar contacts that are missing in the C165A mutant active site due to the replacement of the polar catalytic cysteine side chain with a non-polar side chain from the incorporated alanine at position 165 (Fig. 3.4 (B)) .

The mutation of the catalytic cysteine (Cys165) to a serine alters the active site environment considerably, as the orientation of Ser165 side chain within the binding pocket results in the displacement of the two active site water molecules that are observed in the WT-Apo structure, which ultimately changes the active site hydrogen bonding network in the C165S-Apo structure (Fig. 3.4 (D)). Unlike the sulfhydryl group of Cys165 which interacts with the carboxyl oxygen atoms of Glu61 through a coordinating water molecule in the WT-Apo NitN structure, the hydroxyl group of Ser165 in the C165S-Apo structure hydrogen bonds to the carboxyl oxygen atoms of Glu61 directly. Despite these changes, the active site conformation of the C165S structure is highly similar to that of the WT-Apo structure (Fig. 3.4 (D)).

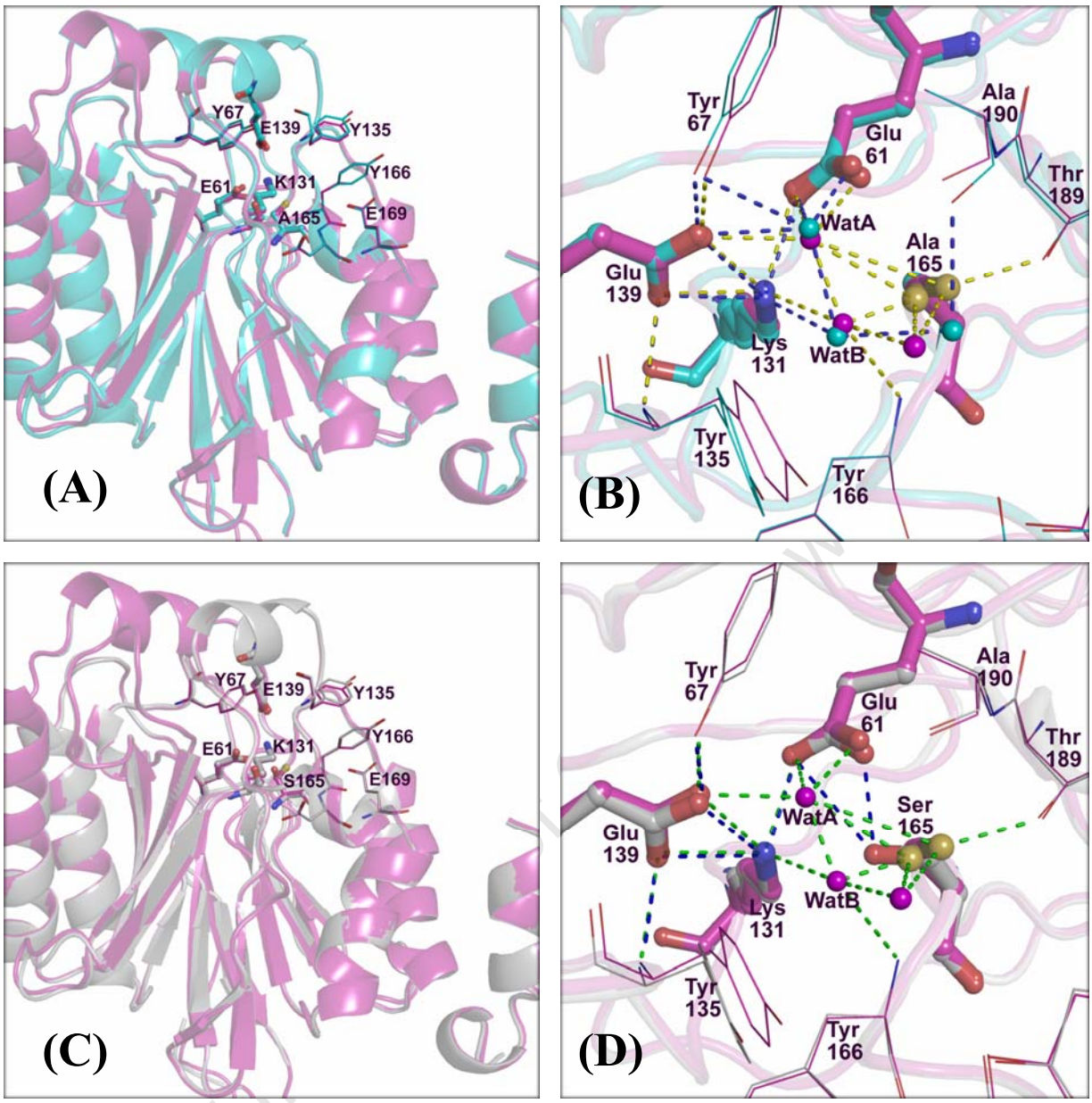


Fig. 3.4: Comparison of the overall structure and the active site environment of the WT NitN and the unbound C165A/S mutants. **(A)** Cartoon representation of the superimposed monomers of the C165A-apo (cyan CPK) and the WT-Apo (magenta CPK; PDB id, 3hkx) structures. There are virtually no differences on the backbone conformation as observed from the near perfect overlay of the cartoon models. The binding site location is indicated by ball and stick rendition of the active site residues. **(B)** Line and ball and stick representation of the active site region of the aligned C165A-Apo (cyan CPK) and the WT-apo (magenta CPK) structures. Magenta non-bonded spheres and yellow dashes are water molecules and potential hydrogen bonding interactions in the WT-Apo active site respectively, while cyan non-bonded spheres and blue dashes are water molecules and potential polar contacts in the C165A-Apo binding pocket respectively. **(C)** Cartoon representation of the backbone alignment of C165S-Apo (grey CPK) and the WT-Apo (magenta CPK) structures. **(D)** Comparison of binding site environment of the C165S-Apo (grey CPK) and WT-Apo (magenta CPK) structures. Magenta non-bonded spheres and green dashes are water molecules and potential hydrogen bonding interactions in the WT-Apo active site respectively, while blue dashes are potential polar contacts in the C165A-Apo binding pocket respectively.

3.4.6 PMD- and BMD-complexed C165A mutant structures

3.4.6.1 The binding of BMD and PMD to the C165A mutants

The first σA-weighted F_{obs} - F_{calc} (difference) electron density maps from the C165A-PMD and C165A-BMD co-crystals showed clear density for the two substrates. This density (Fig. 3.5 (A) and (C)) persisted throughout the model building and refinement process, with the substrate molecules only being incorporated after the protein component had been adjusted and all ordered water molecules added. Two PMD molecules are bound in the active site, one deep in the pocket and a second one close to the mouth of the pocket (Fig. 3.5 (A) and (B)). The BMD C165A co-crystals had ordered density for one BMD molecule deep in the pocket, and some weak (incomplete) density (Fig. 3.5 (C)) for a second molecule in a location similar to the second PMD molecule. The second BMD molecule was however abandoned in the final refinement cycle as it had high atomic B-factors, indicating high flexibility (instability) or partial occupancy.

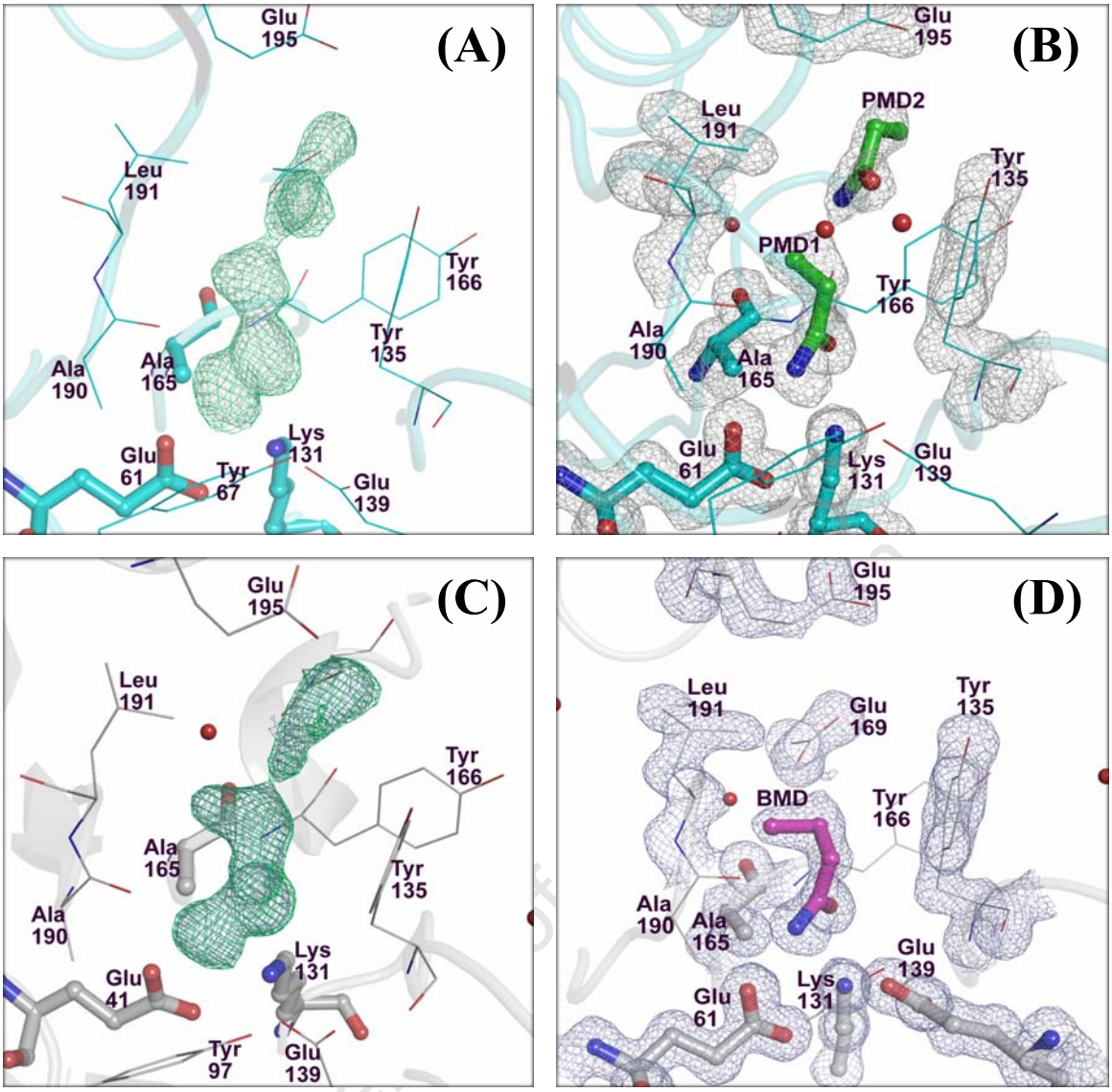


Fig. 3.5: Electron density for PMD and BMD in the active site of C165A NitN mutant. **(A)** σA-weighted $2F_{obs}$ - F_{calc} (grey mesh, contoured at 1σ) and F_{obs} - F_{calc} positive difference (green mesh, contoured at 3σ) maps clearly showing the density for omitted PMD molecules in the active site. **(B)** σA-weighted $2F_{obs}$ - F_{calc} map (grey mesh, at 1σ) for the active site environment of the refined C165A-PMD co-crystal structure. PMD molecules are rendered as green CPK balls and sticks. **(C)** σA-weighted $2F_{obs}$ - F_{calc} (purple mesh, at 1σ) and F_{obs} - F_{calc} positive difference (green mesh, at 3σ) maps of the omitted BMD molecules. The density for the second BMD molecule is incomplete. **(D)** σA-weighted $2F_{obs}$ - F_{calc} map (light purple mesh, at 1σ) for the refined C165A-BMD co-crystal structure. BMD molecule is shown as magenta CPK balls and sticks and the residues surrounding the active site are labeled. The second BMD molecule was omitted from the final refinement cycle as some atoms were disordered (missing), and others had high temperature factors.

The binding of the first propionamide (PMD1) and butyramide (BMD) displaces three water molecules (WatA, WatB and WatC) that are present in the unliganded C165A structure (Fig. 3.6); the carbonyl oxygen of the amide molecules displaces WatB that is located in the two-pronged (amino group of Lys131 and the peptidic NH group of Tyr166) oxyanion hole, the amide amino group displaces WatA, while WatC is replaced by the atoms of the side chain

moieties of the amide substrates. Although the bound and the unbound structures are in virtually identical conformations, small side chain shifts can be observed in some of the active site residues as seen in figure 3.6 below.

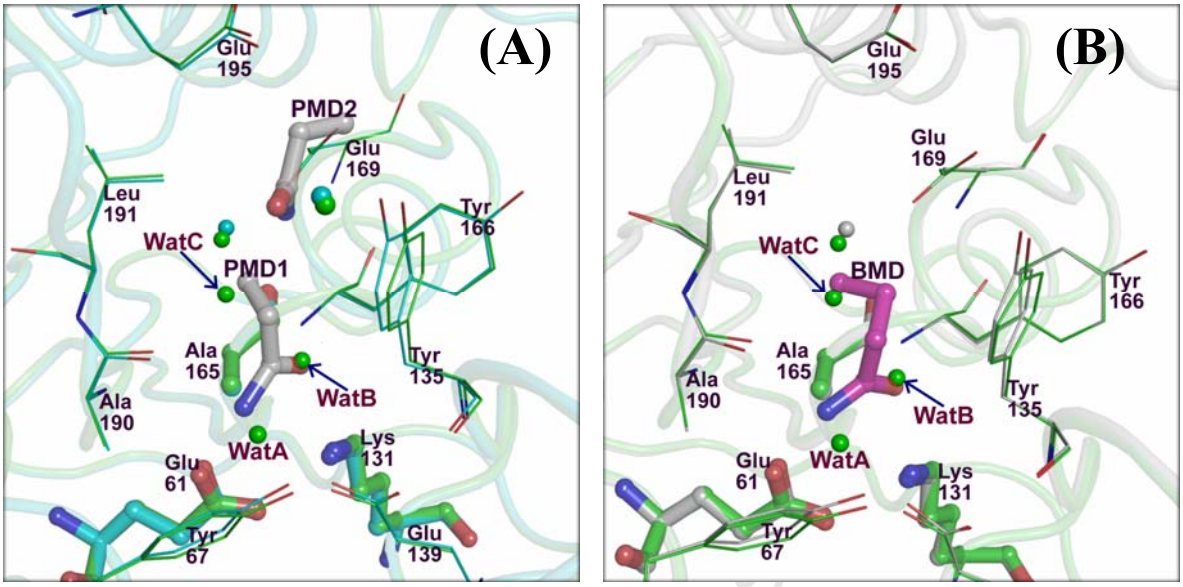


Fig 3.6: Comparison of the PMD- and BMD-complexed C165A mutant structures with the unliganded C165A NitN structure. **(A)** A superposition of the C165A-Apo (green CPK) and C165A-PMD (cyan CPK) active site pockets. The two PMD substrate molecules are shown in grey CPK ball and stick representation and water molecules are shown as green and cyan non-bonded spheres for the C165A-Apo and C165-PMD structures respectively. WatA, WatB and WatC in the unbound structure are displaced upon the substrate binding. **(B)** Overlaid binding pockets of the C165A-Apo (green CPK) and C165A-BMD complex (grey CPK) structures. BMD is shown as magenta CPK ball and stick model while the water molecules belonging to the unliganded C165A structure are shown as green non-bonded spheres. BMD also displaces WatA, WatB and WatC upon binding. Small side chains shifts can be seen between complexed and unbound structures.

3.4.6.2 Interactions of PMD and BMD with the active site atoms

(A) The first binding sub-site (catalytic sub-site)

BMD and the first PMD molecule (PMD1) bind in a similar manner at the bottom of the active site pocket where they are positioned through hydrogen bonding with the key active site residues (Fig. 3.7). The amide carbonyl oxygen is located in the oxyanion hole, interacting with the backbone NH group of Tyr166 and the side chain amino group of Lys131 in both cases. The amide amino group on the other hand is hydrogen bonded to the backbone carboxyl oxygen of Ala190, the OE_2 atom of Glu61 and OE_1 of Glu139 (Table 3.4 and Fig. 3.7). The substrate side chain moieties are stabilized through hydrophobic interactions involving Tyr135 and Leu191 as shown in table 3.4. The observed binding similarity, even though the PMD and BMD side chain

moieties are of different lengths, is an indication that aliphatic amidases only recognize the amide moiety (carbonyl moiety and the amide amino group specifically); the rest of the substrate docks into the space of the binding pocket that might differ in size and amino acid composition depending on the nature and size of the substrate. While the carbonyl oxygen of the amide substrates has a high affinity for the oxyanion hole, the conserved active site glutamic acid residues (Glu61 and Glu139) play an important role in positioning the amide amino group, thus ensuring optimal positioning of the substrates for nucleophilic attack by the reactive sulfhydryl group of the catalytic cysteine. The short aliphatic amides binding site at the bottom of the NitN pocket is referred to as the catalytic sub-site in this study, since it is the site where catalysis takes place.

Table 3.4 Polar contacts and hydrophobic interactions between PMD1 and BMD and the active site residues of the C165A NitN mutant.

Nature of interaction	**Substrate atom**		**Interacting residue, atom**	**Distances (Å)[a]**	
	PMD1	**BMD**		**PMD1**	**BMD**
Hydrogen bonding	N	N	Ala190, O	3.14	3.11
			Glu61, OE_2	2.66	2.66
			Glu139, OE_1	3.32	(3.46)
	O	O	Tyr166, N	2.97	3.07
			Lys131, NZ	2.95	2.90
Hydrophobic contacts	Cβ	C1	Leu191, CD_1	3.83	3.73
	C1	C1 and Cα	Tyr135, CB	3.86	3.90 and 3.76
	Cα	Cα	Tyr135, CG	3.79	3.64
	-	Cα	Tyr135, CD_2	-	3.8

[a]Value in parentheses represent distances that is greater than the maximum hydrogen bond length of 3.4 Å.

(B) The second binding sub-site

The second PMD molecule (PMD2) binds to the Glu169 and Tyr135 side chains at a site that is close to the mouth of the pocket. The amide amino group hydrogen bonds to the OE_2 atom of Glu169, while the carbonyl oxygen interacts with the OH atom of Tyr135 as shown in figure 3.7. The side chain moiety of PMD2 is stabilized by hydrophobic interactions involving Tyr135 and Glu169 atoms as can be seen in table 3.5.

PMD2 has an average B-factor of 35.18 $Å^2$, compared to 23.88 $Å^2$ for PMD1 that is bound deep in the pocket. PMD2 is positioned by fewer (two versus five in PMD1) hydrogen bonds, which

might explain the instability indicated by a higher B-factor value. The occupation of the second binding sub-site is likely to be temporary and a substrate bound to this site in the WT NitN enzyme is likely to descend to the catalytic sub-site when it becomes free during hydrolysis. The binding of two PMD molecules is a clear indication of how large and deep the NitN binding pocket is. Although the second substrate molecule may greatly contribute to the stability of the aliphatic amide molecule bound to the catalytic sub-site hence enhancing the efficiency of the reaction, its presence serves to confirm the speculations by Nel and Colleagues (2011) that short aliphatic amides may not be closely related to the unknown natural substrates of NitN.

Table 3.5 Polar contacts and hydrophobic interactions between PMD2 and the active site residues of C165A NitN mutant.

Nature of interaction	**Substrate atom**	**Interacting residue, atom**	**Distances (Å)**
	PMD2		**PMD2**
Hydrogen bonding	N	Glu169, OE2	2.69
	O	Tyr135, OH	3.37
Hydrophobic contacts	C1	Tyr135, CZ and CE	3.76 and 3.53
	Cβ	Leu191, CD1	3.49
	Cβ	Tyr135, CE1	3.89
	Cβ	Glu169, CD	3.48

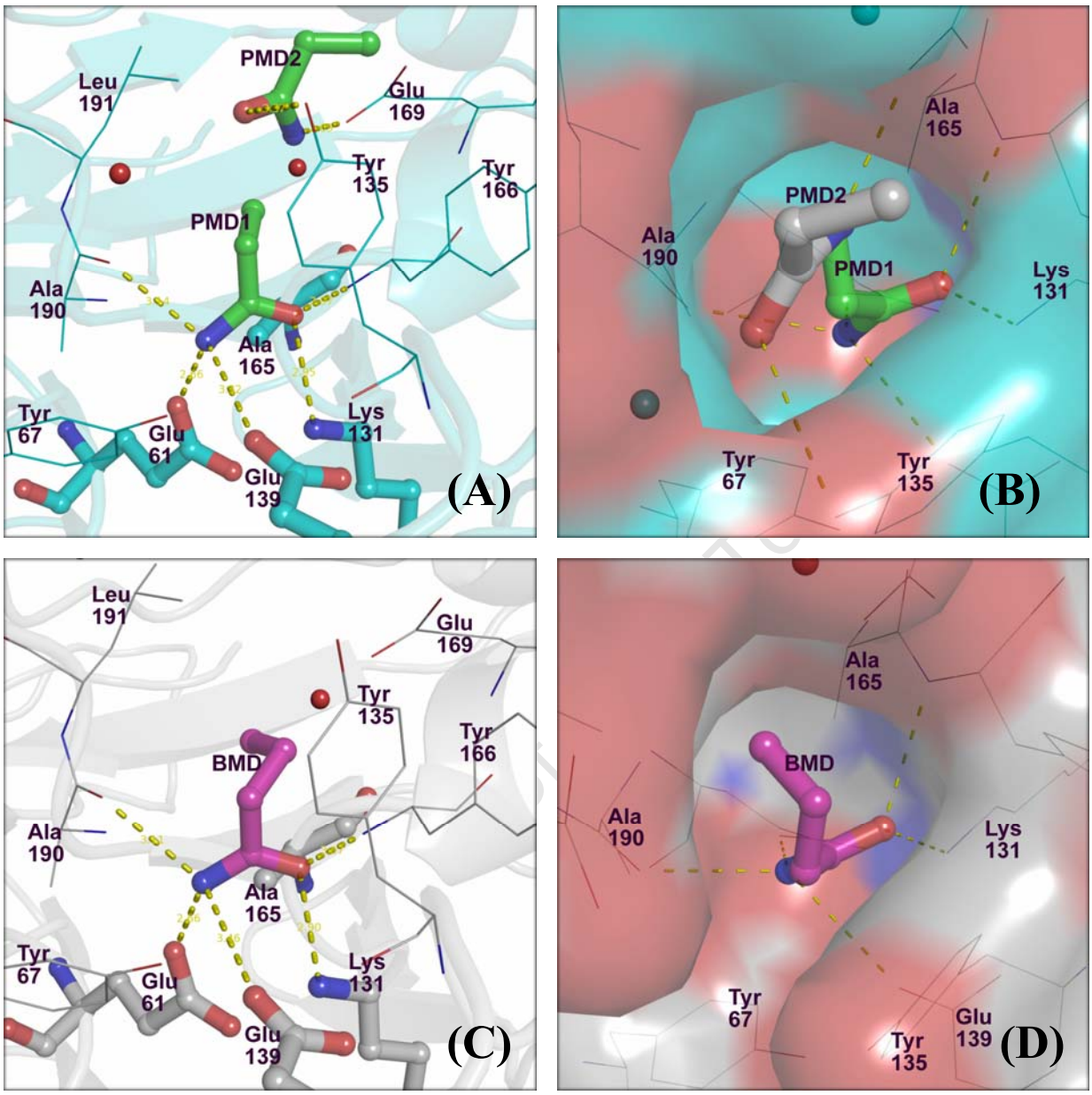


Fig. 3.7: Binding of PMD and BMD in the active site of C165A NitN amidase. **(A)** Line and ball and stick representation of the binding and positioning of PMD molecules in the C165A mutant active site. The two PMD molecules are colored in green CPK and the yellow dashes show potential hydrogen bonds between the substrates and the active site residues. **(B)** A transparent surface showing the two PMD substrate molecules in the pocket. PMD1 at the bottom of the pocket (catalytic sub-site) is shown in green CPK, while PMD2 (grey CPK) is bound close to the mouth of the pocket. **(C)** Positioning and coordination of BMD (magenta CPK) in the binding pocket of C165A mutant, where yellow dashes represent potential hydrogen bonds to the protein atoms. BMD binds in a similar manner to PMD1 as evidenced by the hydrogen bonding network. **(D)** A transparent surface showing the bound BMD (magenta CPK) molecule deep in the binding pocket of the C165A mutant. The pocket appears more open in the region of Tyr67 as compared to the PMD co-crystal structure active site.

3.4.6.3 Possible determinants of preference for PMD over BMD by NitN

NitN WT enzyme has 20% activity for BMD relative to propionamide (100%), which is the best-known substrate *in vitro* (Nel et al. 2011). A comparison of the C165A-PMD and C165A-BMD co-crystal structures does not immediately reveal why PMD is a better substrate than BMD. The two substrates make the same polar contacts with the protein and have similar orientations of the side chain moiety (Fig. 3.8) even though BMD has one extra carbon atom.

Although the protein component of the PMD- and BMD-C165A co-crystal structures is essentially in an identical conformation to the substrate-free C165A monomeric structure, small atomic variations can be observed that might partly explain why PMD is a better substrate. The root mean square deviation (rmsd) values of various atom categories (all-atoms, α-carbons, backbone and side chains) for both the superimposed monomeric structures and the overlaid active site regions are slightly higher for the BMD co-crystal structure than for the PMD-complexed structure, as compared to the unbound C165A structure (Table 3.6). This could be an indication that the binding and hydrolysis of BMD by NitN, as well as the release of the acid product may require some small (extra) conformational changes, which might result in reduced catalytic rates. The observed conformational variations are however not significant, and may only have small contribution (if any) towards the reduced NitN activity with BMD.

A catalytic factor that may contribute significantly towards the preference of PMD over BMD by NitN could be the rate of the release of the acid product. BMD, which has one extra carbon in the side chain moiety may have its corresponding acid (butyric acid) retained in the active site for longer (delayed release) than the propionic acid from PMD, due to the extra hydrophobic interactions that the side chain moiety makes with the protein atoms (Table 3.4). Delayed release of butyric acid would reduce the overall catalytic rates of NitN with BMD, partly explaining why PMD is a better substrate.

The two substrate binding sub-sites in the NitN pocket may have some effect on the catalytic efficiency of NitN on short aliphatic amide substrates. Both sub-sites are stably occupied in the PMD co-structure, while the second sub-site is only partially occupied in the C165A-BMD co-crystal structure as indicated by high B-factors of most BMD2 atoms and lack of ordered density for some of the atoms. While the NitN binding pocket is large and unconstrained (Nel et al. 2011), a single short aliphatic amide molecule at the catalytic sub-site is likely to be relatively

unstable (or unsteady), which would greatly affect the catalytic efficiency of the enzyme. The binding of a second PMD molecule at the second sub-site is likely to confer order on the entire active site pocket and to contribute greatly to the stability of the PMD molecule at the catalytic sub-site, which would enhance the nucleophilic attack by the catalytic cysteine. This phenomenon may partly provide a satisfactory explanation as to why PMD is a more preferred substrate by NitN.

Table 3.6 Rmsd values (Å) for the overall monomeric structures and the active site region between the unliganded C165A structure and co-crystal structure of C165A with PMD (C165A-PMD) and with BMD (C165A-BMD).

Structures aligned	RMSD (Å)							
	Overall structure				Active site[a]			
	All atoms	Cα	Back-bone	Side-chains	All atoms	Cα	Back-bone	Side-chains
C165A-Apo *vs.* C165A-PMD	0.46	0.09	0.10	0.67	0.16	0.09	0.12	0.18
C165A-Apo *vs.* C165A-BMD	0.47	0.16	0.15	0.67	0.18	0.11	0.15	0.19
C165A-PMD *vs.* C165A-BMD	0.37	0.15	0.14	0.51	0.10	0.06	0.07	0.11

[a]Residues within 4Å of the substrates in the active site (Glu61, Tyr67, Lys131, Tyr135, Glu139, Ala165, Tyr166, Glu169, Ala190, Leu191 and Glu195).

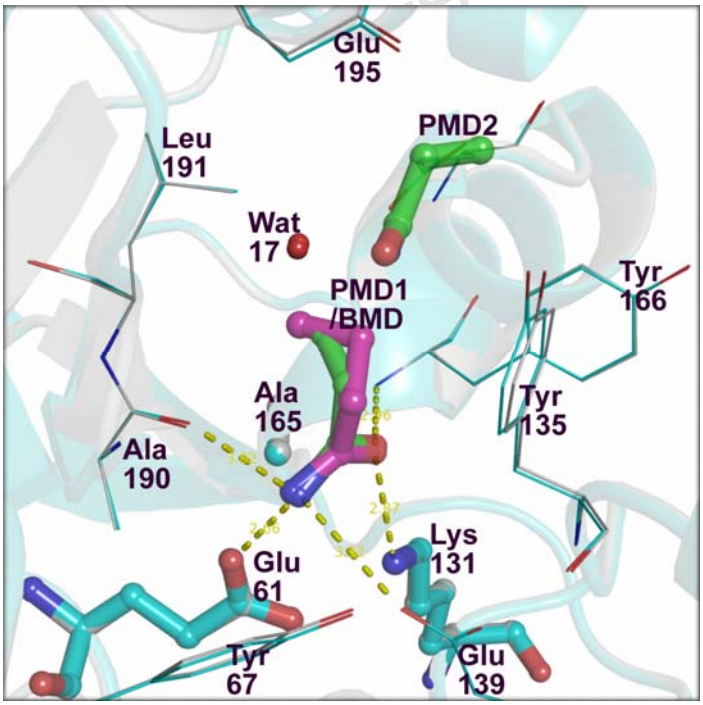


Fig. 3.8: A superposition of the C165A-PMD (cyan CPK) and C165A-BMD (grey CPK) co-crystal structures showing similarities in binding, positioning and orientation of PMD (green CPK) and BMD (magenta CPK) in the active site of the C165A NitN amidase. Potential hydrogen bonds are shown as yellow dashes. PMD1 and BMD substrates make the same polar contacts (yellow dashes) to the protein atoms and the side chain moieties adopt a similar orientation. It is therefore not possible to deduce substrate specificity based on the binding mode. Wat17 that is shown as non-bonded red spheres is conserved in the two co-crystal structures.

3.4.6.4 Comparison of the PMD- and BMD-complexed C165A NitN structures with other substrate-complexed amidase structures

The C166S mutant of the branch 2 formamidase from *Helicobacter pylori* has been co-crystallized with formamide (C166S-FMD; PDB id, 2e2l) (Hung et al. 2007), and a number of crystal structures for the C171A/V236A mutant of the *Agrobacterium* sp. N-carbamyl-D-amino acid amidohydrolase (DCase), in complex with a number of D-amino acid carbamyl substrates (PDB ids, 1uf5, 1uf7 and 1uf8) are available (Hashimoto et al. 2003).

A superposition of the C165A-PMD NitN co-crystal structure with that of the formamide-complexed C166S *H. pylori* formamidase revealed the expected similarities in the substrate binding mode (Fig. 3.9). Similar to PMD, the amino group of formamide (FMD) mainly interacts with the two conserved glutamic acid residues (Glu60 and Glu141 in the formamidase), while the carbonyl oxygen hydrogen bonds to the catalytic lysine (Lys133) and the backbone nitrogen of His167, which replaces Tyr166 in NitN. FMD is however in a different orientation compared to PMD, and this is caused by its additional interactions with the side chain hydroxyl group of Ser166 (Fig. 3.9 (A)), which is a replacement of the catalytic cysteine (Cys166) in the C166S formamidase mutant. The catalytic cysteine would not make such contacts to the substrate, and therefore the orientation of FMD in the *H. pylori* formamidase structure is probably not a true representation of optimal substrate positioning for nucleophilic attack.

The active site pocket gating residues, Tyr192 and Trp137, in the *H. pylori* formamidase (Hung et al. 2007) correspond to Ala190 and Tyr135 respectively in the NitN amidase. It is clear from figure 3.9 (B), how the configuration and orientation of the side chains from these gating residues greatly reduces the size of the binding pocket in the formamidase, allowing only formamide to bind.

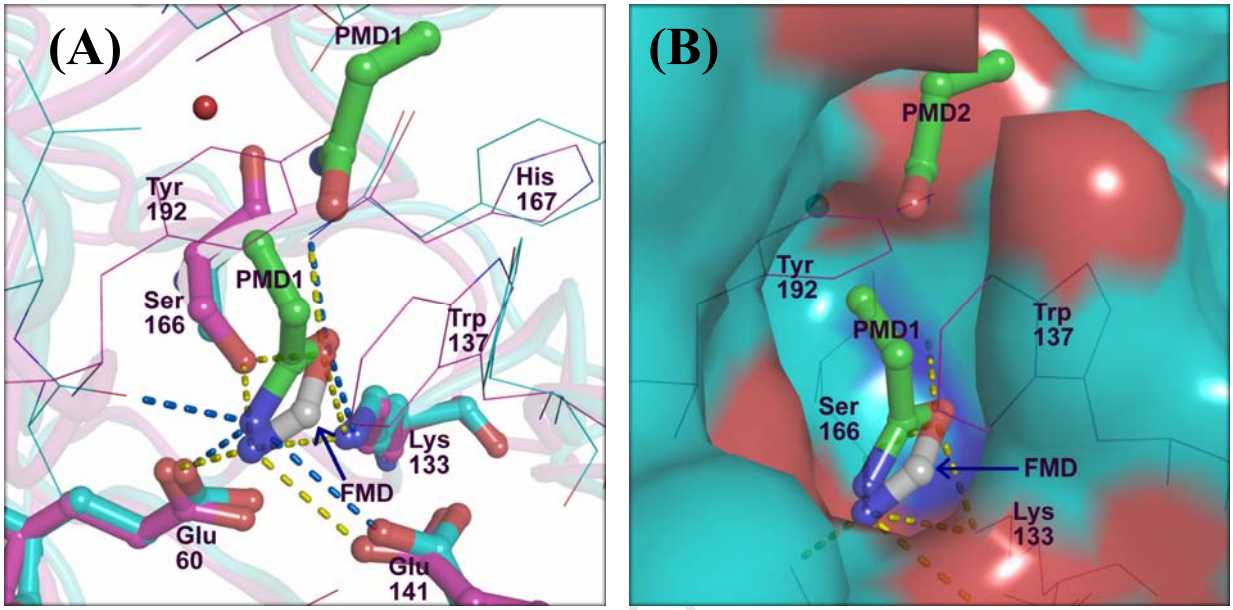


Fig. 3.9: Fig. 3.9: Binding modes of PMD and FMD in the amidase binding pockets. **(A)** An overlay of the active site pockets in the PMD-complexed C165A-PMD NitN structure (cyan CPK) and the FMD-complexed *H. pylori* C166S formamidase structure (magenta CPK; PDB id, 2e2l). The pocket residues are labeled following the formamidase numbering, the potential contacts between FMD and the formamidase active site residues are shown as yellow dashes, while interactions between PMD and the NitN pocket atoms are shown as blue dashes. FMD and PMD molecules are rendered as grey CPK and green CPK ball and stick models respectively. The orientation of FMD is biased by the hydrogen bonding interactions between its amino group and the carbonyl oxygen atom, and the OH of Ser166. These interactions would not occur with the catalytic cysteine in position 166. **(B)** A transparent surface representation of the NitN C165A-PMD structure (cyan CPK) and a sticks model of the overlaid FMD-complexed C166S formamidase co-crystal structure (magenta CPK), with PMD and FMD being rendered as green CPK and grey CPK balls and sticks respectively. The pocket residues are labeled following the formamidase numbering. The binding pocket gating residues in the formamidase, Tyr192 and Trp137, can be seen protruding into the NitN pocket, giving an indication of the pocket size-constraints in the formamidase and explaining its strict specificity for FMD only.

A comparison of the co-crystal structures of NitN C165A-BMD and the C171A/V236A mutant of DCase in complex with N-carbamyl-D-methionine (Dmet; PDB id, 1uf5) revealed interesting similarities and key differences in the binding of the two substrates to their respective active sites (Fig. 3.10). In the superimposed structures, the carbonyl group in the carbamyl moiety of the Dmet substrate aligns with the carbonyl entity of BMD. The rest of the Dmet substrate then extends along the side chain portion of BMD, bringing one of the oxygen atoms from its carboxyl moiety in close proximity to the Cγ atom of BMD (Fig. 3.10 (A)). The methionine side chain moiety of Dmet docks into an area of the pocket that is large enough to accommodate different D-amino acid side chains, explaining the broad substrate specificity that is observed in the bacterial DCases (Ogawa et al. 1994).

The carbonyl group in the carbamyl moiety of Dmet interacts with the protein in a similar manner as the carbonyl moiety of BMD (Fig. 3.10 (A)). In particular, the carbonyl oxygen atom of Dmet interacts with the amino group of the catalytic lysine (Lys126 in DCase) and the backbone NH group of Asn172, which replaces Tyr166 in NitN. On the other hand, the amino group that is attached to the carbonyl group in Dmet interacts with the two conserved glutamate residues (Glu46 and Glu145) as well as the backbone carboxyl oxygen of Asn196, which replaces Ala190 in NitN amidase. The side chain of Glu46 in DCase is in a slightly different conformation, allowing both of its carboxyl oxygen atoms to interact with the amino group of the Dmet substrate, hence contributing an extra hydrogen bond (Fig. 3.10 (A)).

Most of the binding pocket space that is occupied by the Dmet side chain moiety in DCase is not available in the NitN pocket as evidenced by an atomic clash between the side chain of Tyr135 in NitN and the Dmet side chain (Fig. 3.10 (A and B)). This provides a satisfactory explanation as to why NitN has no enzymatic activity on the carbamyl substrates that were tested *in vitro* (Nel et al. 2011). Tyr135 in NitN is replaced by Pro130 in DCase, which provides the required space for binding of carbamyl substrates with different sizes of side chain moiety.

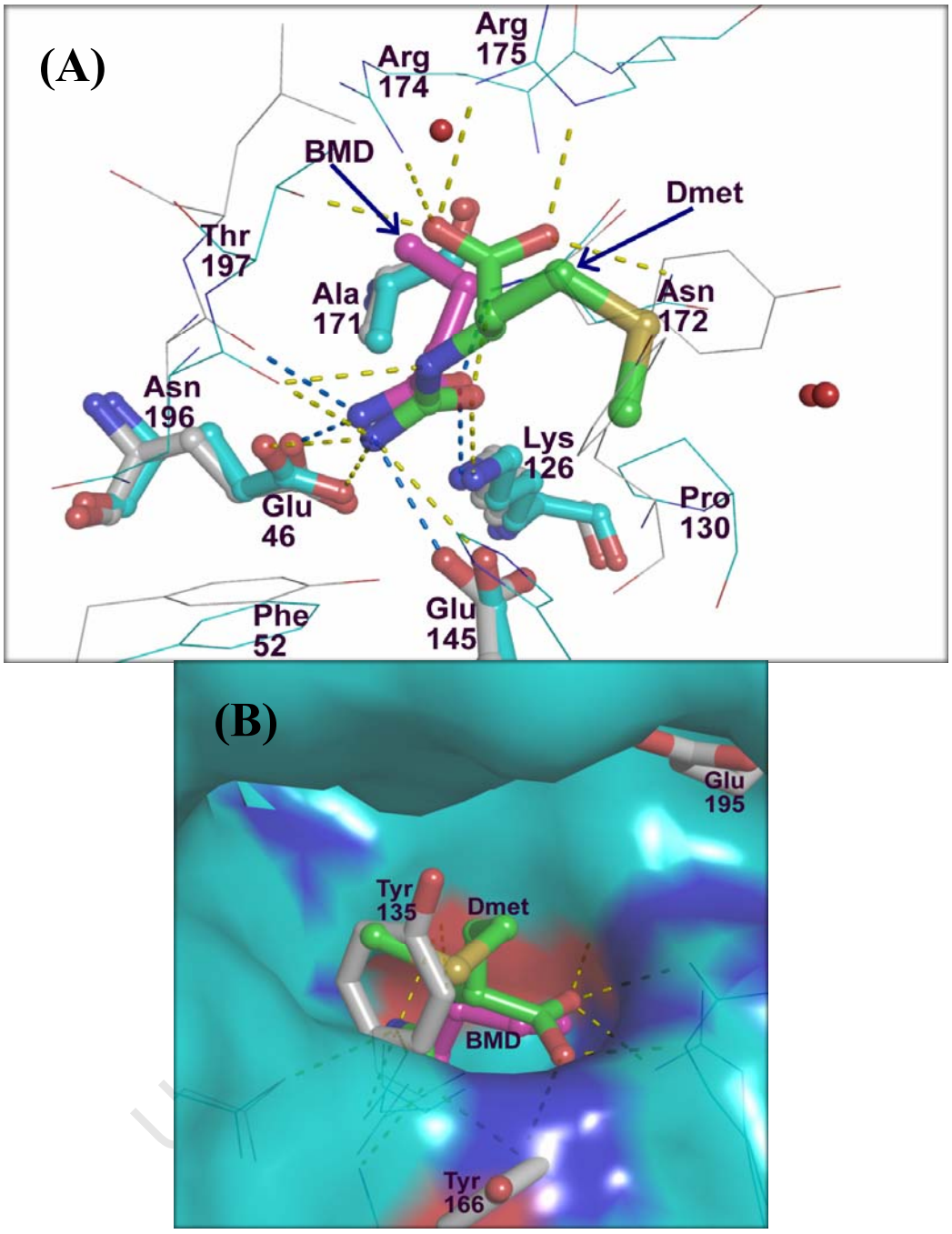


Fig. 3.10: A comparison of the binding modes between Dmet (in DCase) and BMD (in NitN). **(A)** A line and ball and stick representation of aligned binding sites of the Dmet-complexed DCase (cyan CPK; PDB id, 1uf5) and the C165A-BMD NitN structure (grey CPK). BMD is shown in magenta CPK while Dmet is shown in green CPK colors. Protein residues are labeled following the numbering of the 1uf5 coordinates file (PDB id, 1uf5). The yellow dashes represent potential hydrogen bonds between Dmet and the protein atoms, while the blue dashes are polar contacts between BMD and the NitN pocket atoms. **(B)** A surface representation of the Dmet-bound DCase pocket (cyan CPK), with some active site residues from the aligned C165A-BMD NitN pocket (grey CPK) being shown as sticks. Dmet is shown as green CPK while BMD is rendered as magenta CPK balls and sticks. The Dmet side chain moiety clashes with the side chain of Tyr135 in NitN, indicating a limited space in the NitN pocket and explaining the lack of activity on carbamyl substrates by NitN.

3.4.7 The fluoroacetamide-complexed C165A mutant structure

The C165A NitN mutant reacted with fluoroacetamide (FAE) crystallized in a number of conditions, but only crystals grown in buffer 38P (Table 3.1) had FAE bound to the active site. Unlike the WT NitN and all the other co-crystal structures, which have $C222_1$ symmetry with one subunit per asymmetric unit (ASU), the C165A-FAE structure is in $P2_12_12_1$ spacegroup with six monomeric subunits in the ASU. The six subunits are organized into three dimers that are arranged in a helical manner as shown in figure 3.11.

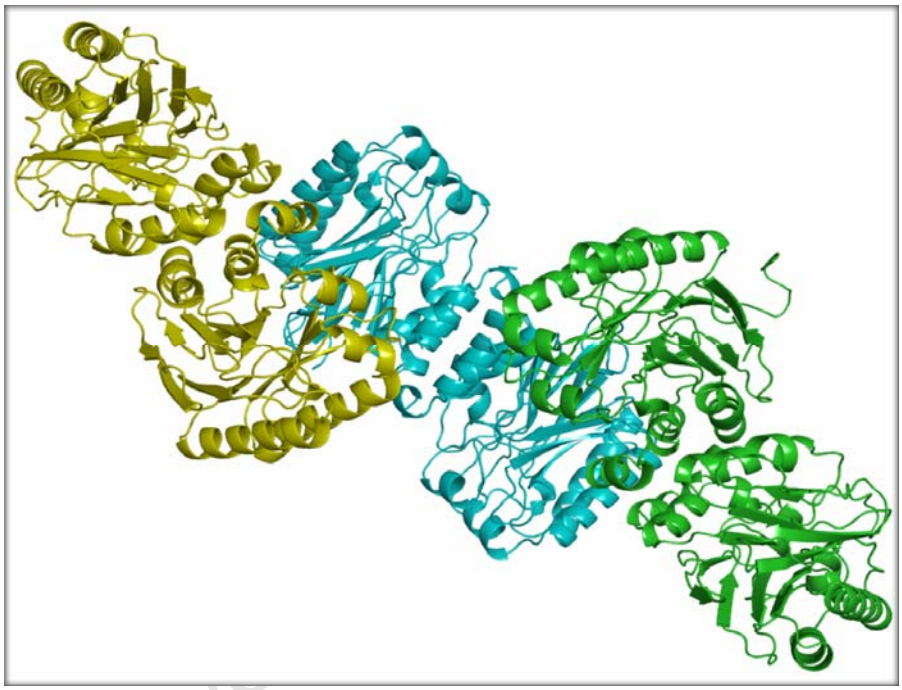


Fig. 3.11: Arrangement of the six C165A-FAE subunits in the asymmetric unit of $P2_12_12_1$ unit cell. The molecules are arranged into three dimers that are shown as green, cyan and yellow cartoons and are viewed down the two fold-axis of the middle (cyan) dimer.

Binding of FAE in the active site

Electron density that resembled a FAE molecule was observed close to the mouth of the pocket (Fig. 3.12 (A)), that is, at the sub-site 2 of the C165A-FAE active site pocket in three out of six subunits (chains A, D and F) in the ASU. A blob of irregular density was also observed half-way along the pocket, close to Ala190, Leu191 and Ala192 (Fig. 3.12 (A)). Although a FAE could roughly fit in the blobby density at different orientations in each of the three protein chains, building and refinement of FAE in this unusual sub-site was not attempted since similar density at exactly the same location was observed in the active sites of most unbound NitN mutant structures. It was assumed that this density belonged to a disordered buffer component or other substances in the crystallization mixture. The orientation and position of the carbonyl moiety of

the bound FAE molecules was approximately conserved, but the side chain moieties adopted different conformations in the three different chains (Fig. 3.2 (B)). It is not clear why this is the case, but there is a possibility of the density being slightly inaccurate and the orientation of the FAE molecules not being fully resolvable at 2.3 Å resolution of the C165A-FAE data.

There was no electron density for FAE molecules at the catalytic sub-site in all the six chains in the ASU; that is, at the site close to the catalytic residues, where PMD1 in the C165A-PMD complex structure is located (Fig. 3.12 (C)). It is not clear why FAE molecules were unable to bind at the catalytic sub-site, but it was speculated that the binding of the unknown molecule belonging to the irregular density close to Ala190, Leu191 and Ala192 (Fig. 3.12 (A)) could have blocked the active site pocket, preventing access to the catalytic sub-site by FAE molecules. The FAE molecules bind in the same location as the PMD2 molecule in the C165A-PMD co-crystal structure. Similar to PMD2, the FAE molecules interact with the carboxyl oxygen of Glu169 and the hydroxyl group of Tyr135 in all the three chains (Fig. 3.12 (D)).

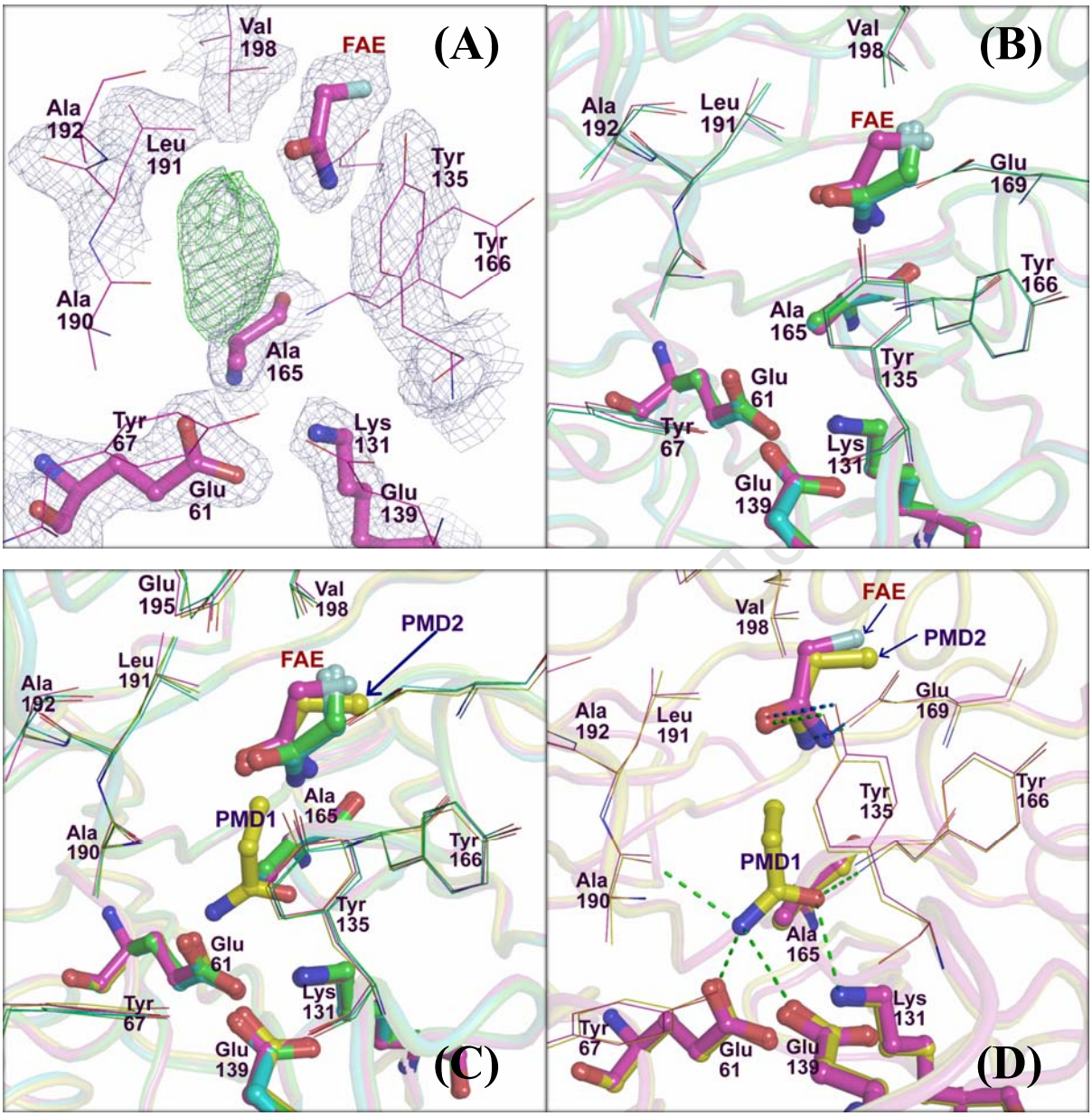


Fig. 3.12: The binding of FAE to the active site pocket of the C165A NitN mutant. **(A)** The σA-weighted $2F_{obs}$ - F_{calc} map (contoured at 1σ; light purple mesh) showing the refined density for the active site environment including a bound FAE molecule, and the positive difference $2F_{obs}$ - F_{calc} map (contoured at 3σ; green mesh) showing the irregular blobby density close to Ala190, Leu191 and Ala191 in the chain F of the C165A-FAE co-crystal structure. **(B)** A superposition of the active sites in the 3 FAE-complexed chains in the ASU; A (green CPK), D (cyan CPK) and F (magenta CPK), showing the position and orientation of the FAE molecule in each chain. The FAE molecules are colored according to the chain coloring. **(C)** A comparison of the binding pockets in the FAE-complexed chains (colored as described in panel B) and the C165A-PMD co-crystal structure (yellow CPK). The FAE molecules did not bind at the catalytic sub-site of the active site pocket, where PMD1 is located. **(D)** An alignment of FAE-complexed chain F (magenta CPK) and C165A-PMD binding pockets (yellow CPK) showing the similarity in the hydrogen bonding network between the protein and the two substrates, FAE and PMD. Green dashes show PMD interactions with the active site residues, while blue dashes show contacts involving FAE and the protein.

3.4.8 The substrate-complexed C165S mutant structures

A number of crystals were obtained from the C165S NitN mutant protein that had been reacted with various amide substrates prior to crystallization. Most of the resultant structures did not have density for substrate molecules in the active site, despite the fact that the same conditions were used for substrates binding in both C165A and C165S NitN mutants. In the few substrate-complexed C165S structures that were solved, none of them had the amide substrates bound to the catalytic sub-site at the bottom of the pocket; instead the substrates were bound in random sub-sites of the NitN pocket.

The mutation of Cys165 to a serine alters the active site configuration and environment of NitN (Fig. 3.4 (D)). The side chain of Ser165 in the C165S NitN mutant is oriented in such a way that, the hydroxyl group protrudes into the pocket, where it hydrogen bonds with the carboxyl oxygen atoms of Glu61, thereby preventing the binding of three water molecules that are present in the WT NitN active site pocket (Fig. 3.4 (D)). These changes may also be unfavorable for the binding of short aliphatic amide substrates; hence C165S NitN mutant is not a good model for substrate binding studies. In the *H. pylori* formamide-complexed C166S formamidase co-crystal structure (PDB id, 2e2l; (Hung et al. 2007), the hydroxyl group of Ser166 makes polar interactions with the formamide (FMD) amino group and the carbonyl oxygen atom (Fig. 3.13 (A)), which ultimately alters the orientation of FMD relative to the PMD1 in the C165A-PMD NitN co-crystal structure (Fig. 3.13 (B)). Since the side chain of Ser165 in the C165S NitN mutant adopts a similar orientation to the side chain of Ser166 in the *H. pylori* FMD-C166S formamidase co-crystal structure (Fig. 3.13 (A) and (B)), the hydroxyl group of Ser165 in NitN is also likely to interact with the amide moiety (carbonyl oxygen atom and the amide amino group) of the short aliphatic amide substrates, which would promote a change in their orientation within the active site pocket. If PMD, FAE, BMD and ACE were to adopt the same orientation as FMD in the *H. pylori* FMD-C166S formamidase binding pocket (Fig. 3.13 (A) and (B)), their side chain moieties would clash with the NitN pocket atoms, which explains why the short-chain amide substrates were unable to bind at the catalytic sub-site.

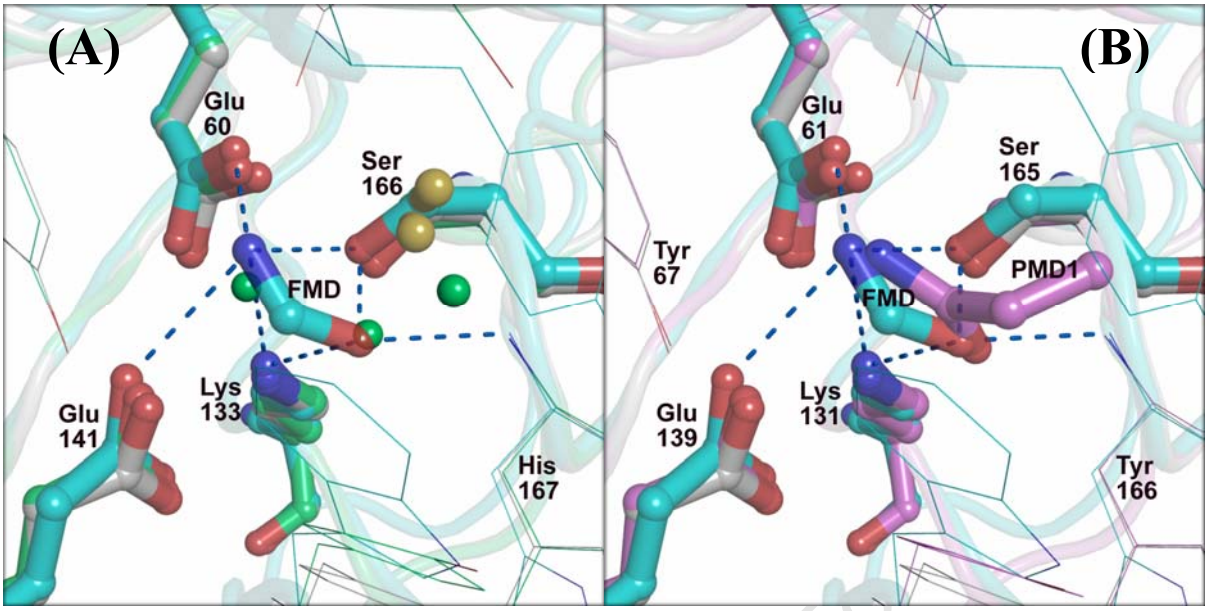


Fig. 3.13: A comparison of the WT, C165S-Apo and C165A-PMD NitN active site pockets with the binding pocket of FMD-complexed *H. pylori* C166S formamidase structure. **(A)** An overlay of the active site pocket regions of the WT NitN (green CPK), C165S-Apo NitN (grey CPK) and the FMD-complexed C166S formamidase from *H. pylori* (cyan CPK; PDB id, 2e2l). The residues are labeled following the numbering of the formamidase coordinates file (PDB id, 2e2l). Green non-bonded spheres are water molecules in the WT Nit binding pocket, while the blue dashes are polar contacts between FMD and the formamidase protein atoms. Ser166 in the formamidase adopts a similar orientation as Ser165 in the C165S-Apo NitN structure, which is different from the orientation of the catalytic cysteine in the WT NitN structure (sulfhydryl groups of Cys165 conformers are shown as yellow balls). **(B)** A superposition of the binding pockets of the FMD-complexed C166S formamidase structure (cyan CPK; PDB id, 2e2l), PMD-complexed C165A-PMD NitN structure (violet CPK) and the unbound (apo) C165S NitN structure (grey CPK). The pocket residues are labeled following the NitN sequence numbering (PDB id, 3hkx), and the blue dashes represent polar contacts between FMD and the formamidase active site atoms. The altered orientation of the FMD molecule due to its interactions with the OH group of the serine can be observed.

In some C165S co-crystal structures, the substrate molecules were bound in different other locations in the active site, while in other cases unidentifiable density, possibly of unknown metabolites from the protein expression host cells (*E. coli* BL21 (DE3)), was observed in the binding pocket. In the co-crystal structure of C165S NitN mutant with acetamide (C165S-ACE), electron density for a possible ACE molecule (Fig. 3.14 (A)) was observed between the positions of the two PMD molecules in the C165A-PMD NitN co-crystal structure (Fig. 3.14 (B)). Both the amide amino group and the carbonyl oxygen atom of the bound ACE molecule hydrogen bond to the carboxyl oxygen atoms of Glu169 (Fig. 3.14 (B)), while its side chain moiety is stabilized through hydrophobic interactions with Leu191 and Tyr135. No ACE molecule was bound at the catalytic sub-site.

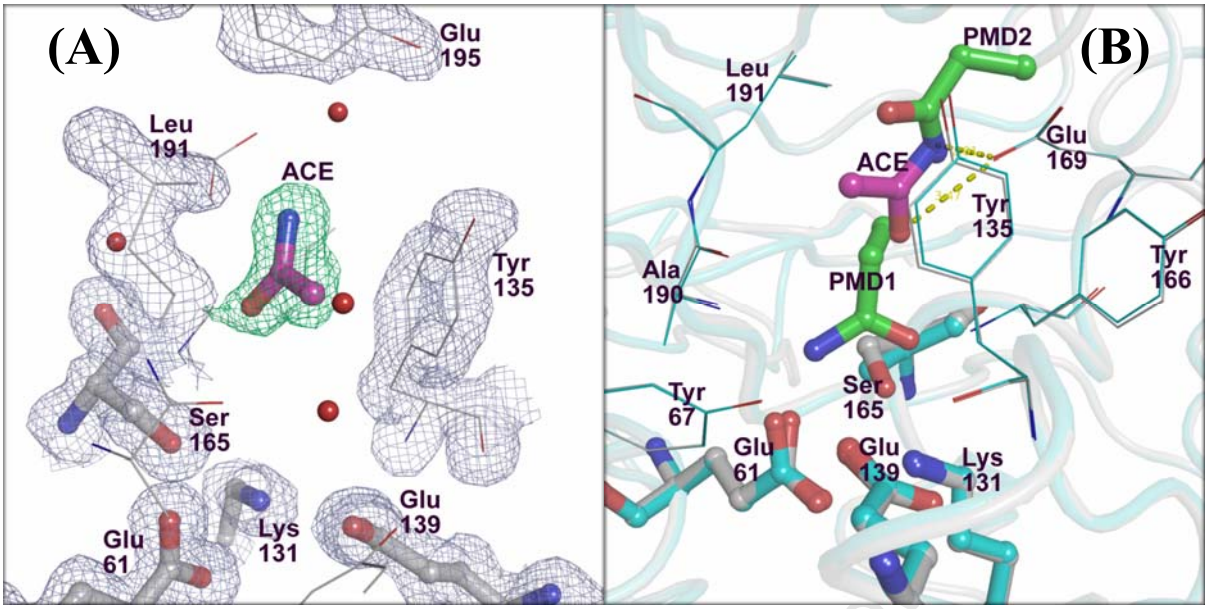


Fig. 3.14: Binding of ACE in the C165S NitN mutant. **(A)** σA-weighted $2F_{obs}$ - F_{calc} map (contoured at 1σ) showing the density for the ACE molecule (green mesh) and the surrounding residues (grey mesh) in the C165S-ACE co-crystal structure. ACE is shown as magenta CPK balls and sticks. **(B)** A superposition of ACE-complexed C165S NitN structure (grey CPK) and the PMD-complexed C165A-PMD NitN crystal structure (cyan CPK). ACE is shown in magenta CPK while the two PMD molecules are rendered as green CPK balls and sticks. Yellow dashes represent possible hydrogen bonds between ACE and the protein atoms. The ACE molecule is in a random position between the two PMD molecules.

In the PMD co-crystal structure of C165S NitN, density that resembled PMD (Fig. 3.15) was found in a unique sub-site of the binding pocket, where it was surrounded by Tyr67, Ala190, Leu191, Ala192 and Leu226 (within 4Å). The density for this PMD molecule was however weak with the atoms having an average B-factor of around 35 Å after refinement of atomic occupancies, an indication of flexibility or instability. In addition to the PMD density, some unidentified continuous density was observed starting from the bottom of the binding pocket around the catalytic residues, and extending towards the mouth of the pocket (Fig. 3.15). Attempts to fit two PMD molecules, HEPES from the crystallization buffer, or any other buffer contents that the protein had encountered during purification, resulted in large blobs of negative difference density in all cases, indicating that wrong molecules were being modeled. The density was speculated to be for an unknown metabolite that was probably bound to the C165S NitN active site during protein expression in *E. coli* cells. Similar density has been observed in another NitN mutant structure that is not reported in this chapter. The ability of NitN to bind larger molecules such as the one represented by this unknown density is supportive of earlier suggestions that its natural substrates are likely to be larger and different from the known short-

chain aliphatic amides; larger substrates are also consistent with the large size of the NitN active site pocket.

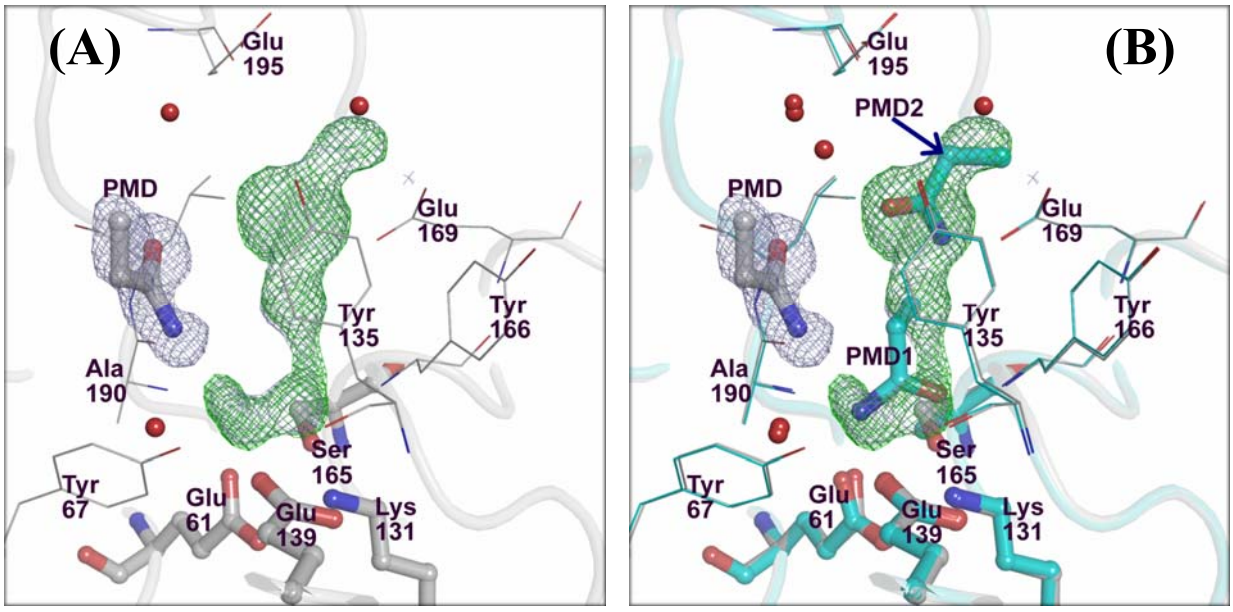


Fig. 3.15: PMD and unidentified density in the active site of C165S NitN that was crystallized in the presence of PMD. **(A)** The σA-weighted $2F_{obs}$ - F_{calc} electron density map (contoured at 0.7σ) for a refined PMD molecule (light purple mesh) in a unique site of the NitN binding pocket and the σA-weighted F_{obs} - F_{calc} positive difference map (contoured at 3σ) for the omitted possible metabolite from *E. coli* (green mesh). **(B)** A superposition of the active site regions of the C165S-PMD (grey CPK) and C165A-PMD (cyan CPK) co-crystal structures. The two PMD molecules from the C165A-PMD complex are rendered as cyan balls and sticks and they superimpose well with the unidentified density from the C165S-PMD maps. Although the PMD2 molecule appears to fit in one part of the strange density, fitting and refinement of one or two PMD molecules in this density resulted in large positive and negative difference densities, indicating that the density represents one continuous molecule, which remains unknown.

3.4.9 Co-crystal structures of the WT and C165A NitN with propionitrile

Co-crystallization experiments of the WT NitN and the C165A mutant with nitrile substrates were carried out in order to understand the basis for the lack of nitrile hydrolysis activity by NitN *in vitro*. The nitriles tested in these experiments were propionitrile (PNT), fluoroacetonitrile (FNT), butyronitrile (BNT) and acetonitrile (ANT). These were all selected based on the known amide substrates for NitN. In most cases, very irregular non-diffracting crystals formed in the presence of these nitriles although a few good crystals were also obtained and a number of datasets collected. The irregularities of these crystals were probably caused by the presence of methanol, which was used to dissolve and dilute out nitrile substrates to the desired concentrations for crystallization. Although most of the nitrile co-crystal structures did

not have any substrates in the active site, WT NitN and the C165A mutant proteins crystallized in the presence of PNT had density that resembled PNT in the binding pocket.

In the WT-PNT co-crystal structure, the catalytic cysteine was modified to a sulfenic acid, by the covalent binding of one oxygen atom to the Sγ atom. This modification is common in enzymes that utilize cysteine for catalysis; most of the available crystal structures of the nitrilase superfamily enzymes have the catalytic cysteine modified to either sulfenic or sulfinic acid. The density for one PNT molecule was found just above the side chain of the modified cysteine (Fig. 3.16 (A)), and analysis of the possible interactions revealed a single hydrogen bond between the nitrogen atom (N) of PNT and the oxygen of Cys165-sulfenic acid (Fig. 3.16 (B)). This interaction would not be expected if the catalytic cysteine was not modified, and therefore the positioning and orientation of PNT in the WT NitN structure is biased by the catalytic cysteine oxidation. Hydrophobic interactions involving the carbon atoms of PNT and Tyr135, Glu169 and Leu191 stabilize the bound PNT molecule within the NitN binding pocket.

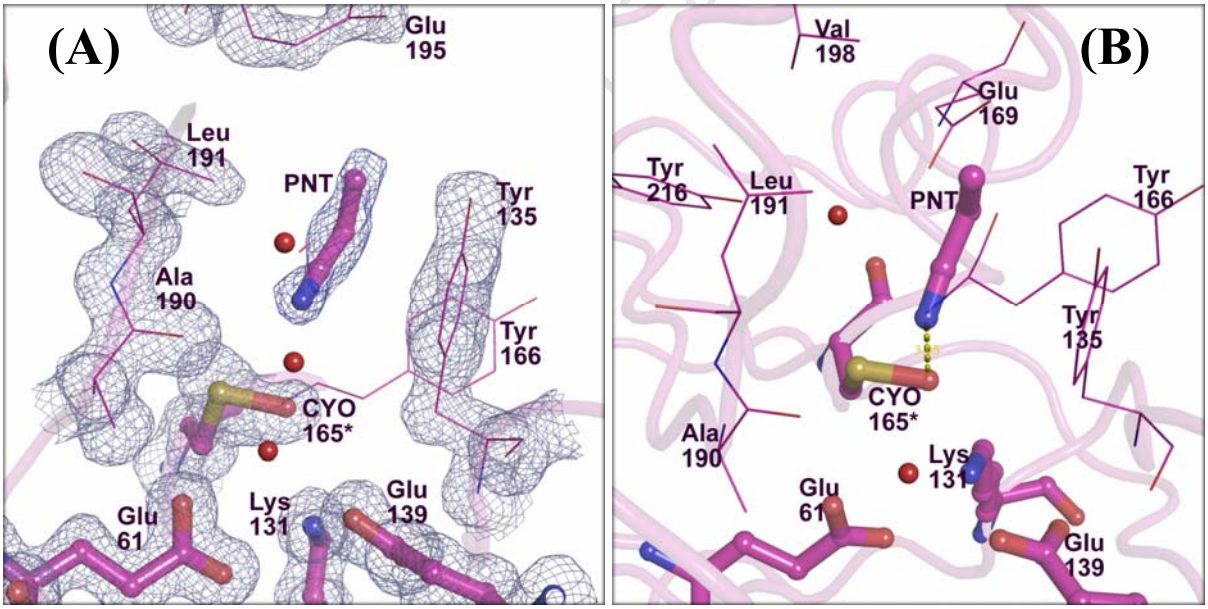


Fig. 3.16: PNT in the binding pocket of the WT NitN enzyme. **(A)** The σA-weighted $2F_{obs}$ - F_{calc} map (purple mesh) showing the density for the refined PNT and the surrounding residues in the active site; all colored in magenta CPK. The catalytic cysteine is covalently modified to a sulfenic acid (CYO165) and the density for the added oxygen atom is quite clear. **(B)** Coordination of PNT in the active site of the WT-PNT NitN structure. The only detected hydrogen bond between the nitrogen atom of PNT and the oxygen of the Cys165-sulfenic acid is shown as yellow dashes.

In the C165A-PNT co-crystal structure, density for two PNT molecules was present in the active site pocket (Fig. 3.17 (A)). The first PNT molecule (PNT1) is bound deep in the pocket at the catalytic sub-site, with its N atom being located in the oxyanion hole, where it interacts with the amino group of Lys131 side chain (2.99 Å) and the backbone NH group of Tyr166 (3.07 Å) (Fig. 3.17 (B)). This coordination suggests a component of general acid catalysis in the hydrolysis of nitriles by nitrilases, where the positively charged amino group of the catalytic lysine (Lys131) attracts electrons away from the C-N triple bond, ultimately making the reactive cyano carbon more electrophilic. PNT1 is further stabilized through numerous hydrophobic contacts between its carbon atoms and various protein atoms from Try135, Ala165, Tyr166, Ala190 and Leu191. The second PNT molecule (PNT2) is close to the mouth of the pocket (Fig. 3.17 (B)) and makes a hydrogen bond with the OE_1 atom of Glu169 (3.05 Å) as well as several hydrophobic interactions with Glu169, Tyr135 and Leu191. The two PNT molecules in the C165A structure differ in orientation and location from the PNT in the WT-PNT co-crystal structure (Fig. 3.17 (C)), whose position and orientation are biased by its interaction with the O atom of the Cys165-sulfenic acid.

Comparison of the binding mode of PNT1 and the butyramide (BMD) molecule in the C165A-BMD co-crystal structure (Fig. 3.17 (D)) provides satisfactory evidence for the observed lack of nitrile hydrolysis activity by NitN *in vitro*. The position and configuration of the electrophilic cyano carbon would not be optimal for nucleophilic attack by the catalytic cysteine. Nitrile-hydrolyzing enzymes are likely to have a unique active site geometry that is favorable for binding and positioning of linear nitrile molecules so that the cyano carbon is optimally aligned and in close proximity with the thiol group of the catalytic cysteine in order to allow nucleophilic attack.

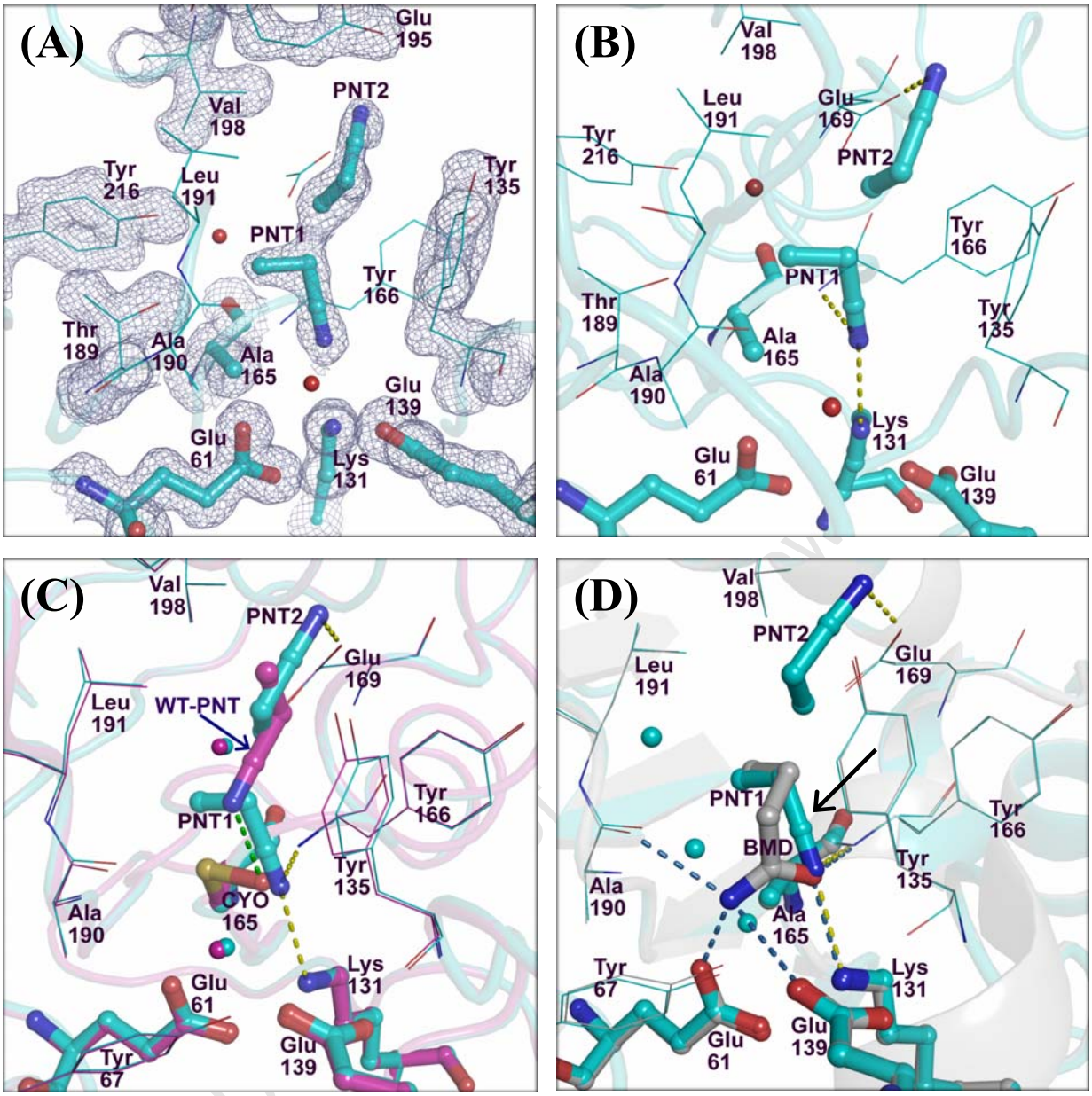


Fig. 3.17: Binding and coordination of PNT in the C165A-PNT co-crystal structure. **(A)** The σA-weighted $2F_{obs}$ - F_{calc} map (purple mesh) showing the density for the two PNT molecules and the surrounding residues in the active site; colored in cyan CPK. **(B)** Polar interactions between the PNT molecules and the active site residues are indicated by the yellow dashes. **(C)** A superposition of the PNT-complexed WT (magenta CPK) and C165A (cyan CPK) NitN structures. PNT in the WT structure (WT-PNT) is shown as magenta CPK balls and sticks and the hydrogen bond between its N atom and the O atom of Cys165-sulfenic acid (CYO165) is shown as green dashes. PNT molecules in the C165A structure are shown as cyan CPK balls and sticks and their potential interactions with the protein are shown as yellow dashes. The position and orientation of the PNT in the WT-PNT structure are biased by its interaction with the O of CYO165. **(D)** A comparison of the binding mode of BMD in the C165A-BMD NitN structure (grey CPK) with that of PNT1 in the C165A-PNT NitN structure (cyan CPK). The interactions between the proteins atoms and the BMD and PNT molecules are shown as blue and yellow dashes respectively. The cyano carbon of PNT1 (pointed at by a black arrow) is not positioned and aligned optimally for nucleophilic attack.

3.4.10 Identification of larger and better amide substrates for NitN

The crystal structure of the NitN catalytic lysine mutant (K131Q) that was crystallized in the presence of acetamide (ACE) had well defined electron density that was attached to the thiol group of the reactive cysteine (Cys165) and stretched all the way to the mouth of the binding pocket as rendered in figure 3.18 (A). While the extra density was clearly not for an ACE molecule, attempts to fit various possible molecules resulted in a tentative interpretation of the density as belonging to a possible adipamide (ADM) reaction intermediate that was covalently attached to the Sγ atom of Cys165. This interpretation appears to match the density (Fig. 3.18 (B) and (C)), although it is not confirmed whether the terminal group was a carboxylate (as in adipamic acid) or an amide (as in ADM). In the absence of mass spectrometric data, the composition of the adduct remains speculative. Also, it has not been possible to ascertain the configuration at the carbon that is attached to the catalytic cysteine; it is likely to be sp2 and thus the structure represents a thioester acyl-enzyme intermediate.

Since the mutant protein had not come into contact with ADM or related compounds during protein purification and crystallization, ADM was assumed to have been incorporated into the protein during expression, even though ADM or related products (adipamic acid and adipic acid) have not been documented as metabolites in *E. coli*. To confirm the interpretation of the extra electron density at the catalytic cysteine, ADM was purchased and the bromophenol blue ammonia assay carried out in order to determine if the WT NitN had enzymatic activity on ADM. The NitN reaction assay with PMD (the best known short aliphatic amide substrate) was used as a control; the amount of ammonia produced was compared to the ammonia generated from ADM hydrolysis under the same reaction conditions, and the results are plotted in figure 3.19 (A). NitN was indeed found to be more reactive with ADM than PMD, as judged from the amount of ammonia produced over a 60 minute reaction period. This supports the interpretation of the extra electron density in the K131Q NitN active site as a possible ADM or adipamic acid covalent intermediate. Although the kinetic parameters of NitN with ADM have not been determined, it is clear that ADM, a 6-carbon molecule (Fig. 3.19 (B)), is a better substrate than PMD, which is consistent with the size of the NitN binding pocket and confirms previous speculations that NitN is likely to have structurally different and larger substrates than the known short aliphatic amides.

The ability of ADM to occupy the entire length of the NitN binding pocket is likely to ensure its stability as well as that of the active site pocket, and this could be the reason for the increased activity and enhanced catalytic efficiency. In addition to the kinetic studies, co-crystallization of ADM with the C165A NitN mutant protein will be necessary in order to visualize non-covalently bound ADM molecules in the NitN pocket.

The ability of NitN to convert ADM to adipic acid efficiently is attractive to the biotechnological industry, as adipic acid is one of the raw materials in the manufacture of nylon-6:6.

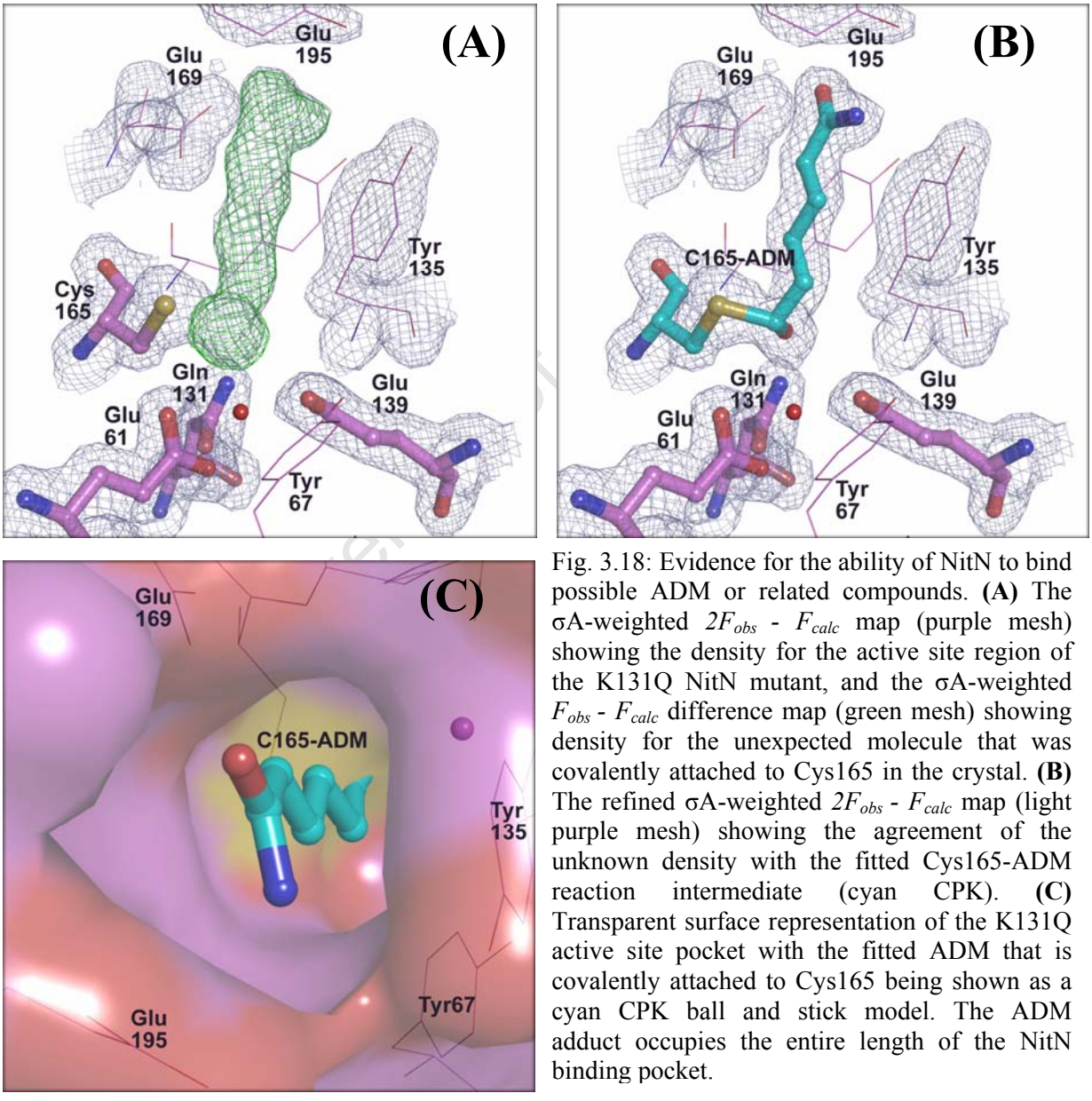
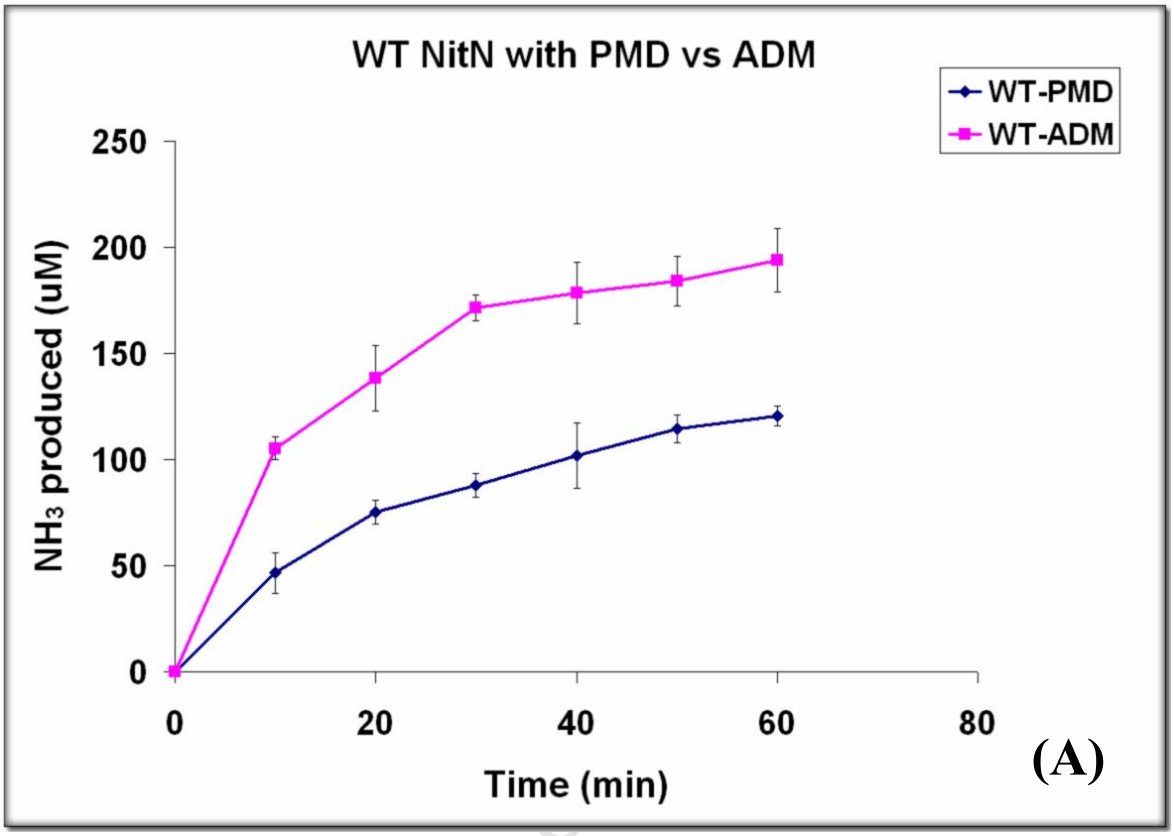


Fig. 3.18: Evidence for the ability of NitN to bind possible ADM or related compounds. **(A)** The σA-weighted $2F_{obs}$ - F_{calc} map (purple mesh) showing the density for the active site region of the K131Q NitN mutant, and the σA-weighted F_{obs} - F_{calc} difference map (green mesh) showing density for the unexpected molecule that was covalently attached to Cys165 in the crystal. **(B)** The refined σA-weighted $2F_{obs}$ - F_{calc} map (light purple mesh) showing the agreement of the unknown density with the fitted Cys165-ADM reaction intermediate (cyan CPK). **(C)** Transparent surface representation of the K131Q active site pocket with the fitted ADM that is covalently attached to Cys165 being shown as a cyan CPK ball and stick model. The ADM adduct occupies the entire length of the NitN binding pocket.



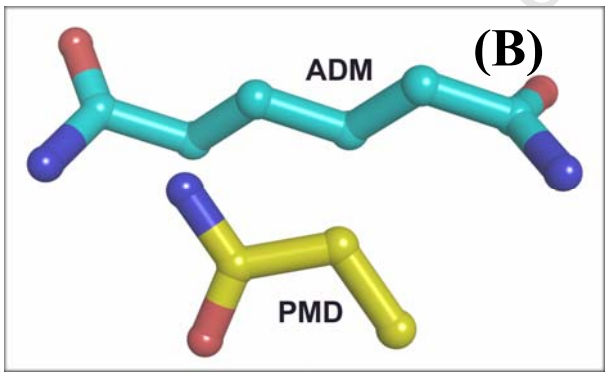


Fig. 3.19 **(A)**: A plot of the amount of ammonia produced when the WT NitN is incubated with ADM and PMD over a 60 minutes reaction period. NitN is more reactive with ADM that PMD. The error bars represent standard deviations from the mean of triplicate readings. Fig. 3.19 **(B)**: Ball and stick representation of the ADM and PMD structures. ADM is much larger and structurally different from PMD and other known short aliphatic amide substrates for NitN.

Chapter 4

The role of the two active site glutamate residues in catalysis

4.1 Abstract

The proposed or suggested roles of the catalytic residues in amidases have never been elucidated in detail. It has been widely proposed that the 'first' active site glutamate (Glu1) acts as the general base catalyst in amidase reactions, while a counterproposal (Kimani et al 2007) is that the 'second' glutamate (Glu2) is the general base catalyst that hydrolyzes the thioester intermediate. Site-directed mutagenesis of the two active site glutamates (Glu61 = Glu1 and Glu139 = Glu2) in NitN, followed by mass spectrometric and X-ray crystallographic analysis of the mutants in the presence of short aliphatic amides were undertaken, in order to probe the role of these residues in catalysis. Both of these residues were mutated in the hope of trapping the thioester after reaction with the amide substrates. Two mutants of Glu61 (E61Q/L), four of Glu139 (E139Q/L/A/D) and two double mutants of Glu61 and Cys165 (E61Q/C165A and E61L/C165S) were prepared, all of which were largely expressed as insoluble aggregates in *E. coli*, indicating the importance of the two glutamates in maintaining the fold of NitN. All mutants involving either Glu61 or Glu139 lacked enzymatic activity showing that both these glutamates are required in the catalytic process. The E139Q mutant could only be crystallized after the active site pocket was stabilized by a cacodylate adduct to Cys165 suggesting that the loop in which the mutation is located was unstable. No Cys165 adducts resulting from reaction of E139Q mutant with the amide substrates were observed by mass spectrometric studies. Mass spectrometry showed that thioester acyl-enzyme intermediates of fluoroacetamide (FAE), acetamide (ACE) and propionamide (PMD) could be trapped by mutating Glu61, and the thioester intermediate of PMD was observed at low atomic occupancy in the crystal structure of E61Q with PMD. Incubation of E61Q/L mutants with FAE resulted in an S_N2 reaction involving the displacement of the fluorine in fluoroacetamide by the catalytic cysteine (Cys165), which was detected by mass spectrometry and visualized in the crystal structures. The formation of the unexpected S_N2 reaction adduct indicated incorrect substrate positioning in the absence of Glu61, and this finding was supported by the crystal structure of the E61Q/C165A double mutant in complex with ACR, in which the orientation of ACR was altered relative to that of BMD and PMD in the C165A co-crystal structures. The occurrence of these reactions suggest that contrary to previous proposals, Glu61 (Glu1) is not required to catalyze nucleophilic attack by Cys165. However the entrapment of thioester intermediates of various substrates indicate that Glu61 is required for activation of a water substrate in the deacylation phase of the reaction. Quantum mechanical calculations showed that the deacylation water is positioned by hydrogen bonding to both glutamates. Other unexpected

reactions observed in this study include the modification of Cys165 to sulfinic acid in the crystal structures of E61Q mutant that was crystallized in the presence of both ACE and PMD, and an ACE molecule that was bound close but not covalently attached to Cys165 in the E61L mutant that was crystallized in the presence of ACE. In summary, the data presented in this section shows that the two active site glutamates are important for enzymatic activity, for maintaining the structural fold of NitN and for substrate positioning in the active site. In addition, Glu61 was found to be required as a catalyst in the deacylation phase of the reaction, but its role in activating the nucleophilic cysteine is questionable. These results support a Cys, Glu1, Glu2 and Lys (CEEK) catalytic tetrad in the nitrilase superfamily enzymes, as opposed to the previously suggested CEK triad.

4.2 Introduction

Reactions catalyzed by amidases of the nitrilase superfamily are proposed to proceed through several enzyme-substrate intermediates (Fig. 4.1) resulting from two nucleophilic attack steps. The proposed amide hydrolysis mechanism is initiated by the formation of non-covalent enzyme-substrate (Michaelis) complex in the active site pocket. This is followed by an acylation phase where the catalytic cysteine attacks the carbonyl carbon of the amide substrate, resulting in the formation of a tetrahedral enzyme-substrate intermediate. The amide amino (NH_2) group then acquires a hydrogen atom (possibly from a base catalyst), leading to the breakdown of the tetrahedral intermediate and the formation of a thioester acyl-enzyme intermediate, with the release of ammonia. In the deacylation phase of the reaction, an activated water substrate attacks the thioester carbonyl carbon atom leading to the second tetrahedral intermediate, which collapses to release the acid product and regenerate the native enzyme (Nakai et al. 2000). Although there is a consensus on the core steps in the pathway, details of each catalytic step, the precise role of the catalytic residues and various factors that influence catalytic efficiency still remain unclear. The proposed mechanism is analogous to the extensively studied ‘ping pong bi bi’ mechanism of serine proteases, which are a large group of enzymes that catalyze hydrolysis of a peptide bond using a conserved Ser, His and Asp (SHD) catalytic triad.

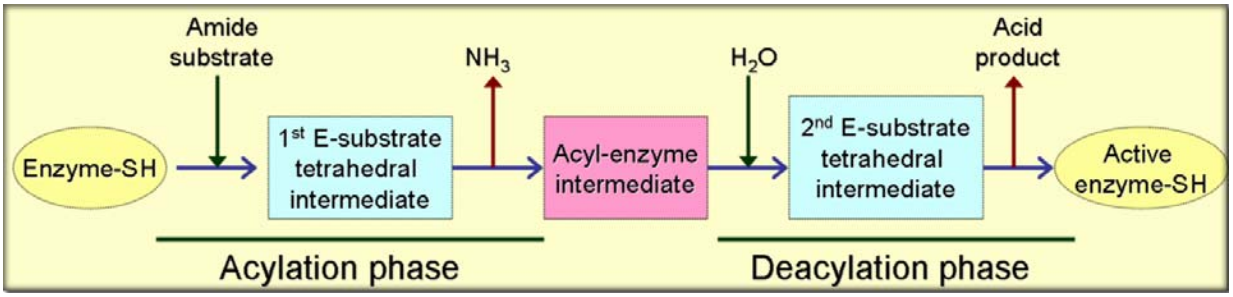


Fig. 4.1: Schematic representation of the 'ping pong bi bi' reaction pathway in amidases. The reaction is proposed to proceed through three covalent intermediates.

Highly conserved Cys, Glu and Lys (CEK) residues are generally accepted as the catalytic triad in amidases, and they are suggested to function at different stages of the catalytic process with the cysteine being the nucleophile, the glutamic acid being the general base catalyst that activates both the catalytic cysteine and the water substrate, while lysine is thought to function by stabilizing the negatively-charged transition states and the two tetrahedral intermediates in the reaction pathway. The roles of the glutamate and the lysine of the CEK triad have however never been investigated or confirmed experimentally. A second, structurally-conserved glutamate residue (Glu2) that is not recognizable from sequence conservation alone is present in the active sites of all nitrilase superfamily enzyme structures (Fig. 4.2), but its role in catalysis has never been elucidated in detail. In the co-crystal structures of C165A *Nesterenkonia* amidase (NitN) mutant with butyramide (BMD) and propionamide (PMD) bound in the active site, the two active site glutamates are involved in the positioning of the substrates, by hydrogen bonding to the amide amino group. Similar interactions have been observed in the crystal structures of the *H. pylori* formamidase and *Agrobacterium* carbamylases in complex with substrate molecules (Chen et al. 2003;Hashimoto et al. 2003;Hung et al. 2007).

Different roles of the second glutamate (Glu2) have been suggested previously during the study of amidases. Hung and colleagues (2007) suggested that Glu2 could be important for maintaining the side chain geometry of the CEK triad, as well as in facilitating the docking of substrate molecules in the active site of the formamidase from *H. pylori* formamidase. Kimani et al. 2007 suggested a catalytic role for the residue in which it acted as a general base catalyst for the hydrolysis of the thioester intermediate. This hypothesis was tested by Soriano-Maldonado and collegues (2011) in their recent biochemical study on the formamidase from *Bacillus cereus*, where they created an inactive mutant in which Glu2 was replaced by an

aspartate. The mutant was confirmed to be folded and the binding of formamide was speculated, but the experiments were unable to determine whether the loss of activity was due to the failure of aspartate to act as a general base catalyst or it was due to other factors. Buddelmeijer and Young (2010) also speculated that Glu2 could play the role of activating the nucleophilic cysteine together with the 'first' glutamate (Glu1), or it could be involved in positioning or stabilizing the phospholipid substrates in the active site of the *E. coli* apolipoprotein N-acyltransferase.

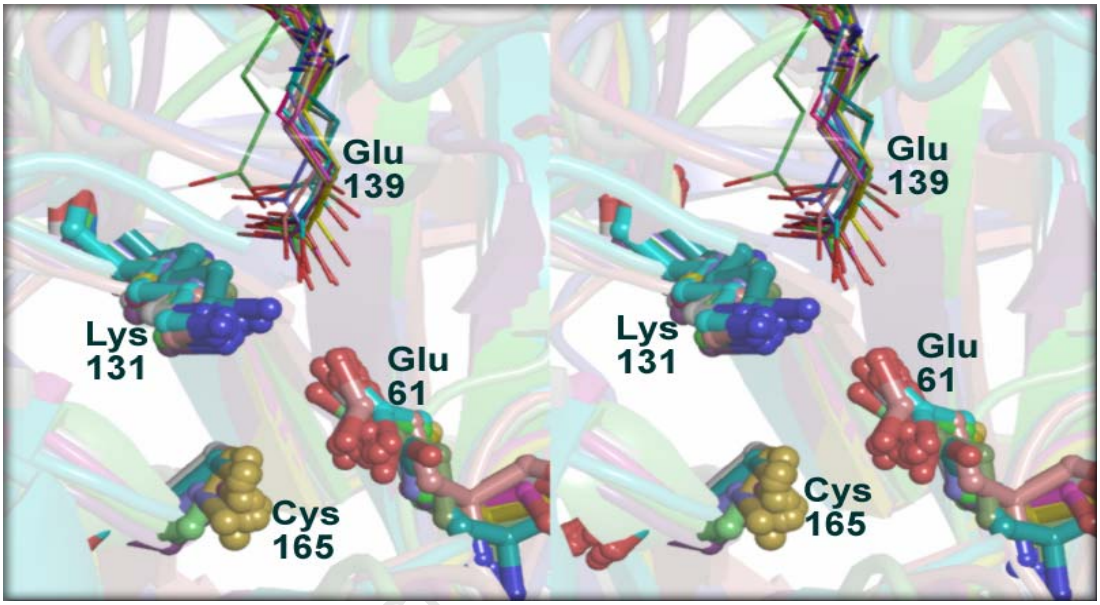


Fig. 4.2: A stereo view of the superposition of the catalytic residues in the 15 unique structures of nitrilase superfamily enzymes. The generally accepted catalytic triad residues are shown in ball and stick representation while the 'second' glutamate is rendered using lines. Labeling follows the numbering of *Nesterenkonia* sp. amidase (NitN; PDB id, 3hkx). The position and configuration of the 'second' glutamate residue is conserved in all structures, except for the worm NitFhit structure (green CPK; PDB id, 1ems) which has a loop where this residue is located in an open conformation, as detailed in section 1.3.2 and figure 1.15.

Various strategies have been utilized in the study of enzymatic mechanisms and the role of catalytic residues in literature. For instance in serine proteases, the hydrolytic enzymes with a well characterized 'ping pong bi bi' mechanism, site-directed mutagenesis has been used extensively to characterize the catalytic residues and understand various aspects of the catalytic process (Amerik et al. 1991;badie-McFarland et al. 1982;Bryan et al. 1986;Carter and Wells 1988;Craik et al. 1987;Hahn and Strauss 1990). Spectroscopic observations using poor substrates and 'partitioning' studies have provided evidence for the formation of acyl-intermediates during serine proteases catalysis (Wilmouth et al. 2001), while use of substrate

analogue inhibitors has allowed visualization of acyl-intermediates using X-ray crystallography (Katona et al. 2002;Wilmouth et al. 1997). Use of transition state inhibitors on the other hand has provided evidence for the two tetrahedral intermediates in the reaction pathway (Sienczyk and Oleksyszyn 2009) and high resolution structures for cryogenically trapped tetrahedral intermediates in the hydrolysis of the acyl-enzyme complex are available (Wilmouth et al. 2001). Quantum mechanical calculations have also contributed immensely to the understanding of the chemistry of serine proteases hydrolytic mechanism (Zhang et al. 2002).

Although details of the catalytic mechanism of nitrilases and amidases are still unclear, various studies have provided useful mechanistic insights. Kinetic studies on a number of nitrilases revealed the 'ping pong bi bi' reaction pathway (Maestracci et al. 1986;Fournand et al. 1998b) and mass spectrometric analysis using poor substrates confirmed the formation of reaction intermediates in these enzymes (Stevenson et al. 1990). Site-directed mutagenesis has been used widely to identify the catalytic residues in various nitrilase enzymes (Bellinzoni et al. 2005;Grifantini et al. 1996; Kobayashi et al. 1992a;Kobayashi et al. 1992b; Novo et al. 1995) and recent biochemical studies by Buddelmeijer and Young (2010) have detected acyl-enzyme intermediates of apolipoprotein N-acyltransferase in the extracytoplasm of *E. coli*. The crystal structure of the amidase from *Pseudomonas aeruginosa* (Andrade et al. 2007) had a trapped tetrahedral intermediate that was formed unexpectedly following the nucleophilic attack by hydroxylamine on the thioester intermediate of acetamide, which provides useful insights into the catalytic process in amidases. No other structural work or mass spectrometry has been attempted in order to fully characterize the reaction intermediates in members of the nitrilase superfamily. Site-directed mutagenesis was the method of choice in this study as it had the potential of allowing characterization of each catalytic glutamate residue separately and hopefully allowing visualization of reaction intermediates upon incubation of the mutants with the amide substrates.

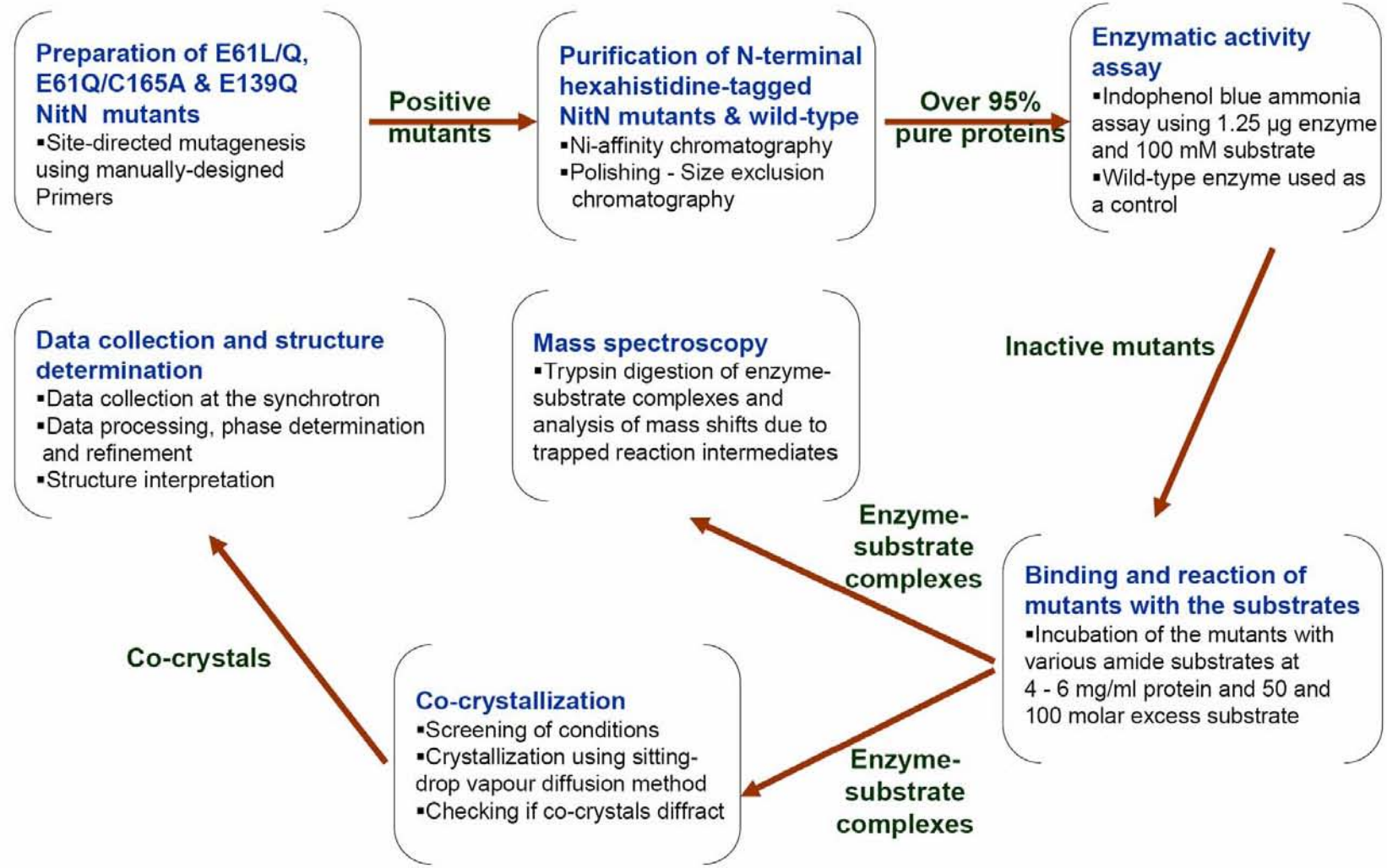
The work presented in this chapter was aimed at investigating the role of the two glutamate residues in the amidase-catalyzed reactions, particularly finding out which one of the two is involved in the second nucleophilic attack step, following the proposed role of the 'second' glutamate by Kimani and colleagues (2007). The experimental design was based on the hypothesis that, if the glutamate that functions as the general base catalyst in the second nucleophilic step is absent, the substrate will react with the cysteine forming the thioester acyl-enzyme intermediate, which will be trapped in the active site. The trapped intermediates can be

detected using mass spectrometry (MS) and visualized through crystallization of the reacted complexes. Another possible consequence of mutating either of the two glutamates, based on the co-crystal structures of C165A NitN with BMD and PMD, could be that the substrates will be mispositioned in the active site making nucleophilic attack at the carbonyl carbon impossible. Attempts were made to mutate the two glutamates to an isosteric residue like glutamine as a similar-sized side chain was likely to preserve the structure of the active site. The maintenance of the negative charge was also considered by replacing glutamate with an aspartic acid in some cases. A change of glutamate to a leucine was also attempted in this study as it has provided a stable mutant of the 'second' glutamate (Glu2) in the aliphatic amidase from *Gebacillus pallidus* RAPc8 (Our unpublished data).

Details of the incorporation of mutations to the two active site glutamates (Glu61 and Glu139) of NitN, reactions with various amide substrates, mass spectrometry, crystallographic determination of the structures of complexes and their analyses are reported in this chapter. Preparation and characterization of the NitN double mutants, E61Q/C165A and E61L/C165S, will also be described. Co-crystallization of these mutants with amide substrates was aimed at trapping non-covalently bound substrates, thereby providing insights into substrate positioning in the NitN pocket in the absence of Glu61.

4.3 Methods - A summary

The steps followed in the preparation of E61Q/L, E61Q/C165A and E139Q mutants, purification, crystallization and structure determination of NitN co-crystal structures. The wild-type NitN enzyme was purified following the same procedure and was used as a control in the characterization and crystallization steps. Details of each step are found in the materials and methods section (Chapter 2.0).



4.4 Results and discussion

4.4.1 Preparation of the glutamate mutants and protein expression

(A) Glu139 mutants

The 'second' active site glutamate (Glu139) was mutated to both leucine and glutamine and the successful incorporation of mutation checked by sequencing. Over-expression of the recombinant mutants in *Escherichia coli* BL21 (DE3) cells resulted in incorrectly folded proteins, which were deposited into insoluble inclusion bodies. These findings clearly indicated that Glu139 is not only important for substrate binding, but is also required for stability and proper folding of the active site and the overall three dimensional (3D) structure. Two more mutants were generated, E139D and E139A, but they were also found to be insoluble upon over-expression.

Several solubility-enhancing strategies were tested with all the four mutants, including: (1), addition of 1% glucose to the expression media to prevent the possibility of leaky expression prior to induction; (2), lowering of the expression temperature (to 16 °C) in combination with transient induction using low concentrations of the inducer, IPTG (0.1 mM) in order to slow down protein expression rates and allow the expressed proteins time to fold correctly; (3), heat- and cold-shocking of the recombinant cells for 20 minutes just before induction, with the hope that the expressed cold- and heat-shock proteins would assist in refolding the expressed target protein; and (4), use of rich *E. coli* growth media (2 times tryptone and yeast extract (2X YT)) to compensate for prolonged expression periods (24 hours) and encourage the cells to grow at 16 °C. Only the E139Q mutant achieved partial solubility as confirmed by SDS-PAGE gels in figure 4.3 below.

Another soluble expression optimization strategy that was attempted with the E139Q and E139L mutants only was co-expression with chaperone proteins. Plasmids expressing a combination of chaperones including *groEL, groES, dnaK, dnaJ, grpE* and *tig* (TaKaRa BIO INC., Japan) were co-transformed with the mutant DNA plasmids into *E. coli* BL21 (DE3) cells. The presence of specific chaperones during expression is known to assist with the refolding of some recombinant target proteins (Thomas et al. 1997), leading to increased

solubility. However, this was not the case with most E139 mutants as improved soluble expression could not achieved.

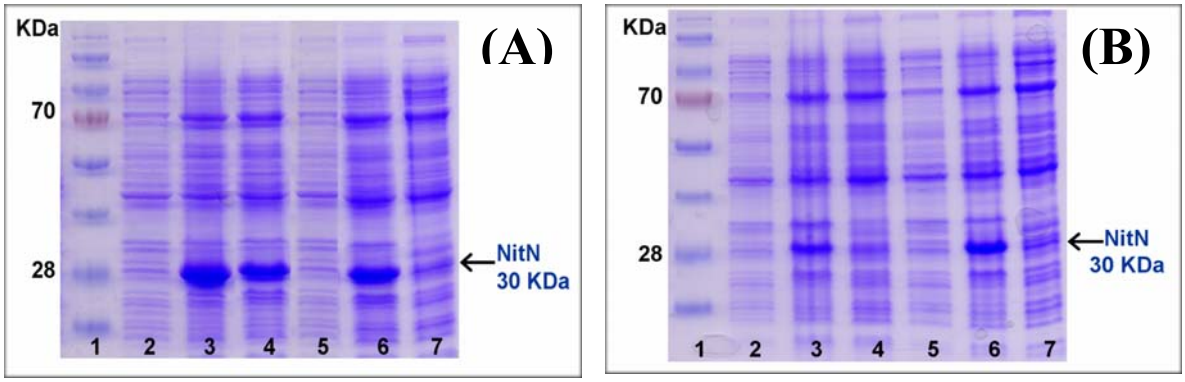


Fig. 4.3: SDS-PAGE gels for detecting soluble expression in the E139Q/L/D/A NitN mutants. **(A)** Gel for optimized expression of E139Q and E139L mutants. Lane 1, molecular weight marker (MWM); lane 2, uninduced cells for recombinant E139Q; lane 3, total cell extract containing expressed E139Q protein; lane 4, soluble cell-free extract containing E139Q protein; lane 5, uninduced cells for recombinant E139L; lane 6, total cell extract containing expressed E139L protein; and lane 7, soluble cell-free extract containing E139L protein. **(B)** Gel for optimized expression of E139D and E139A mutants. Lane 3, total cell extract containing expressed E139D protein; lane 4, soluble cell-free extract containing E139D protein; lane 6, total cell extract containing expressed E139A protein; and lane 7, soluble cell-free extract containing E139A.

In a further bid to improve soluble expression of the mutants, E139Q and E139L mutants were cloned into pCold™ cold shock expression plasmids (pCold™ I and pCold™ TF) from TaKaRa BIO INC. (Japan). These plasmids encode a cold shock protein A (*cspA*) promoter that up-regulates target protein production at lowered incubation temperatures (down to 15 °C), allowing high protein yields and increased solubility during expression in *E. coli*. The pCold™ TF plasmid is a fusion cold shock vector that expresses Trigger Factor (TF) chaperone as a solubility tag. TF is a prokaryotic ribosome-associated chaperone that facilitates co-translational folding of newly expressed proteins. Expression of the two mutants in the pCold™ I vector did not improve solubility, however co-expression of the E139Q mutant with the TF solubility tag in the pCold™ TF vector produced fully soluble fusion protein (Fig. 4.4 (C)). The cleavage of the tag was not attempted however as the fusion protein was found to be very unstable and prone to degradation (Fig. 4.4 (D)). Co-expression of E139L with the chaperones (*groEL-groES-tig*) resulted in improved protein yield and solubility (Fig. 4.4 (A)), but the soluble mutant protein was unstable as it was found to precipitate after overnight storage at 4 ˚C. Although the yield of E139Q improved when the

mutant was co-expressed with chaperones (d*naK-dnaJ-grpE-groES-groEL*), the percentage of the soluble protein fraction did not increase significantly (Fig. 4.4 (B)).

Only the E139Q mutant yielded soluble, stable protein, hence it was studied further by mass spectrometry and X-ray crystallography. Enough protein was obtained by expression of several litres of culture in different batches, with the expressed protein from each batch being pooled during the purification step.

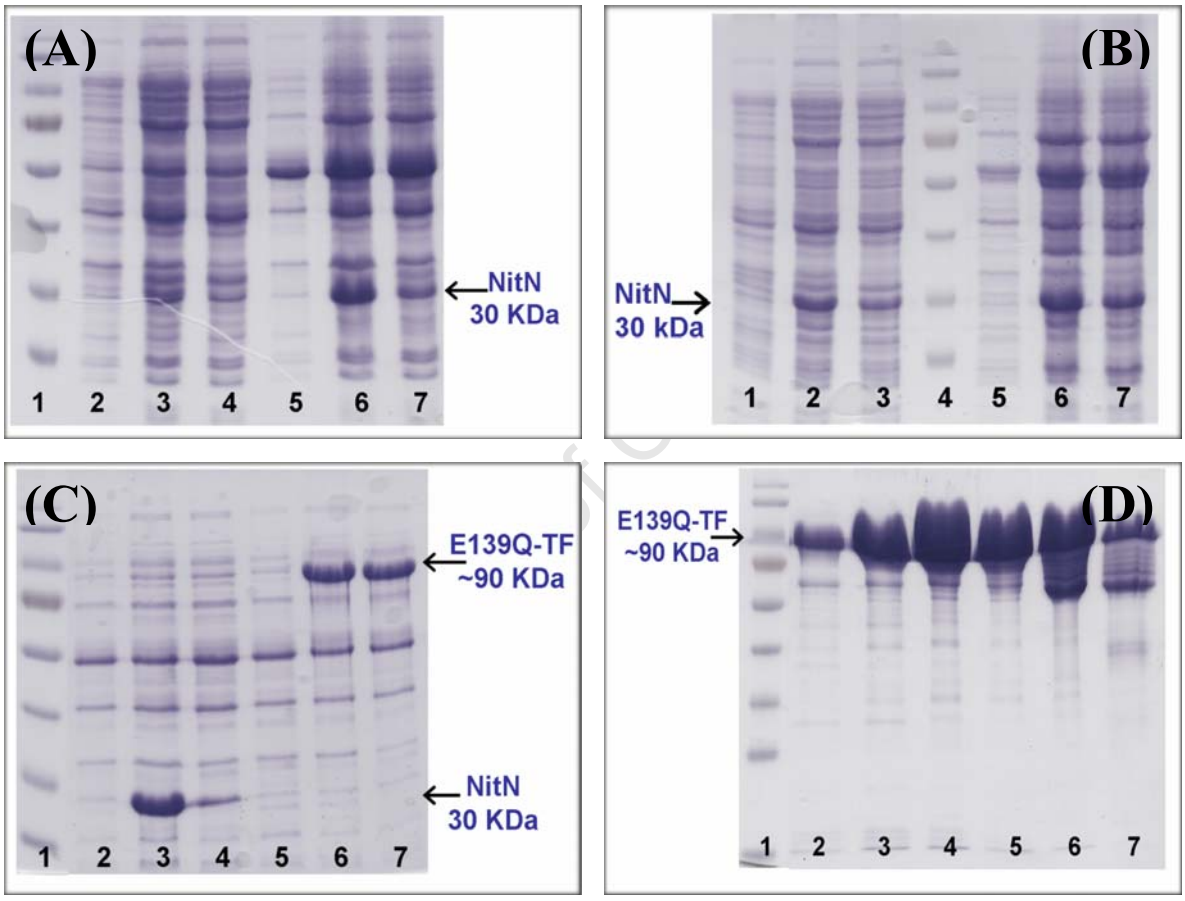


Fig. 4.4: SDS-PAGE gels for the expression of E139Q/L mutants in the cold shock pCold™ plasmids and co-expression of the same mutants with chaperones. **(A)** Gel for co-expression of E139L mutant with *groES-groEL-tig* chaperones. Lanes 3 and 4 show total and soluble E139L proteins expressed without chaperones respectively, while lanes 6 and 7 represent total and soluble proteins expressed in the presence of chaperones respectively. Improved expression and solubility is observed with chaperones. **(B)** Gel for co-expression of E139Q with *dnaK-dnaJ-grpE-groES-groEL* chaperones. Lanes 2 and 3 represent total and soluble proteins expressed without chaperones respectively, while lanes 6 and 7 show total and soluble proteins respectively in the presence of chaperones. The yield of the expressed proteins increase but the soluble fraction is not improved. **(C)** Gel for detecting expression of E139Q in the pCold™ I and pCold™ TF vectors. Lanes 3 and 4 show total and soluble proteins from pCold™ I plasmid respectively, while lanes 6 and 7 show total and soluble E139Q-TF fusion protein respectively from pCold™ TF vector. **(D)** Gel for fractions from purified E139Q-TF fusion protein showing degradation (many low molecular weight bands) upon overnight storage after elution from the gel filtration column.

(B) Glu61 mutants

The 'first' active site glutamate (Glu61) of NitN was mutated to both glutamine and leucine. Both mutants expressed relatively well, with only partial solubility being achieved in both cases, an indication of some degree of instability in the 3D structure. The soluble expression optimization strategy included use of 2X YT growth medium supplemented with 1% glucose, transient expression induction with 0.1 mM IPTG and overnight expression at 16 °C. As for the E139Q mutant, the mutants were expressed in large volumes of bacterial culture in order to obtain enough material for biochemical and biophysical characterization, and structural studies.

(C) Glu61/Cys165 double mutants

Mutations at position 61 were incorporated in the already existing C165A/S NitN mutant plasmids to generate plasmids with mutations at the two positions. Glu61 was changed to glutamine and leucine in both C165A and C165S mutants, generating double mutants of all four possible combinations. Only E61Q-C165A and E61L-C165S mutant plasmids were suitable for protein expression based on the sequencing results; the plasmid DNA for the other two mutants was found to form secondary loop structures, which made sequencing difficult. Although initial attempts to express the E61Q-C165A and E61L-C165S mutants resulted in insoluble proteins, partially soluble protein was obtained using the same optimized expression conditions as the E61Q/L mutants.

4.4.2 Purification and characterization of the mutants

All the mutant proteins were purified using Nickel-affinity chromatography, and polished by a gel filtration step as described in the materials and methods section (Chapter 2). E61Q and E61Q-C165A mutants eluted as single symmetric peak off gel filtration column, while E139Q, E61L and E61L-C165S mutants had other high molecular weight (MW) peaks in addition to the symmetric dimer peak (Fig. 4.5). Some of the high MW peaks were broad, suggesting a possibility of different oligomeric forms or semi-aggregated proteins co-eluting. As was also the case for the wild-type (WT) NitN and C165A/S mutants, the estimated molecular weight of the dimeric E61L/Q and E139Q mutants from size exclusion chromatography was 45 kDa instead of the expected 60 kDa. This is thought to be due to

interactions of the protein with column gel, resulting in the proteins not showing a classical Stokes elution behavior. SDS-PAGE gels revealed over 95% purity and homogeneity in the dimer peaks, while the mutant proteins from the high MW peaks were contaminated by other proteins (Fig. 4.5 (D)). Only protein from the dimeric peaks was used for biochemical characterization and structural studies.

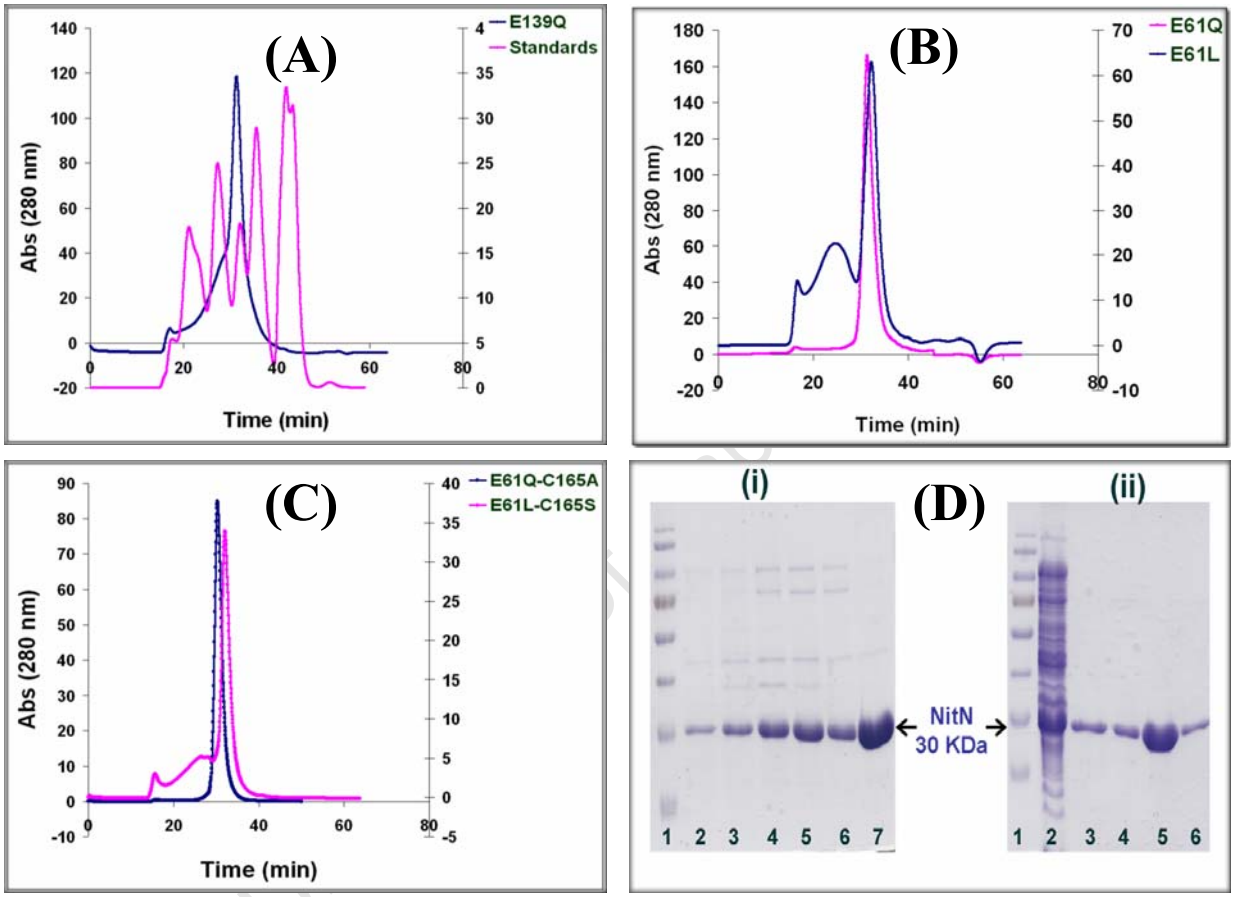


Fig. 4.5: Gel filtration purification profiles of the E61L/Q, E61Q-C165A, E61L-C165S and E139Q NitN mutants. **(A)** Superimposed elution profiles for the E139Q mutant and the gel filtration protein standards. The major peak corresponds to the dimeric E139Q form while other smaller high MW peaks are also observed. **(B)** Gel filtration elution charts for E61Q and E61L NitN mutants. E61L has additional high MW peaks. **(C)** Elution profiles for the E61Q-C165A and E61L-C165S mutants. The E61L-C165S profile has high MW peaks in addition to the major dimeric peak. **(D)** SDS-PAGE gels of the gel filtration fractions from the E61L (i) and E61Q-C165A (ii) mutants. Lanes 2-6 from the E61L fractions (i) are from the high MW peaks and faint bands for higher MW contaminating proteins are visible, while lane 7 is from a fraction in the dimeric peak. Lanes 3-6 in the E61Q-C165A mutant (ii) show fractions from the dimeric peak; high purity levels are observed. The gel was loaded with equivalent volumes of the protein samples.

All mutants were found to be unstable under elution conditions, with the E139Q, E61L and E61L-C165S being more prone to precipitation during handling. 10% glycerol was included in all purification and storage buffers, which helped reduce precipitation and enhance solubility at relatively high protein concentrations. Glycerol was however omitted from all biochemical, biophysical and crystallization studies of the mutants, as it was found to interfere with the binding of the substrates in the WT NitN enzyme. None of the mutants had enzymatic activity as no ammonia was detected using the bromophenol blue assay. This provides evidence that the two glutamates are important for enzymatic activity and also supports the proposed roles, where one or both residues are thought to be involved in catalysis. In addition, the two glutamate residues are clearly important in maintaining the configuration of the active site and in enhancing stability of the overall 3D structure based on the partial insolubility during expression and the observed instability during handling.

In the native NitN structure (WT-Apo), the two active site glutamates (Glu61 and Glu139) are involved in a hydrogen bonding network (Fig. 4.6) that is important for maintaining the side chain conformation of the active site residues. Removing or changing either of the two glutamates would disrupt the structure of the active site pocket, which would destabilize the entire 3D structure; this may explain why the two glutamate mutants were unstable. In addition to maintaining the active site structure, the two glutamates ensure a viable environment for catalysis. The interactions of the catalytic lysine (Lys131) with the negatively-charged carboxylates of the two glutamates would raise the pKa of its amino group, ensuring that it remains protonated throughout the reaction. As suggested by Andrade and colleagues (2007), the positive charge of the lysine withdraws electrons from the carbonyl group of the amide, promoting the nucleophilic attack by the cysteine. Moreover, in the two subsequent tetrahedral intermediates, the catalytic lysine acts together with the backbone amide of the residue next in sequence after the cysteine (Tyr166 in NitN) to stabilize the carbonyl oxyanion.

The instability of E61Q/L NitN mutants was not surprising as similar protein instabilities have been observed in mutants of the equivalent glutamate residue in other amidases. Site-directed mutagenesis of the Glu1 residue (E59Q/N/V) in the amidase from *Pseudomonas aeruginosa* resulted in unstable protein, with the hexameric quaternary structure disintegrating to inactive trimers or dimers depending on the mutation (Karmali et al. 2000;Novo et al.

2002). This clearly indicates the importance of Glu59 in catalysis and in maintaining the quaternary structure of the *P. aeruginosa* amidase.

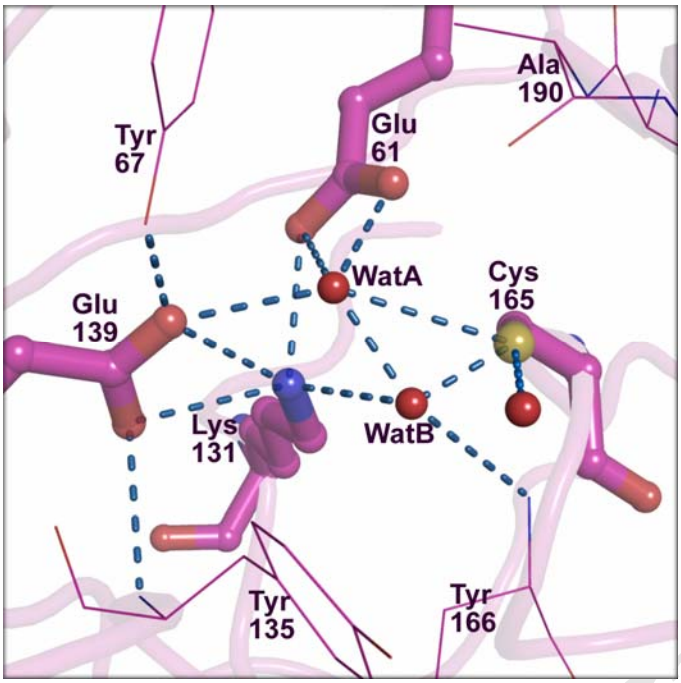


Fig. 4.6: Configuration of the catalytic unit in the WT NitN active site pocket. The Cys, Glu1, Glu2 and Lys (CE_1E_2K) tetrad residues are rendered in ball and stick representation, possible hydrogen bonding interactions within the active site are shown as blue dashes and water molecules are shown as red non-bonded spheres. The two glutamates are involved in a hydrogen bonding network that maintains the active site structure.

4.4.3 Mass spectrometry of the mutants with short aliphatic amides

Mass spectrometry (MS) was used to detect the formation of reaction intermediates covalently bound to the catalytic cysteine, should any occur due to the removal of either of the two glutamic acid residues. A tryptic digestion fragment of 17 amino acids bearing the reactive cysteine (QLSLLV**C**YDVEFPEMVR) has a predicted molecular weight of 2041.0133 Da as shown in table 4.1 (A). Reaction intermediates and adducts covalently attached to the catalytic cysteine would result in a fragment with increased mass, with the mass difference giving an indication of the molecular weight (MW) and hence the nature of the adduct that is trapped in the active site. Four NitN short aliphatic amide substrates (propionamide, fluoroacetamide, butyramide and acetamide) were reacted with the E139Q and E61L/Q mutants using the same conditions of pH (7.5), salt (150 mM) and temperature (25 °C) as the biochemical enzymatic assays. Parallel experiments were performed with the wild-type (WT) enzyme as a control. The reactions were carried out for two hours before being stopped by transferring the reaction mixtures to 4 °C. The reacted proteins were then trypsinized and subjected to MALDI-TOF

MS to check for modifications at the catalytic cysteine. The interpretation of the MS data was achieved through manual comparison of the mass spectra from the reacted mutant proteins with those of the unreacted (apo) mutants and the reacted WT enzyme. The expected mass shifts and total masses of the tryptic fragments corresponding to thioester acyl-enzyme intermediates of the four amide substrates are presented in table 4.1 (B).

Table 4.1 (A) Predicted peptide masses for 93.3% of the NitN sequence

Mass (Da)	Position	Peptide sequence
6805.2848	207-274	AVENGITLAYANHCGPEGGLVFDGGSVVVGPAGQPLGE LGVEPGLLVVDLPDQSQDAGSDSADYLQDR
3893.9377	49-85	ASEQGAQLLLTPELFGFGYVPSQICAQVSAEQVDAAR
2895.4893	23-48	IALMQHTARPLDPQHNLDLIDDAAAR
2558.4551	181-206	GAQLVLVPTALAGDETSVPGILLPAR
2041.0133*	159-175	QLSLLV**C**YDVEFPEMVR
2031.0029	113-131	GITAELADEHGEVLASYQK
1907.0173	141-158	AAFVPGEQPPPVLSWGVR
1899.8892	1-17	MGSSHHHHHHSSGLVPR
1805.9544	96-112	GIALVWSLPGPEGPEQR
1062.5466	132-140	VQLYGPEEK
625.3416	276-280	AELHR
587.2718	18-22	GSHMR

*2041 is the mass of the tryptic peptide bearing the catalytic cysteine (Cys165).

Table 4.1 (B) Expected mass changes in the tryptic fragment bearing the catalytic cysteine if covalent thioester intermediates of the four substrates are trapped in the active sites of the E139Q and E61Q/L mutants.

WT, E139Q//E61L/Q NitN and reacted substrate	Mass of the substrate (Da)	Expected mass shift (Da)	Total mass of expected tryptic fragment (Da)[a]
Apo mutants and WT	-	-	2041
Mutants + Propionamide	73	+56	2097
Mutants + Fluoroacetamide	77	+60	2101
Mutants + Butyramide	87	+70	2111
Mutants + Acetamide	59	+42	2083

[a]Masses are rounded to the nearest whole number.

No reaction or trapped adducts were observed in the mass spectra of the E139Q NitN mutant reacted with the four amide substrates; a high intensity 2041 Da peak corresponding to unreacted (free) cysteine is observed in all cases (Fig. 4.7 (A)). This may be due to the role of Glu139 in positioning the substrates for nucleophilic attack, or/and in maintaining the stability and configuration of the active site pocket, which would ensure its accessibility by amide substrates. The lack of biochemical activity or any detectable reactions in the E139Q mutant may also indicate an involvement of Glu139 in the first nucleophilic attack, probably in maintaining the charge environment that is essential for catalysis or in assisting in general base catalysis. In the co-crystal structures of the C165A NitN mutant with propionamide (PMD) and butyramide (BMD), both Glu139 and Glu61 make hydrogen bonds to the amide amino group of the substrates (section 3.4.6.2 and figure 3.7), which ensures optimal positioning that would allow stereo-electronic alignment of the carbonyl carbon and the thiol group of the reactive cysteine for nucleophilic attack. If the removal of Glu139 renders the active site pocket inaccessible or results in incorrectly positioned substrates, then the nucleophilic cysteine would not be able to attack, and therefore no reactions would take place. Similarly, no reactions would occur if Glu139 is involved in the first nucleophilic attack.

The E61L mutant reacted with fluoroacetamide (FAE) yielding a weak signal peak of the unreacted cysteine fragment at *m/z* 2041 and a higher intensity 2101 Da tryptic fragment peak (Fig. 4.7 (A)), corresponding to a covalent FAE thioester intermediate. These observation has a number of implications: (1), even though Glu61 is involved in the positioning of the substrate molecules, some substrates are still able to align correctly in its absence to result in nucleophilic attack at the carbonyl carbon; (2), another active site residue (probably Glu139) or the substrate molecule itself is able to catalyze the nucleophilic attack by the catalytic cysteine; and (3), the trapping of the acyl intermediate is an indication that Glu61 is required as a general base catalyst in the second nucleophilic attack step that involves activation of a water substrate leading up to the release of acid products. E61L did not however react with any of the other three substrates even after several repeats, except for a low intensity acyl-enzyme intermediate peak (*m/z* 2097) that was observed occasionally with PMD (appendix IV). The inability of the E61L mutant to form substantial amounts of acyl-enzyme intermediate with PMD (which is the best known short aliphatic amide substrate for NitN) even after prolonged incubation times is a clear indication that Glu61 is essential in providing a favorable environment for substrate binding and catalysis.

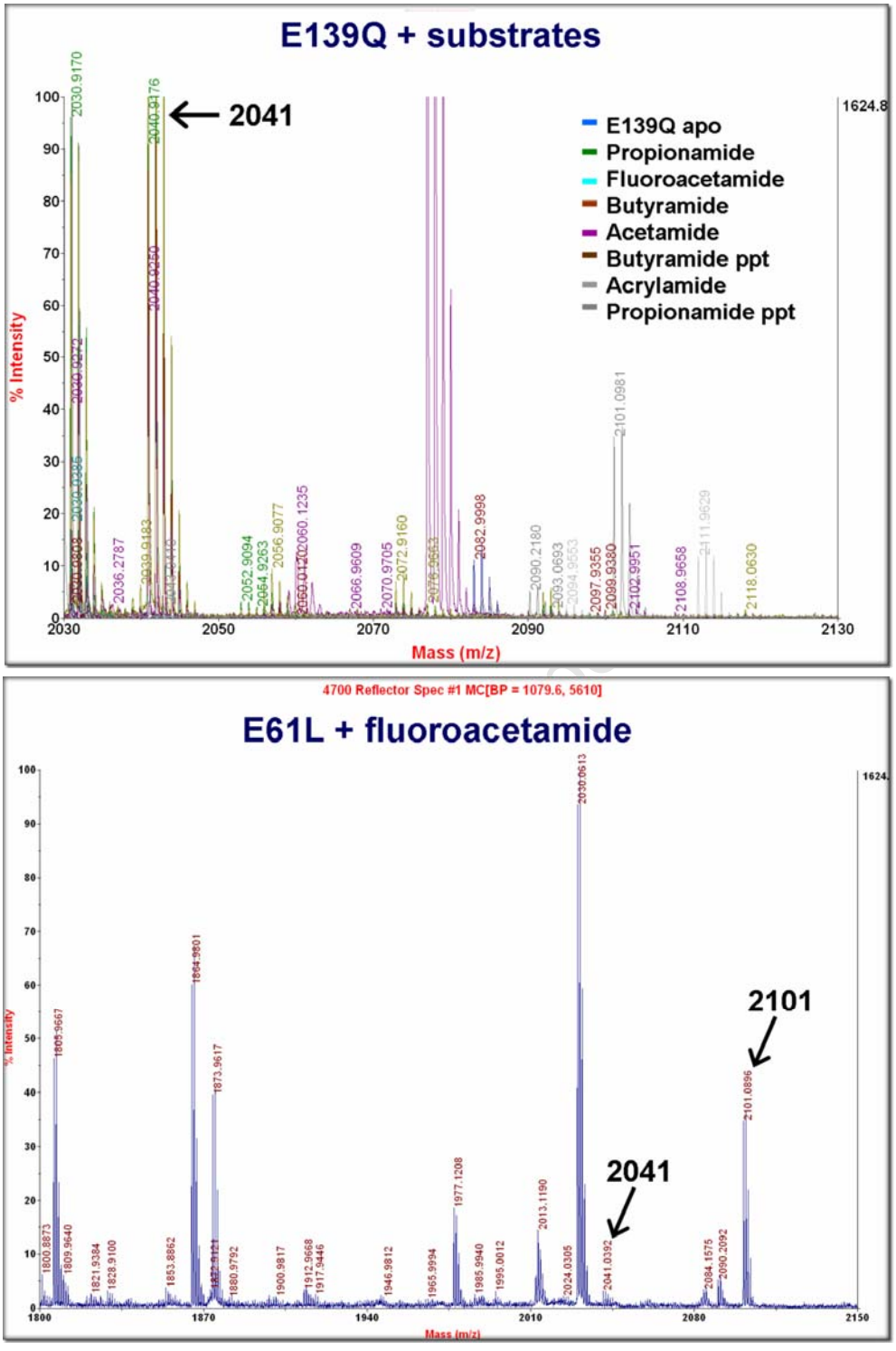


Fig. 4.7 (A): MALDI-TOF mass spectra for the E139Q and E61L NitN mutants. Top panel - an overlay of mass spectra of the E139Q NitN mutant with various amide substrates. The unreacted 2041 Da peak is visible in all cases and there is no other peak that represents a reaction of any kind with Cys165. Bottom panel - mass spectrum of E61L reacted with FAE. A small unreacted protein tryptic fragment peak (2041) is observed, as well as a 2101 peak, which is representative of a FAE acyl-enzyme intermediate trapped in the active site.

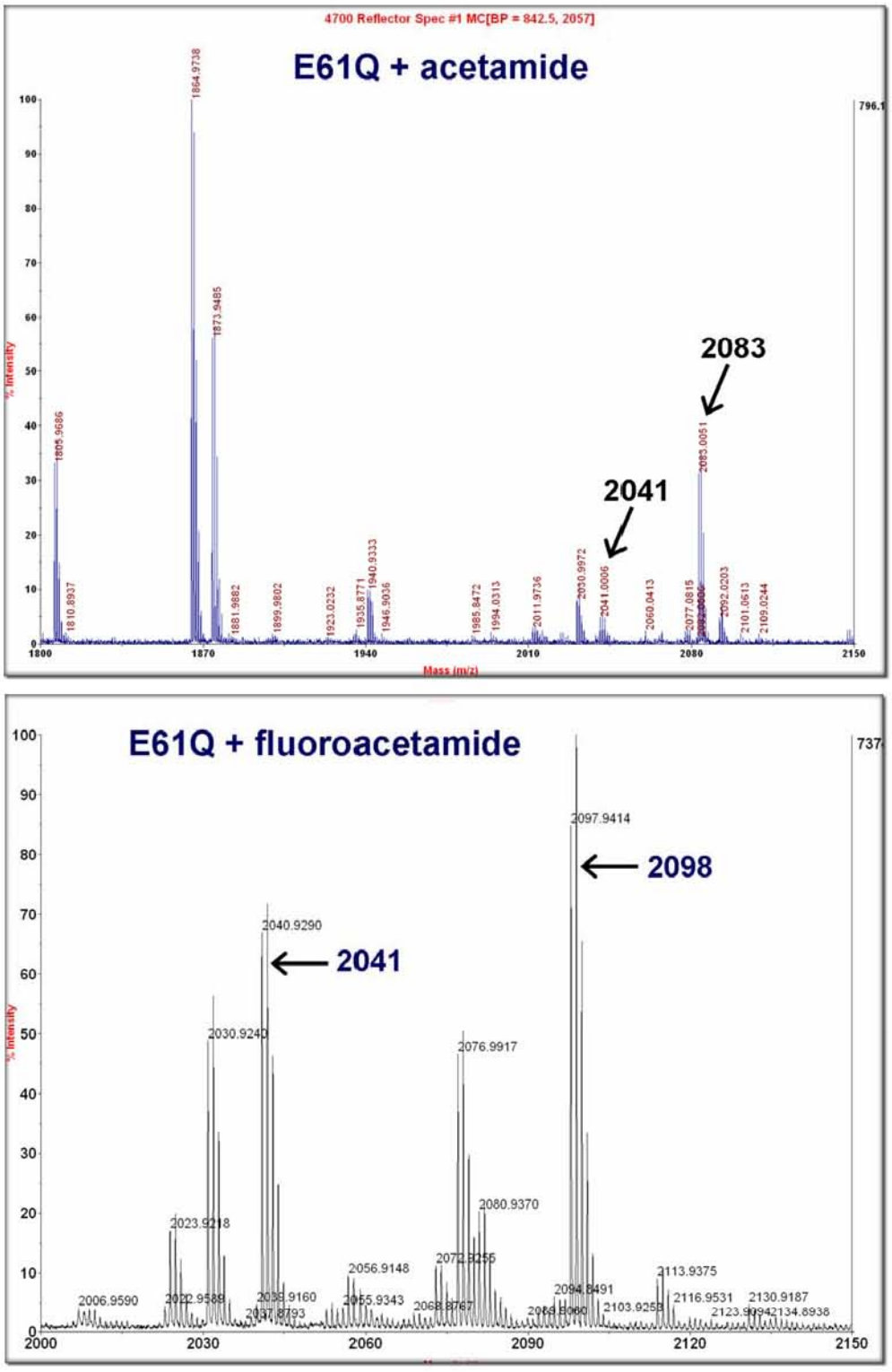


Fig. 4.7 (B): MALDI-TOF mass spectra for the E61Q NitN mutant with ACE and FAE. Top panel - mass spectrum of the E61Q reacted with ACE, showing a weak (small) free cysteine tryptic fragment peak (*m/z* 2041) and an ACE thioester intermediate fragment peak at 2083 Da. Bottom panel - mass spectrum of the E61Q reacted with FAE, showing a free cysteine fragment peak (*m/z* 2041) and an S_N2 reaction Cys-FAE adduct peak (*m/z* 2098).

The E61Q NitN mutant reacted with acetamide (ACE) to yield an acyl-enzyme reaction intermediate with a corresponding tryptic fragment peak of 2083 Daltons, in addition to a weak signal peak of unreacted cysteine at *m/z* 2041 (Fig. 4.7 (B)). On the other hand, the E61Q mutant in the presence of FAE had a dominant tryptic fragment peak of 2098 Daltons with a corresponding mass shift of +57, as well as a medium signal peak of free protein (*m/z* 2041) (Fig. 4.7 (B)). A 2098 Daltons peak is clearly not representative of the expected 2101 Da FAE thioester intermediate or of any other known intermediate in the FAE hydrolytic pathway, and it points to the possibility of another kind of reaction. The reaction accounting for the 2098 Da peak was only worked out when the crystal structure of the E61Q NitN with FAE (discussed in section 4.4.8.2 below) was observed to have an adduct at the catalytic cysteine (Cys165) that had the fluorine atom (with a mass of 19 Da) of FAE missing, implying that Cys165 was able to attack at the Cα carbon center of FAE (instead of the carbonyl carbon in the normal amidase reaction case), eliminating the fluorine atom in the process. This kind of a reaction process is referred to as S_N2 (bimolecular, nucleophilic, substitution) (Mikosch et al. 2008).

In the S_N2 reaction, the lone pair of electrons in a nucleophile attacks an electrophilic carbon center and bonds to it, expelling a leaving group in the process. An S_N2 reaction is therefore a concerted (one step) substitution process where the formation of the carbon-nucleophile bond and the breakage of the carbon-leaving group bond happen simultaneously without an intermediate step, and it proceeds via a transition state in which the carbon under nucleophilic attack is partially bonded to five groups (penta-coordinate) (Fig. 4.8). Since the nucleophile attacks at 180° to the leaving group, this substitution is termed as a backside attack and it proceeds with an inversion of configuration in chiral substrates (Jacobs 1997). S_N2 reactions are bimolecular, meaning they follow second-order kinetics as the reaction rate is dependent on the concentration and nature of both the nucleophile and the electrophilic substrate (Mikosch et al. 2008).

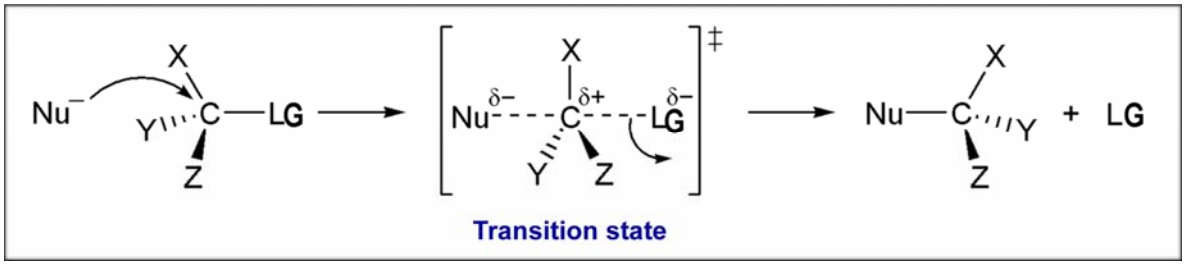


Fig. 4.8: Mechanism of S_N2 reaction. Nu is the nucleophile, while LG is the leaving group. The reaction has one transition state with the carbon under nucleophilic attack being penta-coordinate. The schematic picture was taken from a WIKIPEDIA page (http://en.wikipedia.org/wiki/File:SN2_reaction_mechanism.png) that lacked references.

All interesting results from the mass spectrometric experiments of the two catalytic glutamate mutants of NitN with various short aliphatic amide substrates, the reactions accounting for the detected adducts as well as an indication of whether the observed reactions have been visualized in structures are summarized in table 4.2.

Table 4.2 Summary of the MS results for reactions observed with the two catalytic glutamate mutants and various short aliphatic amide substrates.

NitN Mutant	Substrate (Mass (Da))	Measured mass of the tryptic fragment (Da)	Mass shift (Da)	Type of reaction or adduct observed	Visualized in Structure?
E139Q/ E61Q/L	Apo	2041[a]	-	-	-
E61Q	Acetamide (59)	2083	+42	ACE thioester intermediate	No[b]
	Fluoroacetamide (77)	2098	+57	FAE S_N2 reaction adduct at Cys165	Yes
E61L	Fluoroacetamide (77)	2101	+60	FAE thioester intermediate	No[c]
	Propionamide (73)	2097	+56	PMD thioester intermediate	Yes

[a]The mass of the tryptic fragment bearing the catalytic cysteine is the same for all unbound (apo) mutant proteins (2041 Da).
[b]The obtained crystals did not diffract.
[c]The obtained crystals had an S_N2 reaction adduct instead of the expected acyl-enzyme thioester intermediate.

4.4.3.1 The suitability of fluoroacetamide (FAE) for S_N2 reaction

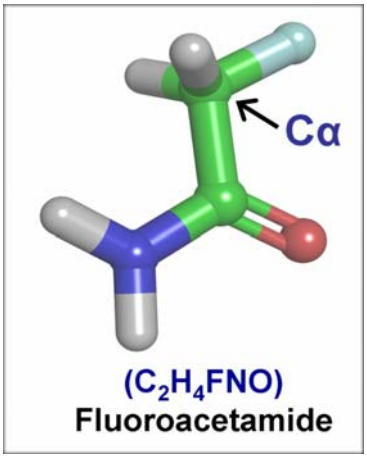


Fig. 4.9: A green CPK ball and stick representation of FAE. The α-carbon atom is labeled.

Some substrate-related structural characteristics that may influence S_N2 nucleophilic attack of FAE by the catalytic cysteine are examined below:

(1) **The steric effects of the targeted carbon center in the substrate:** For fluorine to be the leaving group, the cysteine would have to attack the Cα carbon of FAE (Fig. 4.9). As one of the requirements for an S_N2 reaction to occur is the ability of the nucleophile to attack directly (Jacobs 1997), it is important that the targeted carbon atom is not highly substituted, that its 'back' is accessible and that it is free to invert its configuration (Jacobs 1997). The Cα carbon of FAE (Fig. 4.9) meets all these requirements as it is sp^3 hybridized with only hydrogen atoms as the substituents. The configuration of the Cα carbon in FAE would therefore be favorable for an S_N2 nucleophilic attack.

(2) **The nature of the leaving group:** As one of the conditions for S_N2 reactions is the availability of a leaving group (Jacobs 1997), fluorine, being a halide, would serve as a leaving group although not a very favorable one. Since S_N2 reactions are dependent on the nature of the leaving group and the nucleophile, the reactions proceed faster with good leaving groups and good nucleophiles (Jacobs 1997). While cysteine is one of the strongest nucleophiles in biological systems (Giron et al. 2010), fluorine on the other hand is poor leaving group, as the carbon-fluoride bond is quite strong. While having fluorine as the leaving group would greatly discourage an S_N2 reaction, the electron-withdrawing property of the carbonyl group adjacent to the Cα carbon (Fig.

4.9) is expected to make the α-carbon centre more electrophilic, which would enhance fluorine as a leaving group.

(3) **The nature of the groups adjacent to the targeted carbon attack center:** As the S_N2 reaction transition state has a negative charge spread over the nucleophile, the central carbon and the leaving group (Fig. 4.8), any factors or adjacent groups that stabilize this species would enhance S_N2 reactions. The carbonyl group that is adjacent to the Cα carbon in FAE would be very effective in stabilizing the transition state (Jacobs 1997), which would encourage S_N2 displacement of fluorine.

While the chemical structure of FAE would not exclude the possibility an S_N2 substitution reaction, this reaction does not occur in the WT NitN enzyme. This implies that the change of glutamate to glutamine in the E61Q mutant creates an environment that facilitates the occurrence of this reaction. One possible explanation would be the altered positioning of the FAE molecules in the active site of the mutant given the crucial role of Glu61 in the binding of the amide amino group, which would position the Cα carbon atom for nucleophilic attack rather than the amide carbon. The co-crystal structure of the E61Q/C165A NitN with ACR (discussed in section 4.4.9.1 later in the chapter) unequivocally confirms that the substrate position is altered when Glu61 is replaced by glutamine. It is however questionable how the cysteine is activated in order to carry out the S_N2 nucleophilic attack in the absence of Glu61 (Glu1). As speculated in the case of thioester formation in the E61Q/L mutants, Glu139 (Glu2) could step in to fulfil this role, probably through a coordinating water molecule or the thiol group deprotonation could be substrate-assisted.

S_N2 reactions are common with some hydrolytic and transferase enzymes including phosphotriesterase, some restriction enzymes (e.g. BamHI), glutathione synthase, DNA topoisomerase II, adenylate kinase and ribonuclease U2, among others (Nagano et al. 2007). While these enzymes utilize a water molecule, the substrate, an acceptor group or a catalytic residue for nucleophilic attack, most of them have a catalytic residue serving as a general base catalyst for activating the nucleophile, while some like the glutathione synthase do not require a general base catalyst. Ribonuclease U2, which hydrolyzes a phosphoric ester bond, is the only enzyme among the ones listed here that uses a catalytic residue as an acid catalyst for donating a proton to the leaving group (Nagano et al. 2007).

An interesting observation is that of different reactions occurring between the E61L and E61Q mutants with FAE. The E61L NitN mutant reacted with FAE to form a trapped acyl-enzyme intermediate through the attack of the carbonyl carbon, while E61Q NitN reacted with the Cα carbon of FAE to form an S_N2 reaction adduct. These variable reactions could be as a result of differences in substrate positioning in the binding pockets of the two Glu61 mutants, as the incorporated glutamine and leucine side chains would affect the charge environment of the active site in differently. A change of Glu61 to glutamine replaces the negatively charged carboxylate group with a polar but uncharged side chain while a replacement with leucine introduces a highly hydrophobic side chain at physiological conditions.

In all the cases where reaction intermediates and adducts were detected in the E61Q/L NitN mutants incubated with the substrates, a 2041 Da fragment peak was also present (Fig. 4.7), implying that a proportion of the protein always remained unreacted despite the fact that the reactions were carried out for prolonged incubation periods (2 hours) prior to mass spectrometric analyses. This observation provides clear evidence that the probability of the reactions occurring in the absence of Glu61 is quite low, and together with the evidence for the occurrence of 'abnormal' S_N2 reaction between the mutants and FAE, it emphasizes the importance of Glu61 in the positioning of amide substrates for nucleophilic attack.

In summary, the mass spectrometry data have revealed that in the absence of Glu61 (Glu1), the catalytic cysteine is able to react with the carbonyl group of some amide substrates to generate a thioester acyl-enzyme intermediate that is trapped in the active site. In addition, the cysteine of the E61Q mutant unexpectedly reacts with the α-carbon of fluoroacetamide (FAE), eliminating fluorine and covalently adducting to the acetamide moiety in an S_N2 substitution reaction. These reactions are summarized in table 4.2 above and their occurrence is supported by structural data presented in sections 4.4.7 and 4.4.8 below. These findings raise questions around the mechanism of the cysteine nucleophile activation (deprotonation) in the light of the widely proposed role of Glu1 that it mediates proton transfer (general base catalysis) in both the acylation and deacylation phases of the reaction. The occurrence of the unexpected S_N2 reaction indicates altered positioning of the substrate molecules within the binding pocket. Mass spectrometry also showed that Glu139 (Glu2) mutant was unable to react or form adducts with any of the four substrate molecules (PMD, FAE, BMD and ACE), strongly suggesting that the substrates were incorrectly positioned for nucleophilic attack.

4.4.4 Crystallization of the mutants reacted with short aliphatic amides

For crystallization experiments, the mutants were reacted with the amide substrates following the same conditions as the biochemical activity and mass spectrometry assays. After incubation, the mutants were centrifuged to remove any precipitated material, and then transferred to ice prior to crystallization. Due to structural instability of the mutants, obtaining diffraction quality crystals was quite challenging and difficult to achieve in most cases. Glycerol, which is popular for improving protein solubility during crystallization, could not be used in the crystallization reaction mixtures, as it has been shown to prevent substrates from accessing the active site in the WT enzyme. Other solubility-enhancing agents like glycine, sucrose, urea, ethylene glycol and propanediol were tested and only ethylene glycol was found not to interfere much with the activity of the enzyme (data not shown). 8% ethylene glycol was therefore incorporated into the protein buffer prior to reaction with amide substrates and crystallization. Known crystallization conditions (for the WT and the C165A/S mutants) were attempted first, with screening for new conditions (using Crystal Screen HR2-110 and PEG/Ion Screen HR2-126 kits from Hampton research) being performed where necessary. The concentration of the protein and the solubility-enhancing additives was varied in some cases to improve crystallization. Table 4.3 (A) shows the reservoir buffers that produced diffraction quality crystals of the mutants.

Although crystallization was attempted with all the mutants (E61L, E61Q, E139Q, E61Q/C165A and E61L/C165S) reacted with a range of short aliphatic amide substrates, only few crystals were good enough for diffraction. In most cases, heavy precipitates were observed in the crystallization drops, while in some instances only salt crystals were obtained. The crystals that were selected for diffraction data collection and structure determination are summarized in table 4.3 (B), and images of some of the crystals within their crystallization drops are presented in figures 4.10 (A) and (B). The Co-crystallization of the E61/C165 double mutants with the substrates was aimed at non-covalent binding of substrate molecules in the active site in order to visualize their positioning and orientation in the absence of Glu61. No crystals were obtained with the E61L/C165S mutant, despite extensive screening efforts of a wide range of crystallization conditions and use of additives.

Table 4.3 (A): Crystallization reservoir buffers for the E61Q/L, E139Q and E61Q-C165A Nit mutants.

Crystallization buffer[a]	Contents
4	0.1 M Tris-HCl pH 8.5, 2.0 M ammonium sulfate
15	0.1 M sodium cacodylate trihydrate pH 6.5, 0.2 M ammonium sulfate, 30% polyethylene glycol (PEG) 8000
39	0.1 M HEPES sodium pH 7.5, 2% PEG400, 2.0 M ammonium sulfate
32	2 M ammonium sulfate

[a]The reservoir buffers are named according to their numbering in the Hampton Research Crystal Screen HR2-110 kit.

Table 4.3 (B): A summary of the crystals that were used for structure determination

Crystal name	NitN Protein	Substrate/ adduct	Crystallization buffer
E139Q-Cac	E139Q	Cacodylate[a]	15
WT-Cac	WT	Cacodylate[b]	15
E61L-FAE	E61L	FAE	39
E61L-PMD	E61L	PMD	39
E61L-ACE	E61L	ACE	39
E61Q/C165A-Apo	E61Q-C165A	ACE[c]	39
E61Q/C165A-ACR	E61Q-C165A	ACR	32
E61Q-FAE	E61Q	FAE	4
E61Q-ACE	E61Q	ACE	39
E61Q-PMD	E61Q	PMD	39

[a]The E139Q-Cac crystal was obtained from the E139Q mutant protein pre-incubated with PMD and crystallized in a cacodylate-containing reservoir buffer. A cacodylate adduct at the catalytic cysteine was observed when the structure was determined as detailed in section 4.3.6.
[b]WT-Cac crystal was obtained from the WT NitN protein that was crystallized in a cacodylate-containing reservoir buffer.
[c]E61Q/C165A-Apo crystal was crystallized from E61Q-C165A double mutant protein pre-incubated with ACE but did not have substrate molecules in the active site when the structure was determined as detailed in section 4.3.10.1.

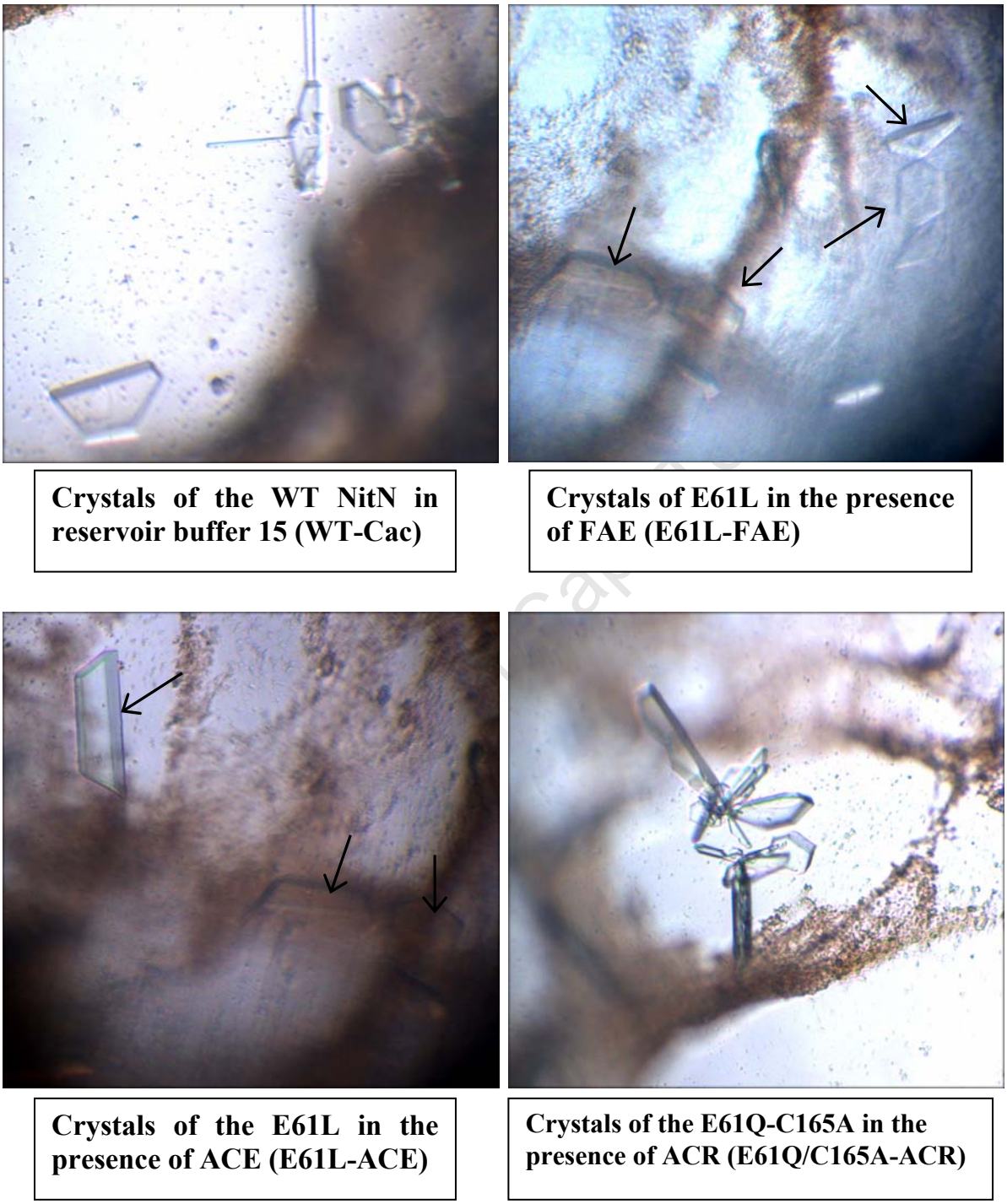
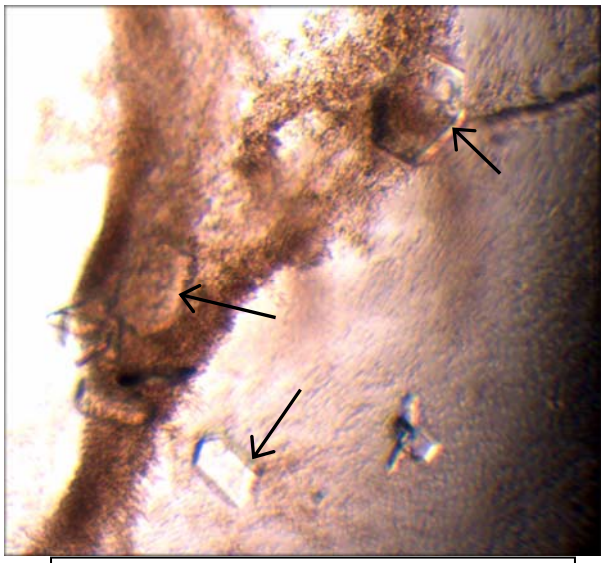
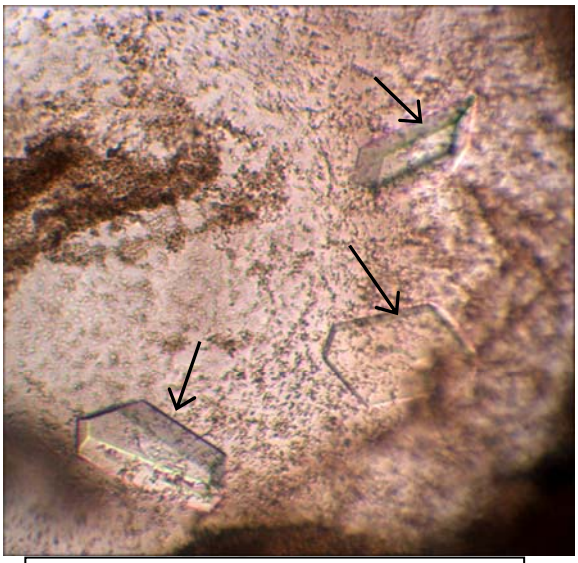


Fig. 4.10 (A): Images of the crystals for some of the structures reported in this chapter. Black arrows are used to point at some of the crystals that were covered by heavy protein precipitate in the crystallization drops.



Crystals of the unbound E61Q-C165A (E61Q/C165A-Apo)



Crystals of the E61Q in the presence of FAE (E61Q-FAE)

Crystals of the E61Q in the presence of PMD (E61Q-PMD)

Crystals of the E139Q in reservoir buffer 15 (E139Q-Cac)

Fig. 4.10 (B): Images of the crystals for some of the structures reported in this chapter. Black arrows are used to point at some of the crystals that were covered by heavy protein precipitate in the crystallization drops.

4.4.5 Structure determination, model building, refinement and validation

Diffraction data of the obtained crystals were collected at the synchrotrons (in France) and initial indexing showed that all crystals had the same space group symmetry ($C222_1$) as the WT NitN structure. The data were merged and scaled at the appropriate highest resolution that was determined based on the completeness and multiplicity of the data, as well as a signal-to-noise ($I/\sigma(I)$) ratio of about 2.0 and merging R-factor (R_{merge}) values of less than 50% in the highest (outer) resolution shell. All datasets were of high quality as indicated by the data collection statistics presented in tables 4.4 (A) and (B). The overall and high resolution shell data had completeness levels approaching 100% in most cases, indicating that almost all of theoretically expected unique reflections had been measured; the observed high redundancy values indicate that the unique reflections had been measured multiple times, which is a requirement for good-quality structures; high average ($I/\sigma(I)$) values indicate high quality data with minimal errors ; and low overall R_{merge} values indicate high accuracy in the measurement of identical reflections. Analysis of the merged data with the *Xtriage* program (Adams et al. 2010) found no possible outliers among acentric and centric reflections (Read 1999), and no ice rings-related problems were detected in the ice ring-sensitive resolution ranges. The space group and unit cell parameters were the same as those of the WT NitN structure (PDB id, 3hkx).

All the structures reported in this chapter were solved by molecular replacement using the atomic coordinates of the WT NitN protein (PDB id, 3hkx) as the search model, with clear solutions (based on cross-correlation scores) being obtained in all cases. The resultant models were improved by alternating cycles of model rebuilding and refinement using maximum likelihood methods. The refinement process was considered complete when the model errors highlighted in the σA-weighted F_{obs} - F_{calc} difference electron density maps had been addressed and when the cross-validation free R-factor (R_{free}) ceased to decrease. Reaction adducts and non-covalent substrates were built into the difference density either attached to the catalytic cysteine or in close proximity to the catalytic residues during the last cycles of refinement.

Table 4.4 (A): Data collection statistics (I)

Data collection statistics					
Dataset	**E139Q-Cac**	**WT-Cac**	**E61L-FAE**	**E61L-PMD**	**E61L-ACE**
Crystallization buffer	15	15	39	39	39
Synchrotron source	BM14[a]	BM14[a]	Proxima I[b]	Proxima I[b]	BM14[a]
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$C222_1$
Unit cell parameters	*a* = 75.960 (Å) *b* = 115.18 (Å) *c* = 64.860 (Å) *α*=*β*=*γ*= 90.0°	*a* = 75.200 (Å) *b* = 115.57 (Å) *c* = 65.510 (Å) *α*=*β*=*γ*= 90.0(°)	*a* = 75.520 (Å) *b* = 114.44 (Å) *c* = 64.820 (Å) *α*=*β*=*γ*= 90.0°	*a* = 75.360 (Å) *b* = 114.83 (Å) *c* = 65.090 (Å) *α*=*β*=*γ*= 90.0°	*a* = 75.550 (Å) *b* = 115.32 (Å) *c* = 65.320 (Å) *α*=*β*=*γ*= 90.0°
Cell content analysis					
Unit cell volume ($Å^3$)	567465.375	569310.000	560207.375	563262.063	569125.813
V_m ($Å^3$/Da)[c]	2.36	2.37	2.33	2.35	2.37
Molecules in ASU[d]	1	1	1	1	1
Solvent content (%)	48.01	48.18	47.34	47.62	48.16
Resolution range (Å)[e]	34.27 - 1.80 (1.90 - 1.80)	43.33 - 1.49 (1.54 - 1.49)	42.87 - 1.60 (1.66 - 1.60)	45.25 - 1.55 (1.61 - 1.55)	22.05 - 1.41 (1.49 - 1.41)
No. of observed reflections	198006	285535	91906	192614	409653
No. of unique reflections	26738	46854	35169	41126	55192
Completeness (%)[e]	100.0 (100.0)	99.9 (100.0)	94.0 (96.9)	99.7 (100.0)	100.0 (100.0)
Redundancy[e]	7.4 (7.4)	6.09 (6.00)	2.61 (2.59)	4.68 (4.76)	7.4 (7.4)
Mean $(I) /\sigma(I)$[e]	17.0 (4.4)	13.7 (4.3)	7.9 (3.0)	12.7 (4.0)	19.7 (5.3)
R_{merge} (%)[e, f]	8.40 (49.50)	5.80 (31.7)	7.60 (33.5)	5.80 (30.2)	5.70 (38.5)

[a]BM14 - Macromolecular beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble.
[b]Proxima I - Macromolecular beamline at the Synchrotron Soleil, Paris.
[c]V_m is the Matthew's coefficient, which allows estimation of the number of protein subunits in the asymmetric unit. $V_m = V/(MW*n)$, where V is the volume of the unit cell, MW is the molecular mass of the molecule and n is the number of asymmetric units in the unit cell.
[d]ASU - asymmetric unit.
[e]Values in parentheses are for the highest resolution shell.
[f]$R_{merge} = [\Sigma_h \Sigma_j |I_j(h) - \langle I(h) \rangle|] / \Sigma_h \Sigma_j |I_j(h)|$, where $I_j(h)$ and $\langle I(h) \rangle$ are the jth and the mean measurements of the intensity of reflection h, respectively.

Table 4.4 (B): Data collection statistics (II)

Data collection statistics Dataset	**E61Q/C165A-Apo**	**E61Q/C165A-ACR**	**E61Q-FAE**	**E61Q-ACE**	**E61Q-PMD**
Crystallization buffer	39	32	4	39	39
Synchrotron source	BM14[a]	BM14[a]	BM14[a]	BM14[a]	BM14[a]
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$C222_1$
Unit cell parameters	*a* = 75.930 (Å) *b* = 115.27 (Å) *c* = 65.280 (Å) *α*=*β*=*γ*= 90.0°	*a* = 75.510 (Å) *b* = 114.10 (Å) *c* = 65.610 (Å) *α*=*β*=*γ*= 90.0°	*a* = 76.640 (Å) *b* = 115.52 (Å) *c* = 64.970 (Å) *α*=*β*=*γ*= 90.0°	*a* = 75.020 (Å) *b* = 115.52 (Å) *c* = 65.440 (Å) *α*=*β*=*γ*= 90.0°	*a* = 75.220 (Å) *b* = 115.35 (Å) *c* = 65.510 (Å) *α*=*β*=*γ*= 90.0°
Cell content analysis					
Unit cell volume ($Å^3$)	571353.688	569759.313	575184.875	567185.250	568397.063
V_m ($Å^3$/Da)[b]	2.38	2.37	2.40	2.36	2.37
Molecules in ASU[c]	1	1	1	1	1
Solvent content (%)	48.37	48.22	48.71	47.99	48.10
Resolution range (Å)[d]	32.61 - 1.80 (1.90 - 1.80)	30.11 - 1.49 (1.53 - 1.49)	32.47- 1.92 (2.02 - 1.92)	20.92 - 1.33 (1.40 - 1.33)	20.93 - 1.35 (1.42 - 1.35)
No. of observed reflections	199799	353974	164280	485044	460485
No. of unique reflections	26903	42947	22375	65436	62789
Completeness (%)[d]	100.0 (100.0)	92.0 (90.0)	100.0 (100.0)	99.9 (99.7)	99.9 (99.8)
Redundancy[d]	7.4 (7.5)	6.5 (5.4)	7.3 (7.4)	7.4 (7.4)	7.3 (7.2)
Mean $(I)/\sigma(I)$[d]	15.8 (5.4)	27.9 (2.3)	13.0 (4.8)	18.0 (5.4)	18.7 (5.1)
R_{merge} (%)[d, e]	9.1 (38.7)	8.6 (52.1)	10.8 (41.3)	6.30 (37.5)	5.70 (38.0)

[a]BM14 - Macromolecular beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble.

[b]V_m is the Matthew's coefficient, which allows estimation of the number of protein subunits in the asymmetric unit. $V_m = V/(MW*n)$, where V is the volume of the unit cell, MW is the molecular mass of the molecule and n is the number of asymmetric units in the unit cell.

[c]ASU - asymmetric unit.

[d]Values in parentheses are for the highest resolution shell.

[e]$R_{merge} = [\Sigma_h \Sigma_j |I_j(h) - \langle I(h) \rangle|] / \Sigma_h \Sigma_j |I_j(h)|$, where $I_j(h)$ and $\langle I(h) \rangle$ are the jth and the mean measurements of the intensity of reflection h, respectively.

The resultant structures were all of high quality as depicted by various statistical parameters in the refinement and model validation statistics in tables 4.5 (A) and (B). The final crystallographic R-factor values are what would be expected for data of the observed resolution range. R_{cryst} values of less than 20% in all cases indicate a good agreement between the final model and the experimental data, and small differences between R_{cryst} and R_{free} values (less than 5% in all cases) are an indication of correct modeling, with very minimal over-fitting of the model to the data. The Ramachandran plots (Ramachandran et al. 1963) from both *RAMPAGE* (Lovell et al. 2003) and *MolProbity* (Davis et al. 2007) had over 99.5% of all residues falling in the favoured and allowed regions, an indication of correctness in the backbone geometry of all structures. Figure of merit values of greater than 0.9 in most cases indicate high accuracy of the final structures. The models had reasonable bond lengths and bond angles, as judged from the root mean square deviation (rmsd) values that are within the acceptable range for well-defined structures of similar resolution. The atomic clashes indicated by the *MolProbity* all-atom clash score involve the hydrogen atoms that were added by the program during the validation process in all cases. Ramachandran plots for all residues case in the final structures reported in this chapter are presented in appendix III.

Residues 1-11 of the hexahistidine tag and part of the associated linker region were missing in all structures reported in this chapter. There was no visible density for residues 259-268 in all cases. Surface glutamate, glutamine and arginine residues had high B-factors; with some of them having alternative conformations. Leu153 and a few arginine residues were found to have poor rotamers in most cases.

Table 4.5 (A): Model refinement and validation statistics (I)

Dataset	**E139Q-Cac**	**WT-Cac**	**E61L-FAE**	**E61L-PMD**	**E61L-ACE**
Refinement statistics					
Resolution range	32.43- 1.80	37.60- 1.49	34.05 - 1.60	31.50- 1.55	22.02- 1.41
No. of reflections in working set (95%)	24027	42115	33210	37073	49593
No. of reflections in test set (5%)	1345	2367	1845	2062	2809
No. of atoms					
Protein[a]	1979	1986	1996	1996	2068
Ligand/ion	22	4	29	12	4
Water	182	288	155	255	274
Missing residues	A1-12 A259-267	A1-12 A16-18 A259-267	A1-12 A259-267	A1-12 A258-267	A1-12 A258-267
R_{cryst} (%)[b]	15.92	13.95	12.72	12.85	13.09
R_{free} (%)[c]	19.25	17.33	17.26	16.41	15.67
Overall figure of merit	0.893	0.906	0.917	0.930	0.928
Average B value (Å^2)	20.03	17.47	17.30	17.92	16.95
Model validation					
Rms[d] deviations from ideality					
Bond lengths (Å)	0.03	0.03	0.03	0.03	0.03
Bond angles (°)	2.50	2.18	2.22	2.34	2.49
Bad rotamers (%)	0.50	0.50	0.50	0.50	1.50
Ramachandran outliers (%)	0.00	0.00	0.00	0.00	0.40
All-atom clashscore	6.61	4.41	5.34	5.12	4.65

[a]Number of protein atoms include duplicated atoms in alternate conformations.
[b]$R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$ where F_o and F_c are observed and calculated structure factor amplitudes of reflection h of the working set of reflections, respectively.
[c]R_{free} is equal to R_{cryst} for h belonging to the test set of reflections.
[d]Rms - root mean square.

Table 4.5 (B): Model refinement and validation statistics (II)

Dataset	**E61Q/C165A-Apo**	**E61Q/C165A-ACR**	**E61Q-FAE**	**E61Q-ACE**	**E61Q-PMD**
Refinement statistics					
Resolution range	31.68 - 1.80	30.31 - 1.50	32.39 - 1.92	20.82 - 1.33	20.89 - 1.35
No. of reflections in working set (95%)	24280	40619	20102	58824	45779
No. of reflections in test set (5%)	1341	2326	1134	3274	2524
No. of atoms					
Protein[a]	1991	2009	2026	2033	-
Ligand/ion	7	15	14	9	9
Water	216	270	200	301	235
Missing residues	A1-12 A258-267	A1-12 A259-267	A1-11 A17-18 A258-267	A1-12 A258-267	A1-12 A16-18 A260-267
R_{cryst} (%)[b]	15.14	15.38	14.72	12.74	13.56
R_{free} (%)[c]	17.99	18.64	18.18	15.86	17.43
Overall figure of merit	0.893	0.883	0.890	0.934	0.892
Average B value (Å^2)	16.68	19.42	18.57	14.71	10.83
Model validation					
Rms[d] deviations from ideality					
Bond lengths (Å)	0.03	0.02	0.03	0.03	0.02
Bond angles (°)	2.41	1.68	2.27	2.44	1.89
Bad rotamers (%)	1.50	1.00	0.00	0.50	0.50
Ramachandran outliers (%)	0.40	0.00	0.00	0.00	0.00
All-atom clashscore	4.11	7.67	3.33	2.32	4.38

[a]Number of protein atoms include duplicated atoms in alternate conformations.
[b]$R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$ where F_o and F_c are observed and calculated structure factor amplitudes of reflection h of the working set of reflections, respectively.
[c]R_{free} is equal to R_{cryst} for h belonging to the test set of reflections.
[d]Rms - root mean square.

4.4.6 Evidence for structural disorder on Glu139 mutation

4.4.6.1 The E139Q NitN mutant structure (E139Q-Cac)

Even though no reaction was detected by mass spectrometry between E139Q NitN and the four amide substrates, crystallization of this mutant in the presence of substrates was attempted. The E139Q mutant was prone to precipitation during crystallization even at low protein concentrations, with small crystals only growing in reservoir buffer 15 (see table 4.3 (A)) that contained sodium cacodylate as the buffering agent. A number of crystals were obtained, but the resultant structures were all the same regardless of which substrate was present during crystallization, hence only one E139Q structure (E139Q-Cac) was refined to completion.

Glu139 is located in a small helix that is in the middle of a long flexible loop (residues 130 - 152) that connects β-stand 4 to β-strand 5 in the NitN monomeric structure (Fig. 4.11). Key active site residues including Lys131, Tyr135 and Glu139 are located in this loop, indicating its importance in the formation of the active site pocket. Glu139 is involved in a hydrogen bonding network that maintains the configuration of the active site, ensuring its stability as well as that of the entire 3D structure. In the branch 2 aliphatic amidases, an equivalent of this flexible loop is also involved in the formation of the active sites in the hexameric quaternary forms, while in branch 1 nitrilases, a similar loop forms part of the active site and also contributes to an interface (the C-surface; (Sewell et al. 2005) that is involved in the assembly of helical oligomeric nitrilase structures.

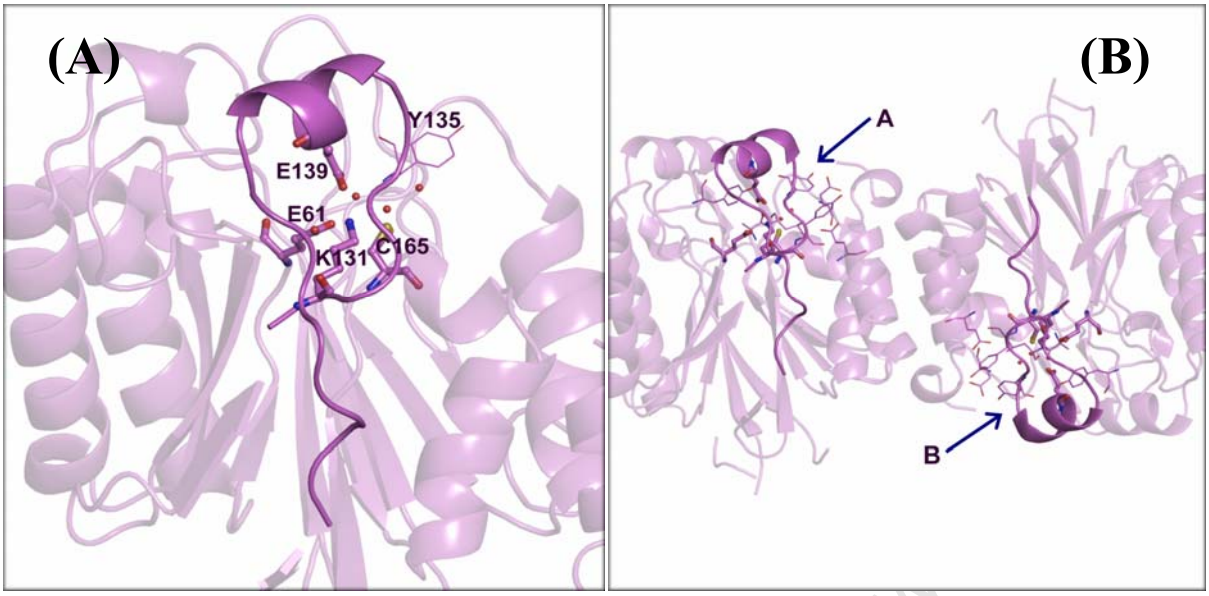


Fig. 4.11: Cartoon representation of the loop structure where Glu139 is located in the NitN amidase model. **(A)** A monomer of NitN showing the flexible loop with a small helix in the middle (non-transparent; residues 130 -152) where Glu139 is located. **(B)** The WT NitN dimeric structure showing the loop bearing Glu139 as non-transparent cartoon and the arrow pointing into the two active site pockets (rendered as lines). The model is viewed down the dimeric two fold axis.

In the E139Q-Cac structure, a portion of the long loop between residues 136 and 143 was found to be slightly disordered as evidenced by noisy difference density in the region (Fig. 4.12 (A)). Attempts to rebuild this loop in *PHENIX* (Adams et al. 2010) did not improve the fit or eliminate the noisy density around the disordered region. A comparison of the refined E139Q-Cac structure with that of the WT NitN enzyme showed that some of the side chains from the E139Q mutant had shifted considerably in the disordered region, and the secondary structure was slightly disturbed resulting in a shorter helix (Fig. 4.12 (B)). The introduction of the E139Q mutation clearly destabilizes part of the loop surrounding the point of mutation, which implies that the active site pocket is slightly disordered. This may explain the instability of the E139Q mutant protein observed during expression, purification and crystallization. Apart from the disordered loop region, the rest of the E139Q structure is highly conserved relative to the wild-type structure (Fig. 4.12 (C)), with a root mean square deviation (rmsd) of 0.57 Å over the Cα atoms between the two structures.

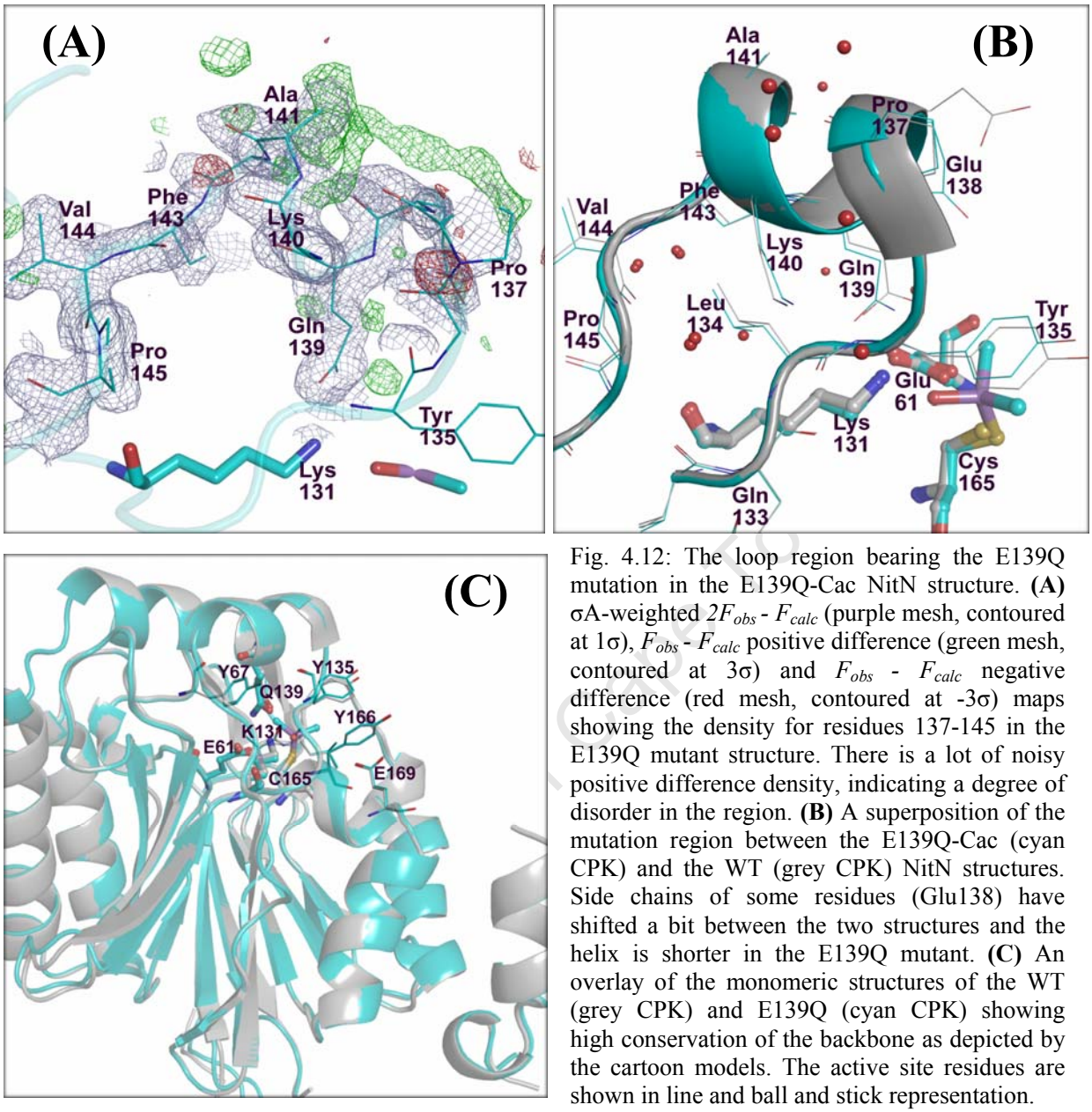


Fig. 4.12: The loop region bearing the E139Q mutation in the E139Q-Cac NitN structure. **(A)** σA-weighted $2F_{obs}$ - F_{calc} (purple mesh, contoured at 1σ), F_{obs} - F_{calc} positive difference (green mesh, contoured at 3σ) and F_{obs} - F_{calc} negative difference (red mesh, contoured at -3σ) maps showing the density for residues 137-145 in the E139Q mutant structure. There is a lot of noisy positive difference density, indicating a degree of disorder in the region. **(B)** A superposition of the mutation region between the E139Q-Cac (cyan CPK) and the WT (grey CPK) NitN structures. Side chains of some residues (Glu138) have shifted a bit between the two structures and the helix is shorter in the E139Q mutant. **(C)** An overlay of the monomeric structures of the WT (grey CPK) and E139Q (cyan CPK) showing high conservation of the backbone as depicted by the cartoon models. The active site residues are shown in line and ball and stick representation.

The catalytic cysteine (Cys165) of the E139Q-Cac structure had very clear density attached to the Sγ atom (Fig. 4.13 (A)) that indicated covalent modification, and which persisted throughout the refinement process. Careful examination of this density showed that it was not from a reaction intermediate from a substrate but it resembled dimethylarsenate (cacodylate) from the crystallization buffer (buffer 15; see table 4.3(A)). Crystallization of the WT enzyme in the same buffer resulted in similar density being attached to the catalytic cysteine (Fig. 4.13 (B)).

Modification of the reactive cysteine by cacodylate is a common reaction. The catalytic cysteine of the nitrilase from *Xanthomonas campestris* reacted with the arsenate of cacodylate from the crystallization buffer to result in a stable Cys-cacodylate adduct that resembles the proposed transition state for the breakdown of the acyl-enzyme intermediate (Chin et al. 2007) in the nitrilase and amidase catalyzed reactions.

Given the instability of Glu139 mutants as observed during protein expression and purification, the cysteine-dimethylarsenate covalent adduct may be responsible for conferring structural stability to the E139Q mutant that permitted crystal packing. Indeed, the oxygen of the cacodylate adduct makes hydrogen bonds to the amino group of the catalytic lysine (Lys131) and the backbone NH group of Tyr166, which form an oxyanion hole (Fig. 4.13 (C)). However, even with the cacodylate in the active site, the mutation-bearing region of the loop (residues 134 - 144) is still relatively disordered as evidenced by noisy density shown in figure 4.12 (A) above. The active site stabilization role of the Cys165-cacodylate adduct explains why the E139Q mutant could not crystallize in reservoir buffers that lacked cacodylate. Although the active site configuration is evidently disturbed by the E139Q mutation, the possibility of the binding pocket being inaccessible to the amide substrates is ruled out by the mass spectrometric data, which provide satisfactory evidence that acrylamide (ACR) is able to bind to the E139Q active site pocket and to be covalently attached to the catalytic cysteine in a Michael-like addition reaction (described in chapter 5).

A comparison of the E139Q active site with that of the unbound WT (WT-Apo) reveals differences in the hydrogen bonding network between the two structures (Fig. 4.13 (C)), which may explain the observed instability in the E139Q structure. While numerous polar contacts exist between active site residues and water molecules in the WT-Apo active site, fewer hydrogen bonds are observed in the cacodylate-bound E139Q (E139Q-Cac) structure. Specifically, the replacement of Glu139 with glutamine in the E139Q structure results in the loss of two active site-stabilizing contacts (between the carboxyl oxygen of Glu139 and the amino group of Lys131 and the hydroxyl of Tyr67; see figures 4.13 (C) and (D)). These hydrogen bonds are present in the cacodylate-complexed WT (WT-Cac) structure (Fig. 4.13 (D)), and may explain why the mutation-bearing loop is still relatively disordered in the E139Q-Cac structure.

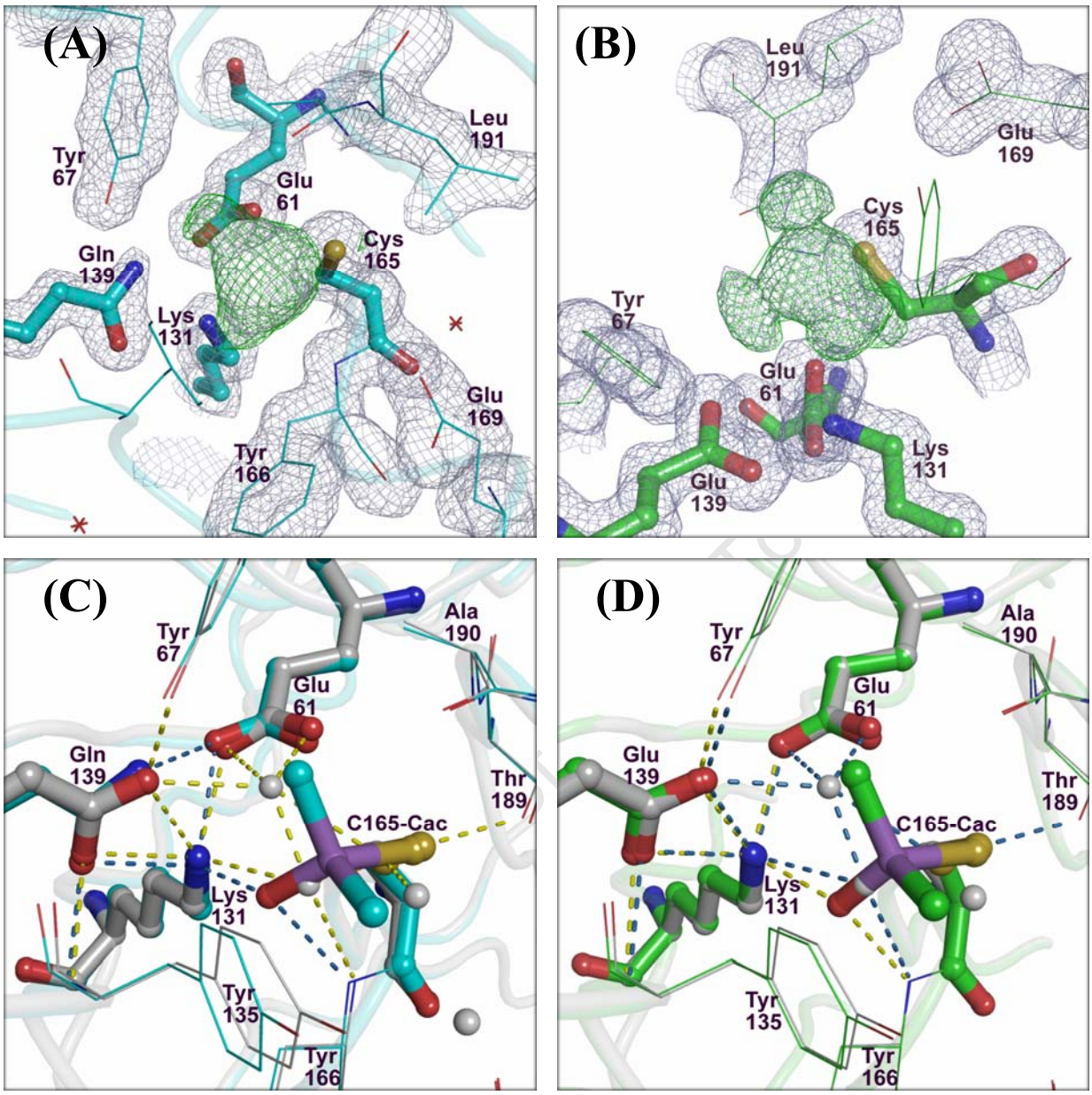


Fig. 4.13: Covalent modification of the catalytic cysteine with cacodylate in the E139Q and WT NitN structures. **(A)** σA-weighted $2F_{obs}$ - F_{calc} (grey mesh, contoured at 1σ) and F_{obs} - F_{calc} positive difference (green mesh, contoured at 3σ) maps showing density for the omitted cacodylate attached to Cys165 in the E139Q-Cac structure. **(B)** σA-weighted $2F_{obs}$ - F_{calc} (grey mesh, contoured at 1σ) and F_{obs} - F_{calc} positive difference (green mesh, contoured at 3σ) maps showing density for the omitted cacodylate attached to Cys165 in the WT-Cac structure. **(C)** A comparison of the hydrogen bonding network in the active sites of E139Q-Cac (cyan CPK) and WT-Apo (grey CPK) structures. Yellow dashes show potential contacts in the WT-Apo while blue dashes are hydrogen bonds in the E139Q-Cac active site. Active site waters of the WT-Apo structure are shown as grey spheres. **(D)** Superposition of the active sites of the cacodylate-modified WT structure (WT-Cac; green CPK) and that of the WT-Apo enzyme (grey CPK). Yellow dashes are potential polar contacts in the WT-Cac while blue dashes are contacts in the WT-Apo active site.

4.4.7 Structural evidence for a trapped thioester intermediate on Glu61 mutation

4.4.7.1 The structure of E61L with propionamide (E61L-PMD)

Although the MALDI-TOF mass spectrum of E61L NitN reacted with PMD had a dominant fragment peak corresponding to the free protein (unbound cysteine; *m/z* 2041), a weak signal peak at *m/z* 2097 was also observed (Appendix IV), which most likely represents a small percentage of PMD thioester acyl-enzyme intermediate. Determination of the structure of E61L mutant protein reacted with PMD (E61L-PMD) resulted in electron density which clearly indicated that the catalytic cysteine (Cys165) adopted two distinct conformations, with one of the conformers being covalently modified at the thiol group. The difference density attached to one of the Cys165 conformers (Fig. 4.14 (A)) resembled a PMD acyl moiety, with a high possibility of including another modification state or water molecules since the PMD thioester intermediate only constituted a small percentage of the E61L protein reacted with PMD, according to the mass spectrometric (MS) data. The best fit was obtained when a PMD acyl moiety attached to one cysteine conformer at 0.3 atomic occupancy was modeled in combination with two waters that were interpreted to be closely associated with the free cysteine conformer (Fig. 4.14 (B)). While the distance between the two water molecules (labeled watA and watB in figure 4.14 (B)) is a little shorter than a normal hydrogen bond length, water molecules in similar locations have been observed in close association with the Cys165 thiol group in the WT Nit crystal structure. The protein model presented here represents the two protein species in the crystal; the PMD thioester intermediate and the free enzyme, as depicted by the MS data (Appendix IV).

The PMD acyl-Cys165 intermediate has the carbonyl oxygen of PMD interacting with the side chain amino group of Lys131 and the backbone NH group of Tyr166, which form the oxyanion hole, while the rest of the acyl moiety docks into the active site pocket where it participates in hydrophobic contacts probably with Leu61 and Tyr135 (Fig. 4.14 (C)). The water molecule in position B (WatB) coincides with the carbonyl oxygen of the PMD in the acyl-Cys165 intermediate, and it interacts with the SH group of the free Cys165 conformer in addition to hydrogen bonding interactions with the oxyanion hole. Position A water molecule (WatA) interacts with the free thiol group and is also coordinated between the OE_1 atom of Glu139 and the backbone carboxylate oxygen of Ala190 as shown in figure 4.14 (C)). The

polar interactions involving the two water molecules (WatA and WatB) probably enhance stability of the active site in the E61L free protein species, while the oxyanion hole contacts and hydrophobic contacts of the PMD thioester ensure stability in the acyl-enzyme intermediate protein proportion of the crystal.

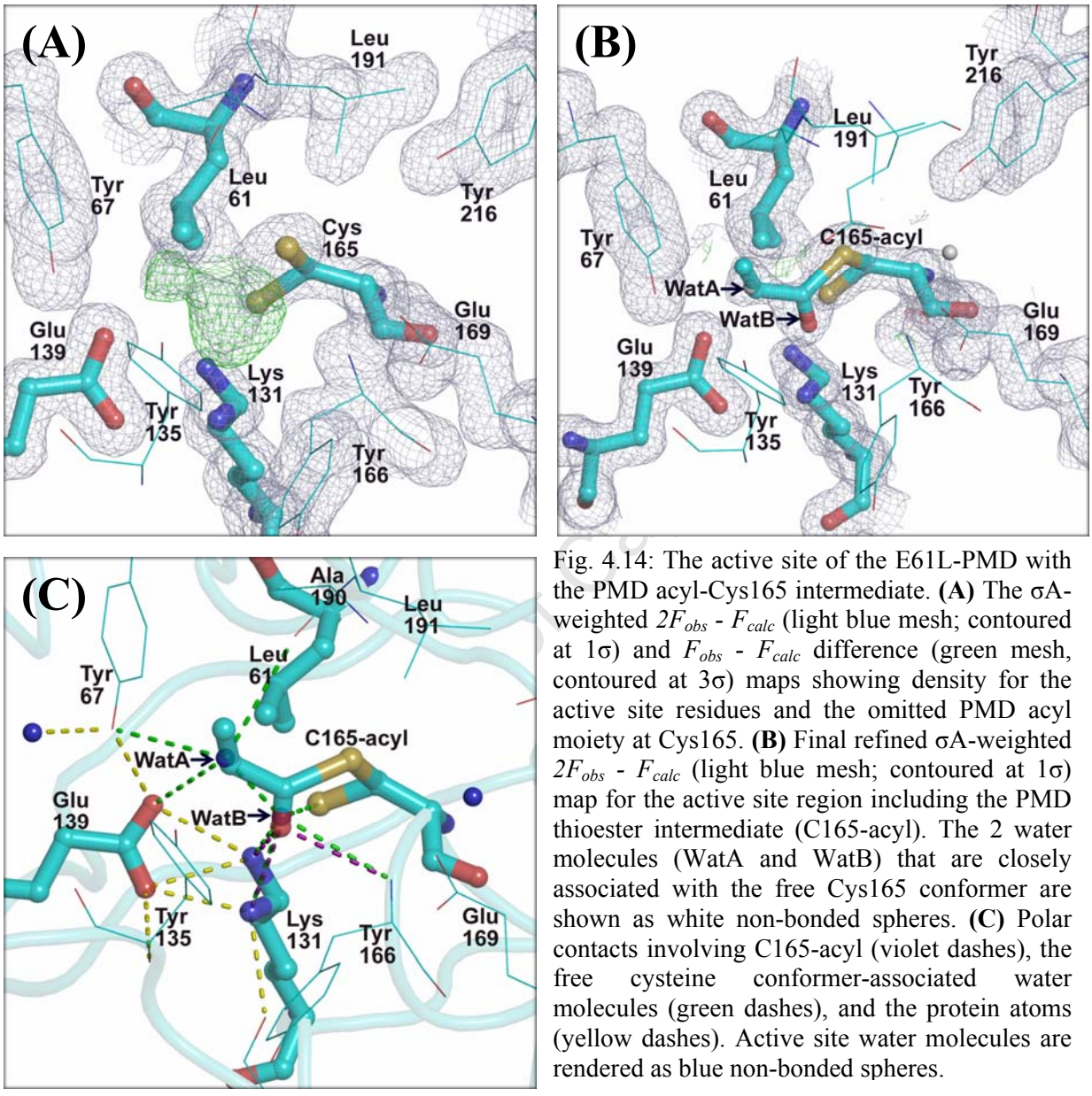


Fig. 4.14: The active site of the E61L-PMD with the PMD acyl-Cys165 intermediate. **(A)** The σA-weighted $2F_{obs}$ - F_{calc} (light blue mesh; contoured at 1σ) and F_{obs} - F_{calc} difference (green mesh, contoured at 3σ) maps showing density for the active site residues and the omitted PMD acyl moiety at Cys165. **(B)** Final refined σA-weighted $2F_{obs}$ - F_{calc} (light blue mesh; contoured at 1σ) map for the active site region including the PMD thioester intermediate (C165-acyl). The 2 water molecules (WatA and WatB) that are closely associated with the free Cys165 conformer are shown as white non-bonded spheres. **(C)** Polar contacts involving C165-acyl (violet dashes), the free cysteine conformer-associated water molecules (green dashes), and the protein atoms (yellow dashes). Active site water molecules are rendered as blue non-bonded spheres.

The location of the thioester carbonyl oxygen in the oxyanion hole is consistent with the suggested role of the catalytic lysine in promoting nucleophilic attack (Andrade et al. 2007). The positive character of the lysine amino group would polarize the electrons at the carbonyl group double bond thus creating a better electrophile at the carbonyl carbon and facilitating attack by the water substrate.

4.4.8 Structural evidence for S_N2 reaction with FAE on Glu61 mutation

4.4.8.1 The structure of the E61L mutant with FAE (E61L-FAE)

While the E61Q mutant reacted with the α-carbon of FAE to yield a FAE-Cys165 S_N2 reaction adduct (2098 Da fragment peak; Fig. 4.7 (B)), the mass spectrum of the E61L NitN reacted with FAE did not have a peak corresponding to the FAE-Cys165 S_N2 reaction adduct even after several repeats; instead, a FAE thioester acyl-enzyme intermediate peak (*m/z* 2101; Fig. 4.7 (A)) was observed with the E61L mutant. Crystallization of the E61L NitN protein pre-incubated with FAE was difficult to achieve as the reacted protein was found to be highly unstable and precipitated extensively before and during crystallization set-ups. A few crystals however appeared in heavy protein precipitate (Fig. 4.10 (A)) after 3-4 weeks of incubation. Upon collection of diffraction data and structure solution, clear continuous density was found attached to the catalytic cysteine (Fig. 4.15 (A) and (B)), indicating covalent modification. The extra density however appeared to resemble an S_N2 reaction covalent adduct rather than the FAE acyl-Cys intermediate observed by MS. While this was surprising, it implied that similar to E61Q mutant, E61L NitN was also able to carry out an S_N2 substitution attack on the α-carbon atom of FAE, eliminating fluorine and resulting in the formation of an adduct at the catalytic cysteine. But the question is, why was the S_N2 reaction not detected by MS? The reason could be one of the two possibilities: (1) if the S_N2 reaction between E61L and FAE had occurred at the same rate as with the E61Q mutant, then the corresponding tryptic fragment peak may have failed to desorb (fly) off the matrix; or (2) the S_N2 reaction may not be favorable in the E61L mutant, probably due to substrate positioning factors, hence the reaction may occur very slowly such that the two hours incubation period is not sufficient to have detectable amounts of the adduct formed. Although the first option is not entirely impossible, MS experiments were repeated a couple of times and in all cases, S_N2 reaction was not detected between E61L mutant and FAE, and hence the second option is more likely to be true.

Why then was the FAE thioester (acyl-enzyme) intermediate not present in the structure and why did S_N2 reaction occur in the crystal? There could be three possible explanations for this:

(1) **Stability of the FAE acyl-enzyme intermediate** - the unbound (apo) E61L mutant is relatively unstable compared to the WT NitN enzyme and the mutant protein at crystallization concentrations (4-6 mg/ml) was found to be highly prone to

precipitation. The solubility problem was further compounded by the possibility that the acyl-enzyme intermediate is likely to be unstable, as the E61L mutant protein pre-incubated with FAE was found to precipitate heavily before and during crystallization (See heavy protein precipitation in the crystallization drop of E61L-FAE in figure 4.10 (A)) . It is therefore likely that the acyl-complexed mutant protein had precipitated out, explaining why it was not incorporated in the crystal. MS experiments were carried out at 1 mg/ml protein, a concentration that is probably low enough for the unstable mutant proteins to remain in solution.

(2) **S_N2 reaction of FAE with the unreacted (free) E61L protein proportion during crystallization** - In all mass spectrometric experiments where reaction adducts were observed, a fragment peak for unreacted protein (2041 Da) was always present even after prolonged reaction incubation periods. With favorable conditions, the unreacted protein is likely to undergo further reactions during crystallization. E61L-FAE protein crystallized in reservoir buffer 39 (Table 4.3 (A)) that is buffered at pH 7.5. This is the optimal pH for NitN activity and is therefore conducive for further reactions during crystallization. The S_N2 reaction adduct observed in the E61L-FAE crystal structure was concluded to have occurred during crystallization with the unreacted fraction of the mutant protein. But if the E61L protein is capable of carrying out an S_N2 displacement of fluorine in FAE, why were the adducts not detected by MS?

(3) **The S_N2 reaction happened over time** - As mentioned previously, it is highly likely that S_N2 substitution attack of FAE is a very slow and an unfavorable reaction in the E61L mutant. But over time, the FAE adduct-complexed mutant protein is likely to accumulate, and if it's stable enough and able to achieve supersaturation, then crystals of the adducted protein would grow. Indeed, small E61L-FAE crystals only appeared after about 3-4 weeks following set-up of crystallization condition trials, and continued to grow for a while after that. It seems that the two hour pre-incubation period during mass spectrometric assays is not sufficient time to form detectable amounts of the S_N2 adducts.

In the E61L-FAE NitN structure, the FAE S_N2 reaction adduct to the catalytic cysteine is stabilized by various residues of the active site pocket (Fig. 4.15 (C)). The carbonyl oxygen atom of the FAE adduct interacts with both the amino group of Lys131 and the backbone NH group of Tyr166 in the oxyanion hole, while the amide nitrogen interacts with the OE_1 atom of Glu139 and the OH group of Tyr67 (Fig. 4.15 (C)). In addition to the observed adduct, a

non-covalently bound FAE molecule located close to the mouth of the active site pocket has its amide nitrogen hydrogen bonding to the OE_1 atom of Glu169, with the rest of the molecule being stabilized by hydrophobic contacts with various atoms of Tyr135.

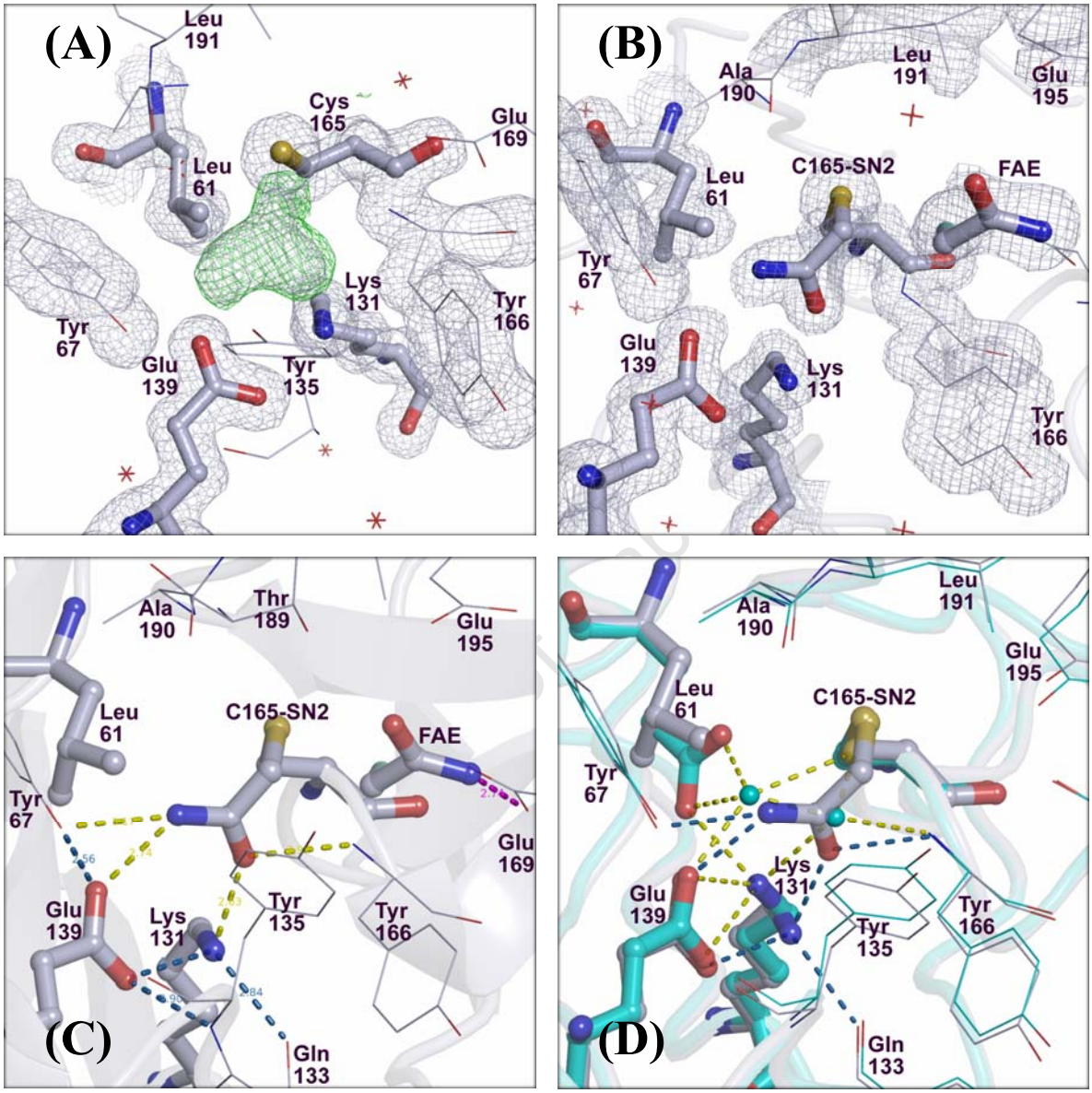


Fig. 4.15: FAE S_N2 reaction adduct to the catalytic cysteine in the E61L-FAE structure. **(A)** σA-weighted $2F_{obs}$ - F_{calc} (grey mesh, contoured at 1σ) and F_{obs} - F_{calc} positive difference (green mesh, contoured at 3σ) maps showing density for the omitted S_N2 reaction adduct of FAE attached to Cys165 as well that of the active site residues in the E61L-FAE structure. **(B)** σA-weighted $2F_{obs}$ - F_{calc} (grey mesh, contoured at 1σ) map showing the final refined density for the active site region of the E61L-FAE structure. There is continuous density for the FAE S_N2 reaction adduct on Cys165, labeled as C165- S_N2 in the picture. **(C)** Hydrogen bonding network in the E61L-FAE active site. Yellow dashes show the interactions between the FAE C165-S_N2 adduct and the active site residues while blue dashes are for the potential hydrogen bonds involving catalytic residues. **(D)** A superposition of the E61L-FAE active site (light purple CPK) with that of the unbound WT NitN structure (cyan CPK). Yellow dashes indicate potential interactions in the WT active site while blue dashes are for contacts in the E61L-FAE structure.

When comparing the E61L-FAE structure with that of the unbound WT NitN, differences in the active site hydrogen bonding network can be observed (Fig. 4.15 (D)). Due to the change of Glu61 to a leucine, a hydrogen bond that is involved in maintaining the orientation of the catalytic lysine side chain is lost. The amino group of the Lys131 side chain in the mutant enzyme is in a different conformation that allows it to form a stabilizing hydrogen bond to the carbonyl oxygen of the Gln133 backbone. It can clearly be seen how the interactions between the FAE S_N2 reaction adduct at Cys165 and the catalytic residues contribute to the stabilization of the active site and hence the entire 3D structure. This may also explain why attempts to crystallize the E61L apo protein in the absence of amide substrates were unsuccessful.

Apart from some small side chain and backbone conformational changes, the overall fold and structure of E61L-FAE is preserved. The mutant structure superimposes well with the NitN WT-Apo structure with a RMSD value of 0.37 Å over Cα atoms of the overall structures and 0.25 Å for Cα atoms of the active site residues. Superimposed side chains in the active site pocket of the two structures have a RMSD value of 0.40 Å. Although leucine and glutamine are not isosteric, the active site geometry is not considerably affected by the E61L mutation (Fig. 4.15 (D)); probably due to the interactions involving the FAE S_N2 reaction adduct at Cys165.

4.4.8.2 The structure of the E61Q mutant with FAE (E61Q-FAE)

Although the E61Q mutant protein is more stable than the E61L NitN mutant, crystallization of the E61Q NitN reacted with substrates was still problematic, with protein precipitation and irregular crystals being the main challenges. But unlike the E61L mutant, unreacted E61Q (apo) protein was able to form crystals that diffracted to 2.8 Å. The structure was however not solved as the data were found to be of poor quality and low completeness.

A few diffraction-quality crystals of E61Q NitN mutant reacted with FAE appeared in thick protein precipitate after 1-2 weeks of incubation. Structure solution by molecular replacement and calculation of the first electron density map revealed clear density for a FAE S_N2 reaction adduct covalently attached to Cys165. This was in agreement with the MS data that consistently showed the occurrence of an S_N2 reaction when E61Q NitN was reacted with FAE. The density for the covalent adduct persisted throughout the model building and

refinement process. Figure 4.16 (A) shows the density for the active site region from the final electron density map of E61Q-FAE structure.

The Cys165-FAE S_N2 adduct of the E61Q-FAE structure is in the same conformation as that of the E61L-FAE structure. However, there are a few active site hydrogen bonding differences between the E61L-FAE and E61Q-FAE active sites (Fig. 4.16 (B)). Unlike Leu61, the amide oxygen of Gln61 side chain is able to hydrogen bond to the amino group of Lys131, hence maintaining Lys131 in the same orientation as the WT NitN enzyme. The adducts make similar contacts with the active site residues in the two structures, with the carbonyl oxygen atoms hydrogen bonding to the backbone NH group of Tyr166 and the amino group of Lys131, in the oxyanion hole.

While the FAE S_N2 reaction adduct at the catalytic cysteine provides key hydrogen bonds that contribute to the stability of the active site in the E61Q-FAE structure, the inability of the amide nitrogen (NE_2) of the Gln61 side chain to make polar contacts around the active site creates a deficit in the valuable stabilizing hydrogen bonds as compared to the WT enzyme (Fig. 4.16 (D)). This could explain why the FAE adduct-complexed E61Q protein was still unstable and prone to precipitation both before and during crystallization relative to the WT protein. The E61Q-FAE crystals diffracted to about 1.8 Å, which is a lower resolution than most of the structures presented in this chapter, and might be indicative of a slight crystal lattice disorder due to the instability of the adduct-bound E61Q protein.

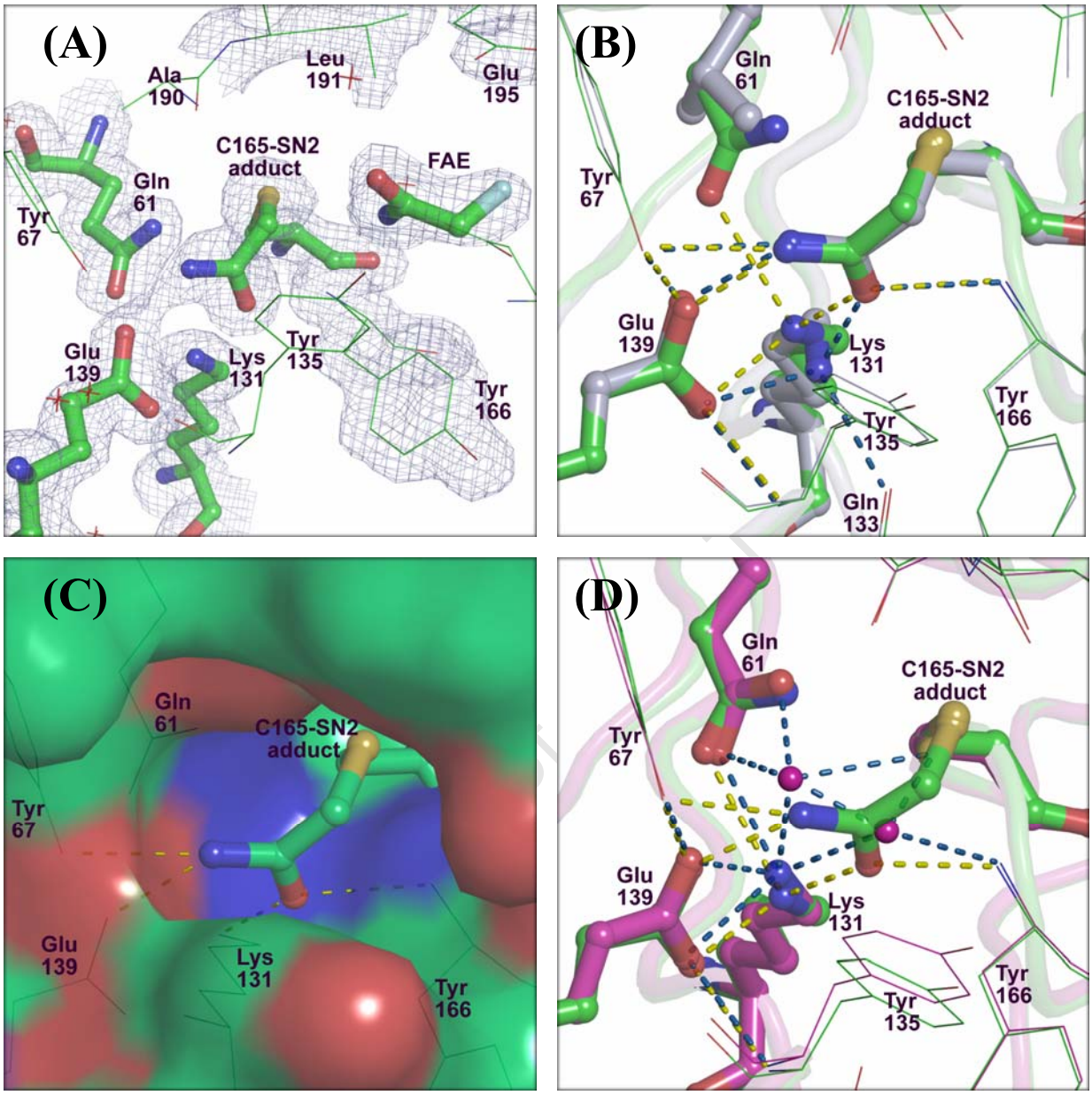


Fig. 4.16: FAE S_N2 reaction adduct to the catalytic cysteine of the E61Q-FAE and E61L-FAE structures. **(A)** Final σA-weighted $2F_{obs}$ - F_{calc} map (grey mesh, contoured at 1σ) showing density for the active site residues and the fitted FAE S_N2 reaction adduct to Cys165 (C165-S_N2) in the E61Q-FAE structure. **(B)** Superposition of the active site regions of the E61Q-FAE (green CPK) and E61L-FAE (light purple CPK) NitN structures. Possible polar contacts in the active are shown as yellow (E61Q-FAE) and blue (E61L-FAE) dashes. **(C)** Surface representation of the active site pocket of the E61Q-FAE NitN showing how the C165-S_N2 reaction adduct (ball and sticks) fits in the active site pocket. **(D)** An overlay of the WT NitN (magenta CPK) active site with that of the E61Q-FAE structure (green CPK). Yellow dashes represent polar contacts in the E61Q-FAE active site while blue dashes show hydrogen bonding network in the WT NitN active site.

4.4.9 Structural evidence for altered substrate positioning in Glu61 mutants

4.3.10.1 The E61Q/C165A-Apo and E61Q/C165A-ACR double mutant structures

Although FAE is a suitable substrate for an S_N2 substitution reaction as discussed previously (section 4.4.3.1), the WT NitN enzyme only reacts with FAE by attacking the carbonyl carbon, leading to the release of fluoroacetic acid and ammonia. On the other hand, the E61Q/L mutants are able to carry out a displacement reaction with FAE, where the fluorine atom is expelled by an attack on the α-carbon atom, and the rest of the substrate (acetamide moiety) remains covalently attached to the Sγ atom of the catalytic cysteine - what is referred to as an S_N2 substitution reaction. Why are these mutants able to attack the α-carbon instead of the carbonyl carbon?

The most reasonable explanation for the occurrence of the S_N2 substitution reaction between the E61L/Q NitN mutants and FAE is the substrate positioning alteration arising from the change of the active site environment after the removal of the carboxylate group of Glu61. To confirm this hypothesis, double mutants of Cys165 and Glu61 (E61Q/C165A and E61L/C165S) were prepared, in order to allow non-covalent binding of the amide substrates, which would give insights into the substrate orientation and positioning in the absence of Glu61. Although the E61Q-C165A double mutant enzyme was relatively unstable, a few diffraction-quality crystals were obtained and a number of crystal structures determined: two structures of the E61Q/C165A apo enzyme (E61Q/C165A-Apo) and one of the E61Q/C165A in complex with acrylamide (E61Q/C165A-ACR). The E61L/C165S mutant was highly unstable and could not pack into crystals in the presence or absence of amide substrates.

The two E61Q/C165A-Apo structures were actually solved from crystals that were grown in the presence of amide substrates, but had no substrates bound in the active site pocket. The E61Q/C165A mutant crystallized in the absence of substrates only formed thick protein precipitates with no crystals, even after prolonged incubation periods. This might therefore imply that even though the substrates did not bind in the active site, their presence in the crystallization mixture had a stabilizing effect, probably through binding to the surface of the molecules or filling up cavities in the crystal lattice. Indeed, some substrate molecules were located in the spaces between symmetry-related molecules.

The E61Q/C165A-Apo active site is stabilized by a hydrogen bonding network involving catalytic residues and water molecules. This is similar to the WT NitN apo structure, except that the water molecules in the E61Q/C165A-Apo structure are in different positions due to the two active site mutations (Fig. 4.17).

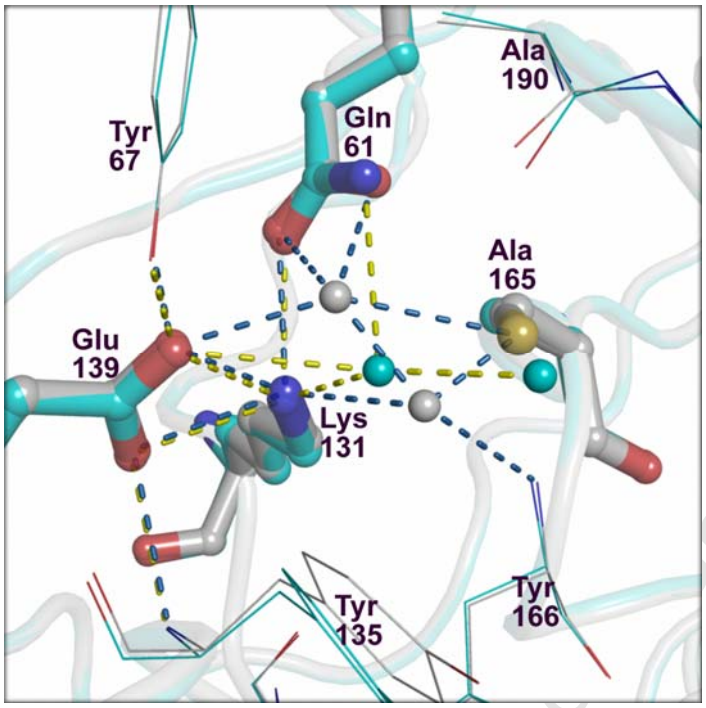


Fig. 4.17: A comparison of the stabilizing polar interactions in the active site pockets of the WT apo enzyme (grey CPK) and the E61Q/C165A-apo (cyan CPK) structure. Blue dashes and grey non-bonded spheres are potential hydrogen bonds and water molecules respectively in the WT apo active site, while yellow dashes and cyan non-bonded spheres are active site contacts and water molecules respectively in the E61Q/C165A-apo structure.

One acrylamide (ACR) molecule was fitted into clear omit difference density (Fig. 4.18 (A)) located close to the catalytic residues (catalytic sub-site) in the E61Q/C165A-ACR structure, which was crystallized in the presence of ACR. The final refined map showing ACR and the surrounding active site residues is shown in figure 4.18 (B).

Similar to butyramide (BMD) and propionamide (PMD) molecules in the C165A NitN co-crystal structures (section 3.4.6.6; figure 3.7), only the carbonyl and the amino groups of ACR make polar contacts with the protein atoms in the E61Q/C165A-ACR structure, while the side chain moiety is stabilized by hydrophobic contacts. The carbonyl oxygen atom of ACR hydrogen bonds to the amino group of Lys131 and the backbone NH group of Tyr166 in the oxyanion hole, while the amide amino group is positioned by polar contacts with the OE_1 atom of Glu139, the OH group of Tyr67 and an active site water molecule (Wat36) (Fig. 4.18 (C)). It is clear from the E61Q/C165A-ACR co-crystal structure that the change of Glu61 to glutamine has altered the way the amino group of the aliphatic amide substrates interacts with

the protein atoms. Tyr67 is not involved in the positioning of BMD and PMD and there are no water molecules in the active site of C165A-PMD and C165A-BMD co-crystal structures. Table 4.6 gives a summary of the interactions involving protein atoms and ACR and BMD in the C165A-BMD and E61Q/C165A-ACR complex structures, and highlights the differences in the binding of the amide nitrogen between the two substrates. The hydrophobic interactions between the two substrates involve different residues indicating differences in the position and orientation of the side chain moieties in Glu61 mutants.

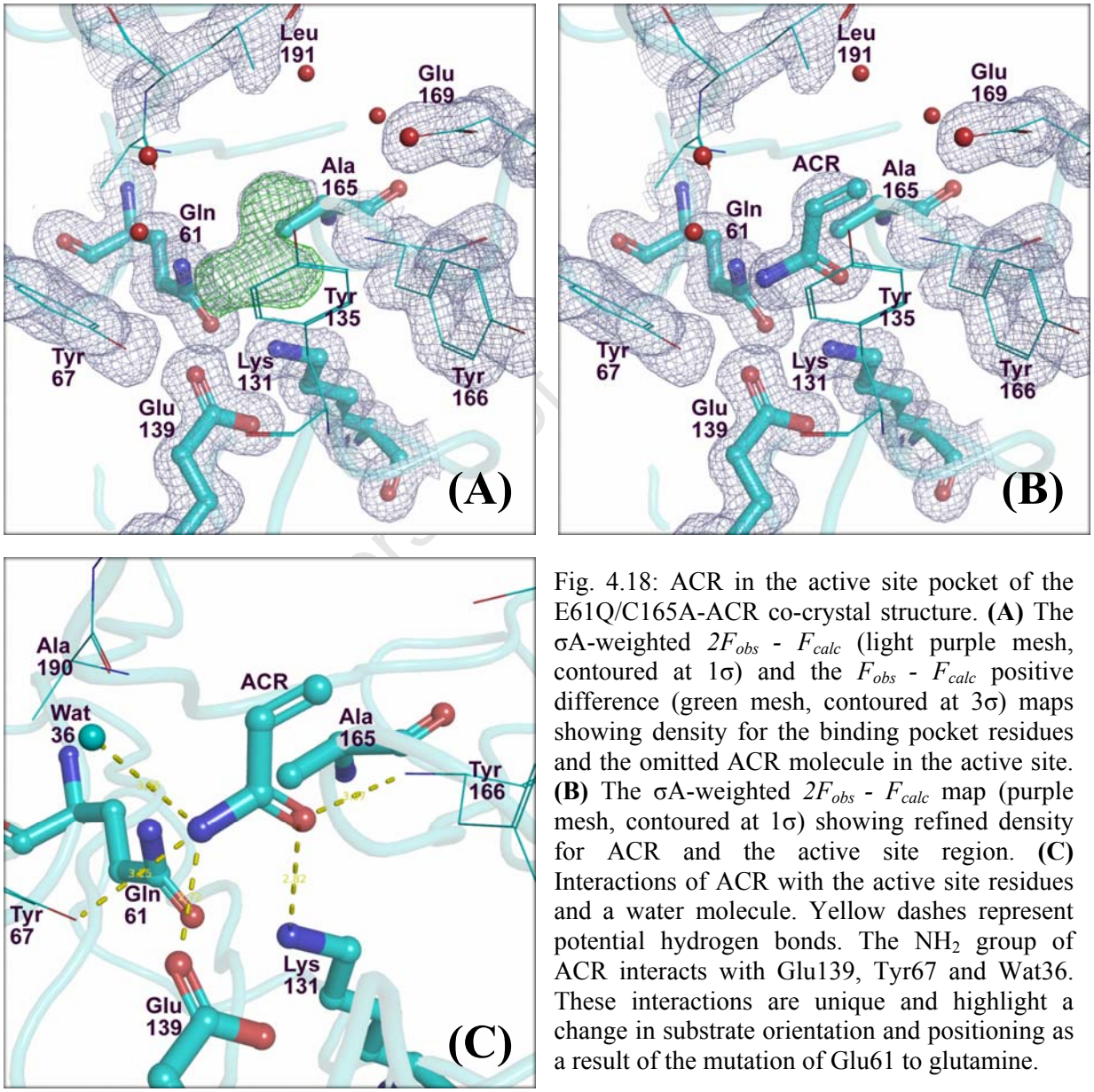


Fig. 4.18: ACR in the active site pocket of the E61Q/C165A-ACR co-crystal structure. **(A)** The σA-weighted $2F_{obs}$ - F_{calc} (light purple mesh, contoured at 1σ) and the F_{obs} - F_{calc} positive difference (green mesh, contoured at 3σ) maps showing density for the binding pocket residues and the omitted ACR molecule in the active site. **(B)** The σA-weighted $2F_{obs}$ - F_{calc} map (purple mesh, contoured at 1σ) showing refined density for ACR and the active site region. **(C)** Interactions of ACR with the active site residues and a water molecule. Yellow dashes represent potential hydrogen bonds. The NH_2 group of ACR interacts with Glu139, Tyr67 and Wat36. These interactions are unique and highlight a change in substrate orientation and positioning as a result of the mutation of Glu61 to glutamine.

Table 4.6 Polar contacts and hydrophobic interactions between ACR and BMD and the active site residues of E61Q-C165A and C165A NitN mutants respectively.

Nature of interaction	Substrate atom		Interacting residue, atom	Distance (Å)[a]	
	ACR	BMD		ACR	BMD
Hydrogen bonding	N	N	Ala190, O	-	3.11
			Glu61, OE_2	-	2.66
			Glu139, OE_1	2.72	(3.46)
			Tyr67, OH	3.25	
			Wat36	3.36	
	O	O	Tyr166, N	3.07	3.07
			Lys131, NZ	2.72	2.90
Hydrophobic contacts	-	C1	Leu191, CD_1	-	3.73
	-	Cα	Tyr135, CG	-	3.64
	-	Cα	Tyr135, CD_2	-	3.80
	C1	C1 and Cα	Tyr135, CB	3.68	3.90 and 3.76
	C1	-	Ala165, CB	3.48	-
	Cα	-	Ala165, CB	3.90	-
	Cβ	-	Tyr166, CA	3.90	-

[a]Value in parentheses represents a distance that is greater than the maximum hydrogen bond length of 3.4 Å.

Superposition of the E61Q/C165A-ACR and C165A-BMD co-crystal structures revealed differences in positioning and orientation between ACR and BMD in the active sites of the two mutant structures (Fig. 4.19 (A)). While the carbonyl oxygen atoms are in similar location, the carbonyl carbon atoms and the amino groups of the two substrates are in different positions, and the side chain moiety of ACR is tilted relative to that of BMD. This clearly demonstrates the importance of Glu61 in positioning the amide substrates in the active site of amidases. Alignment of the WT-Apo, E61Q/C165A-ACR and C165A-BMD structures, as well as fitting of FAE in place of ACR in the E61Q/C165A-ACR co-crystal structure (Fig. 4.19 (B) and (C)) provides insights into the alignment and positioning of the catalytic cysteine relative to the carbonyl carbon of the substrates and offers a satisfactory explanation as to why the catalytic cysteine of the E61L/Q mutants was able to attack at the α-carbon of FAE leading to an S_N2 substition reaction. A detailed discussion is given in the next paragraph.

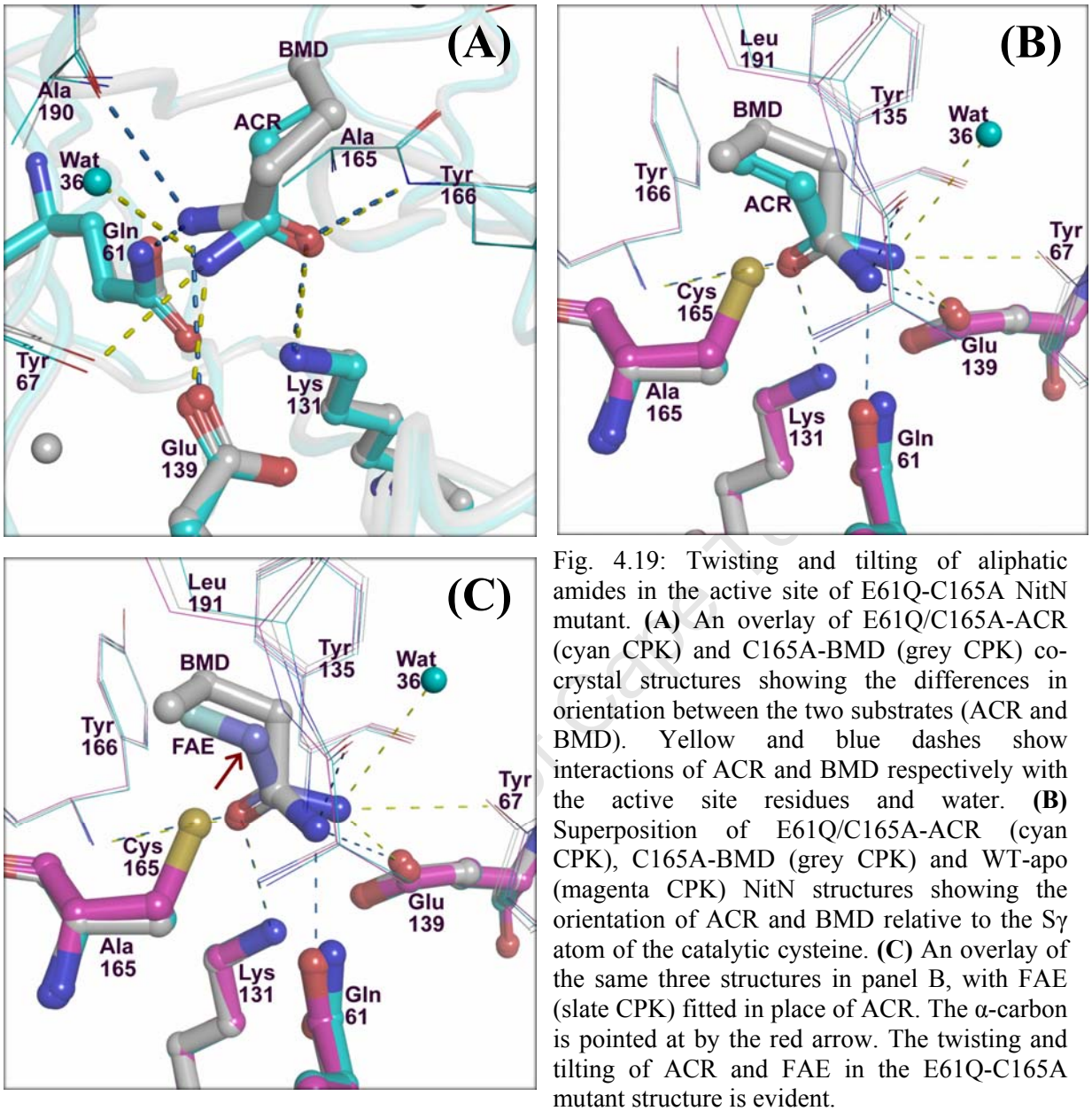


Fig. 4.19: Twisting and tilting of aliphatic amides in the active site of E61Q-C165A NitN mutant. **(A)** An overlay of E61Q/C165A-ACR (cyan CPK) and C165A-BMD (grey CPK) co-crystal structures showing the differences in orientation between the two substrates (ACR and BMD). Yellow and blue dashes show interactions of ACR and BMD respectively with the active site residues and water. **(B)** Superposition of E61Q/C165A-ACR (cyan CPK), C165A-BMD (grey CPK) and WT-apo (magenta CPK) NitN structures showing the orientation of ACR and BMD relative to the Sγ atom of the catalytic cysteine. **(C)** An overlay of the same three structures in panel B, with FAE (slate CPK) fitted in place of ACR. The α-carbon is pointed at by the red arrow. The twisting and tilting of ACR and FAE in the E61Q-C165A mutant structure is evident.

In the amide hydrolysis reaction, the formation of non-covalent Michaelis complex (Fodor et al. 2006) between the substrate and the active site residues precedes the nucleophilic attack by the cysteine, which constitutes the interaction of the highest occupied molecular orbital (HOMO) of the cysteine Sγ atom and the lowest unoccupied molecular orbital (LUMO) of the carbonyl carbon electrophile. One of the requirements for this interaction is optimal stereo-electronic alignment of the reacting orbitals, which would not only require accurate positioning but also 'steady' (restrained) binding of the substrate molecules within the active site. From the E61Q/C165A-ACR structure, it is clear that the LUMO of the amide carbonyl

carbon and the HOMO of cysteine Sγ atom would not be optimally aligned as the entire ACR molecule is twisted away from the thiol group of Cys165 (Fig. 4.19 (B)). The altered substrate orientation (twisting and tilting) would instead favor alignment of the Cα carbon with the Sγ atom of the cysteine (Fig. 4.19 (B) and (C)), which together with the structural properties of FAE (see section 4.4.3.1) would favor occurrence of the S_N2 substitution reaction.

The adopted substrate positioning and orientation in the absence of Glu1 (Fig. 4.19) would imply that the nucleophilic attack at the carbonyl carbon would be greatly reduced or completely eliminated. This is however not entirely the case as a thioester intermediate of acetamide (ACE) was detected by mass spectrometry following incubation of the E61Q mutant with ACE (See figure 4.7 (B) and table 4.2). There are two possible explanations to this observation: (1), the small side chain moiety of acetamide (having 1 carbon atom, see figure 3.1) may permit an orientation that would favor stereo-electronic alignment of its carbonyl carbon and the thiol group of the nucleophilic cysteine; and (2), the bound ACE molecules are likely to be less restrained ('wobbly') within the binding pocket, which would facilitate occasional attack at the carbonyl carbon center to yield trapped thioester intermediates.

As observed previously, the E61L mutant was able to carry out two kinds of reactions with FAE: (1), it reacted with the carbonyl carbon atom resulting in a thioester acyl-enzyme intermediate that was positively identified by MS (Fig. 4.7 (A) and table 4.2); and (2), it unfavorably reacted with the Cα carbon atom generating small amounts of S_N2 reaction adducts that were only observed in the crystal structure (section 4.4.8.1) after prolonged incubation periods. These observations indicate that the substrate molecules were likely to be less constrained ('wobbly') within the binding pocket, which allowed the catalytic cysteine to react with both carbon centers. However, since the S_N2 reaction was much less favorable (section 4.4.8.1), it permits us to speculate that positioning of the substrate molecules in the E61L mutant is different from that observed when Glu61 is changed to a glutamine as visualized in the E61Q/C165A-ACR structure (Fig. 4.19). The mass spectrum of E61L reacted with propionamide (PMD) also had a minor peak for a thioester intermediate (Appendix IV), indicating that the positioning of PMD molecules within the active site permitted nucleophilic attack at the carbonyl carbon, although this reaction was relatively infrequent.

4.4.10 The E61Q/L structures with unexpected Cys165 modifications

A number of crystal structures determined from the E61Q/L mutants crystallized in the presence of some short aliphatic amides had the catalytic cysteine (Cys165) either modified by oxidation or closely associated with a bound substrate molecule within the active site. This was observed even in cases where a reaction intermediate or adduct was detected by mass spectrometry. Three structures, E61Q-ACE, E61Q-PMD and E61L-ACE, are described in this section.

4.4.10.1 The E61Q-ACE and E61Q-PMD crystal structures

The E61Q NitN mutant reacted with ACE formed crystals in reservoir buffer 4 (Table 4.3 (A)), but the crystals either diffracted to very low resolution (4.3 Å) or a number of small crystals were stacked together, making it difficult to isolate single crystals for diffraction. Collection of diffraction data on one of the seemingly 'single' crystals resulted in a 1.7 Å dataset that showed clear overlap of spots from a number of crystals, with no success in all attempts to process the data. Crystallization optimization strategies centered on reservoir buffer 4 and aimed at obtaining single and more ordered E61Q-ACE crystals were not successful. Since mass spectrometric data had shown that E61Q was able to react with ACE to form a thioester acyl-enzyme intermediate, it is possible that the presence of the ACE intermediate in the active site may have caused unusual crystal lattice packing resulting in irregular crystals. Exploration of new crystallization conditions lead to the formation of a few E61Q-ACE crystals in reservoir buffer 39 (Table 4.3 (A)), which only appeared in drops that were dominated by thick protein precipitates after 2-3 weeks following set-up of crystallization. Upon diffraction data collection from a number of these crystals and determination of structures, the extra difference density attached to the thiol group of Cys165 did not resemble the expected ACE acyl moiety, but was indicative of a possible modification to sulfinic acid (Fig. 4.20 (A) and (B)). It is possible that the ACE thioester intermediate may have precipitated out during crystallization, while the cysteine of any free (unreacted) E61Q protein was oxidized to sulfinic acid, resulting in stable protein that was able to pack into crystals. The crystal structure of E61Q pre-incubated with PMD (E61Q-PMD) also had the catalytic cysteine modified to sulfinic acid. In this case, no reaction intermediate or adduct was anticipated as none was detected by mass spectrometry.

The two structures, E61Q-ACE and E61Q-PMD, had very similar active site characteristics. In addition to Cys165-sulfinic acid modification, 3 conserved water molecules and a highly-ordered active site hydrogen bonding network (Fig. 4.20 (C)) appear to have conferred stability to the otherwise unstable mutant proteins resulting in highly-ordered crystals that diffracted to higher than 1.3 Å in both cases. The Cys165-sulfinic acid adduct (Cys-O(OH) in figure 4.20 (C)) is stabilized by six polar contacts, with its O atom interacting with the OG_1 atom of Thr189, Wat29 and the backbone NH group of Ala190, while the OH group hydrogen bonds to the backbone NH group of Tyr166, the amino group of Lys131 (in the oxyanion hole) and a water molecule (Wat163) (Fig. 4.20 (C)). One of the active site water molecules (Wat163 in figures 4.20 (B) and (C)) is closely associated with the OH group of the sulfinic acid adduct, such that even though the two atoms are roughly 2.5 Å apart, the electron density between them appears almost continuous.

Superposition of the E61Q-ACE structure with that of the native NitN enzyme (WT-Apo) shows that the active site regions of the two structures are in a similar conformation (Fig. 4.20 (D)). The active site stabilizing effect of the cysteine-sulfinic acid adduct can thus be appreciated. The OH group of sulfinic acid in E61Q-ACE replaces an oxyanion hole water molecule in the WT-Apo structure, while the coordinating water between the SH group of Cys165 and the carboxylates of Glu61 and Glu139 (WatA) in the WT-Apo structure is preserved in the E61Q-ACE active site.

Due to high reactivity of the thiol group, modification of the reactive cysteine residues by oxidation to sulfenic (Cys-SOH), sulfinic (Cys-SO(OH)) and sulfonic ($Cys-S(=O)_2OH$) acids is common in enzymes that utilize sulfhydryl for catalysis. As a matter of fact, almost all crystal structures of the nitrilase superfamily enzymes that have been determined to-date are modified to various oxidation states at the catalytic cysteine. Although some oxidation states are reversible, their occurrence in sulfhydryl proteins inactivates the enzyme, hence it is common practice to purify and store nitrilase and amidase enzymes that are intended for biochemical activity studies in the presence of reducing agents like dithiothreitol (DTT). NitN is active for up to three months if stored at 4 °C in a buffer containing 2 mM DTT, but it progressively loses activity in the absence of DTT, such that all activity is abolished within three weeks of storage at 4 °C.

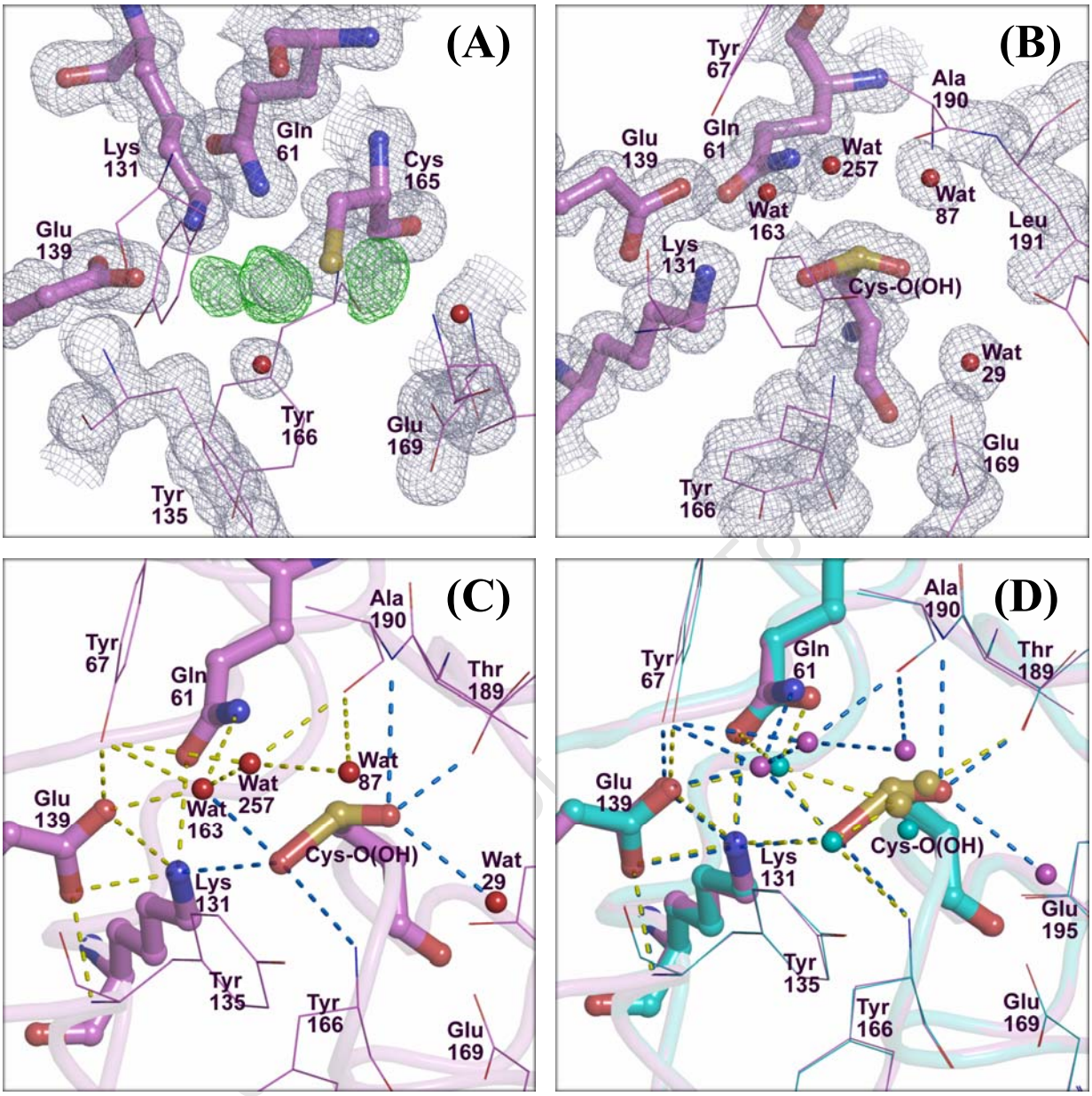


Fig. 4.20: Cysteine sulfinic acid (Cys-O(OH)) modification and the active site polar network in the E61Q-ACE structure. **(A)** The σA-weighted $2F_{obs}$ - F_{calc} (light purple mesh, contoured at 1σ) and F_{obs} - F_{calc} positive difference (green mesh, contoured at 3σ) maps showing density for the binding pocket residues and the omitted sulfinic acid oxygen atoms and a closely associated water molecule. **(B)** The σA-weighted $2F_{obs}$ - F_{calc} map (purple mesh, contoured at 1σ) showing refined density for Cys-O(OH) and the active site region. **(C)** Polar interactions in the active site pocket. Blue dashes represent contacts involving Cys-O(OH), while yellow dashes are potential hydrogen bonds between active site residues and water molecules. The highly ordered polar interactions network greatly stabilizes the active site and the entire 3D structure. **(D)** Superposition of E61Q-ACE (violet CPK) and the WT-apo NitN (cyan CPK) active sites. The violet non-bonded spheres and blue dashes are waters and potential polar contacts respectively in the E61Q-ACE active site while cyan spheres and yellow dashes are waters and potential hydrogen bonds respectively in the WT-apo structure.

4.4.10.2 The E61L-ACE structure

Based on the mass spectrometric data, no reaction intermediates or adducts were formed when the E61L mutant was incubated with acetamide (ACE) and therefore none were anticipated in the crystal structure. However, the structure of the E61L mutant crystallized in the presence of ACE (E61L-ACE) had positive difference electron density which indicated that the catalytic cysteine (Cys165) adopted two different conformations, and that one of the conformers was likely to be covalently attached to an unknown adduct (Fig. 4.21 (A)). Careful examination of the electron density maps revealed that the extra density resembled an ACE molecule that was not necessarily bound to Cys165, but was otherwise close enough to the thiol groups of the two cysteine conformers to appear covalently attached. Modeling and refinement of a non-covalently bound ACE molecule in the extra density resulted in a near perfect fit as shown in figure 4.21 (B)). The α-carbon atom of the ACE molecule is 2.13 Å and 2.33 Å away from the SH groups of the two Cys165 conformers, ruling out the possibility of covalent attachment as these distances are larger than the normal S-C covalent bond, which is approximately 1.8 Å in length. Furthermore, there is no known evidence or possibility of the catalytic cysteine attacking the Cα atom (which is effectively a methyl group) of ACE to form a covalent bond, or for the formation of a H-bond between the SH group of Cys165 and the CH_3 group of ACE.

In the structure, the carbonyl oxygen of the bound ACE molecule is located in the oxyanion hole, hydrogen bonding to the backbone NH group of Tyr166 and the amino groups of the two Lys131 conformers (Fig. 4.21 (C)). Although there are no possible contacts between the amide nitrogen of ACE and the protein atoms, the carbonyl carbon and the Cα atoms of ACE are stabilized by hydrophobic interactions with the Cβ atom of Tyr135, Cβ atom of Cys165 and the CD_1 atom of Leu191. The binding of the ACE molecules in close proximity to the catalytic cysteine may have caused steric clashes that probably lead to the displacement of the entire cysteine residue, which in turn caused the region of the active site loop between Val164 and Tyr166 to adopt two different conformations (Fig. 4.21 (D)). The catalytic lysine also has its side chain adopting two alternative conformations (Fig. 4.21 (D)).

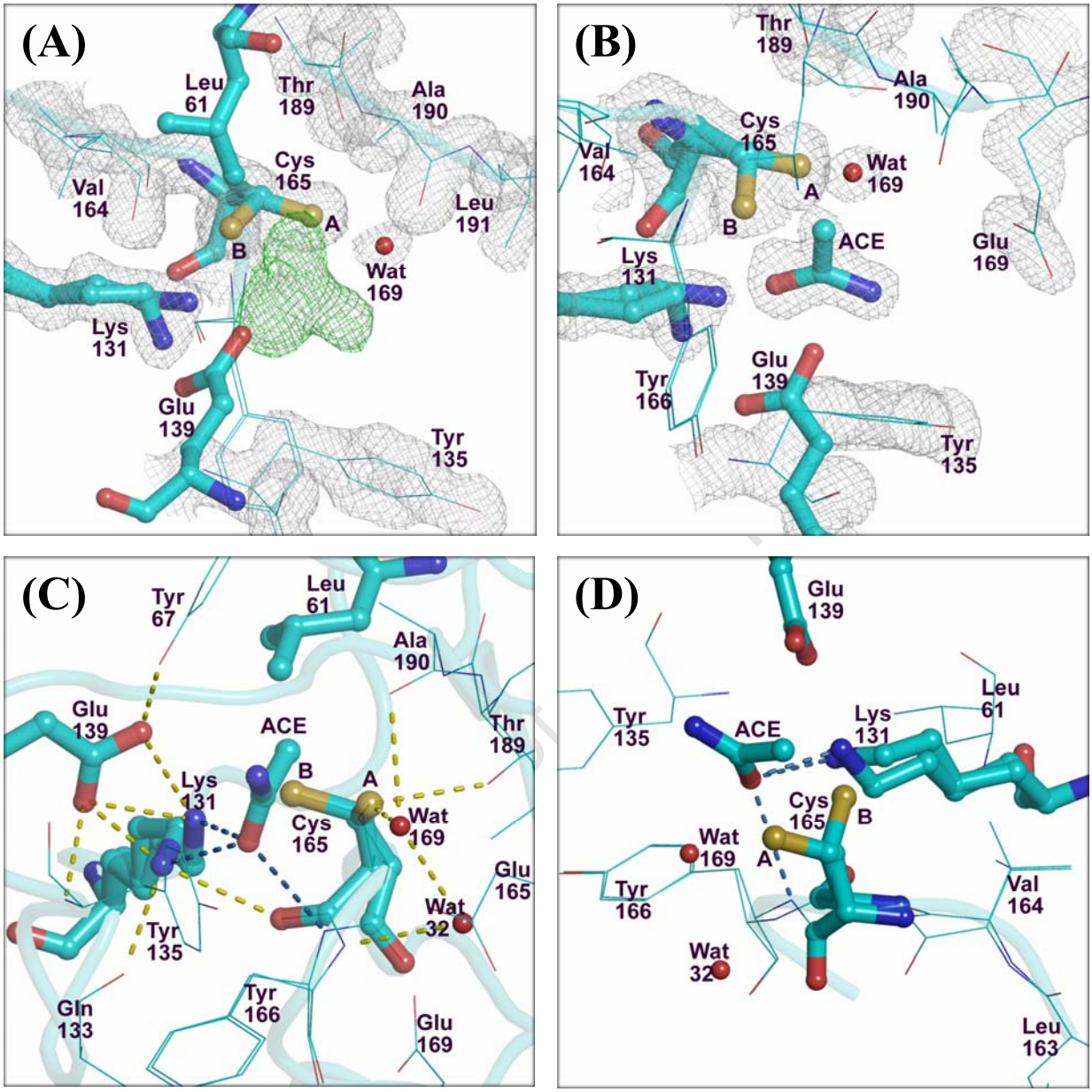


Fig. 4.21: Acetamide (ACE) molecule closely associated with the catalytic cysteine and the active site polar network in the E61L-ACE structure. **(A)** σA-weighted $2F_{obs}$ - F_{calc} (grey mesh, contoured at 1σ) and F_{obs} - F_{calc} positive difference (green mesh, contoured at 3σ) maps showing density for the active site residues and the omitted ACE molecule that is closely associated with Cys165. The SH atoms of the two Cys165 conformers are labeled A and B. **(B)** σA-weighted $2F_{obs}$ - F_{calc} map (grey mesh, contoured at 1σ) showing refined density for the ACE molecule and the active site region. **(C)** Polar interactions in the active site pocket. Blue dashes represent contacts involving the ACE molecule, while yellow dashes are potential hydrogen bonds involving active site residues and water molecules. **(D)** Active site residues adopting alternative conformations, probably due to steric interaction between ACE and the catalytic cysteine (Cys165) side chain.

The binding of ACE and the observed conformational changes in the region surrounding Cys165 in the E61L-QCE structure appeared to have conferred stability to the otherwise unstable mutant protein as the crystals were quite orderly, diffracting to higher than 1.3 Å resolution. Indeed, the alternative conformers stabilize the active site through their polar interactions with the substrate, other protein atoms and water molecules, as shown in figure 4.21 (C). The thiol group of the A conformer of Cys165 interacts with both the OG_1 of Thr189 and a water molecule (Wat169), which in turn interacts with Wat32. Wat32 on the other hand makes a hydrogen bond to the backbone NH group of Tyr166 B conformer. The amino group of the A conformer of Lys131 interacts with the carbonyl oxygen of ACE, the carboxyl oxygen of Cys165 B conformer and the carboxyl oxygen of Gln133. The amino group of Lys131 B conformer on the other hand interacts with the carbonyl oxygen of the bound ACE, as well as the carboxyl oxygen atoms of Glu139 (Fig. 4.21 (C)).

In the E61L-ACE structure, Cys73, which is located in a small helix close to the surface of the molecule, was found to have extra density attached to the Sγ atom, which was interpreted as a possible oxidation to sulfenic acid (Fig. 4.22). The exposure of Cys73 to the solvent environment makes its thiol group accessible to oxidation by free radicals probably arising from the radiation exposure during diffraction data collection.

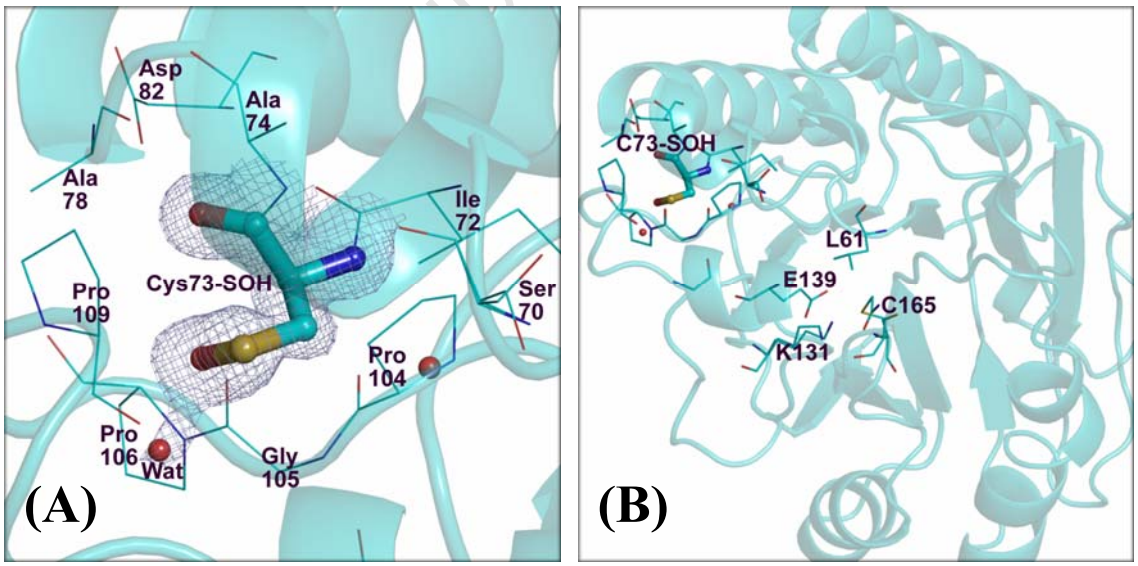


Fig. 4.22: Location and oxidation modification of Cys73 in the E61L-ACE structure. **(A)** Oxidation of Cys73 to what is possibly sulfenic acid. The light purple mesh is the σA-weighted $2F_{obs}$ - F_{calc} (contoured at 1σ) showing the refined density for Cys73-sulfenic acid (Cys73-SOH) and a closely associated possible water molecule at 2.3 Å from the OH atom of the sulfenic acid. **(B)** A cartoon representation of the E61L-ACE NitN monomer, showing the location of Cys73 in the structure. Cys73, now oxidized to Cys73-SOH is located on a small helix close to the surface of the molecule.

4.4.11 Insights into the catalytic mechanism

4.4.11.1 The mechanism for the catalytic cysteine activation

The 'first' glutamic acid (Glu61 in NitN) is generally accepted as the general base catalyst in the amidase-catalyzed reactions, where it has been proposed to facilitate proton transfer by: (1) abstracting a proton from the -SH group during the nucleophilic attack at the amide substrates (acylation step); (2) donating the acquired proton to the amide amino group (NH_2) in the tetrahedral enzyme-substrate intermediate leading to the release of the first product, ammonia; (3) deprotonating the water substrate that attacks the thioester intermediate in the deacylation phase; and (4) donating the proton back to the second tetrahedral intermediate, allowing the release of the acid product and regeneration of an active enzyme, which completes the reaction cycle. The proposed role of Glu1 was based on the location of its carboxyl group relative to the thiol group of the catalytic cysteine. However in the native NitN structure (WT-Apo), the inter-atomic distances between the carboxyl oxygens of the Glu61 (Glu1) and the Sγ atom of the nucleophilic cysteine are larger than the longest possible hydrogen bonding distance of approximately 3.4 Å (Fig. 4.23 (A) and Table 4.7), which may imply that Glu61 is unlikely to activate the cysteine through direct interaction with the thiol group in NitN, but possibly through a bridging water molecule. In the latter mechanism, the carboxyl group of Glu61 would deprotonate a water molecule, which would in turn abstract the thiol proton from Cys165. Contrary to this theory, in the substrate-complexed C165A-BMD and C165A-PMD NitN co-crystal structures (section 3.4.6.2), there are no visible water molecules in the binding pocket (Fig. 3.7) that would possibly fulfill this role, and furthermore, there is not enough space to accommodate a coordinated water molecule between the carboxyl group of Glu61 and the thiol of Cys165. Similarly, there are no water molecules in the substrate-complexed structures from the *Helicobacter pylori* formamidase (Hung et al. 2007) and the *Agrobacterium* sp. DCase (Hashimoto et al. 2003). It therefore appears quite unlikely that a water molecule would be involved in the activation of the catalytic cysteine in amidases. Since the side chain of Cys165 has been observed to adopt different conformations in various NitN structures (see aligned structures in figure 4.23 (B) for example), it is possible that the enzyme may utilize this flexibility to facilitate a direct contact between Glu61 and Cys165 side chains during the thiol group activation.

Table 4.7 Distances of carboxyl oxygens of Glu61 and Glu139 to the -SH group of the catalytic cysteine in the WT-Apo NitN structure.

Glutamate residue, atom	Distance from Sγ of Cys165 (Å)
Glu61	
OE_1	4.81
OE_2	3.48
Glu139	
OE_1	5.35
OE_2	6.29

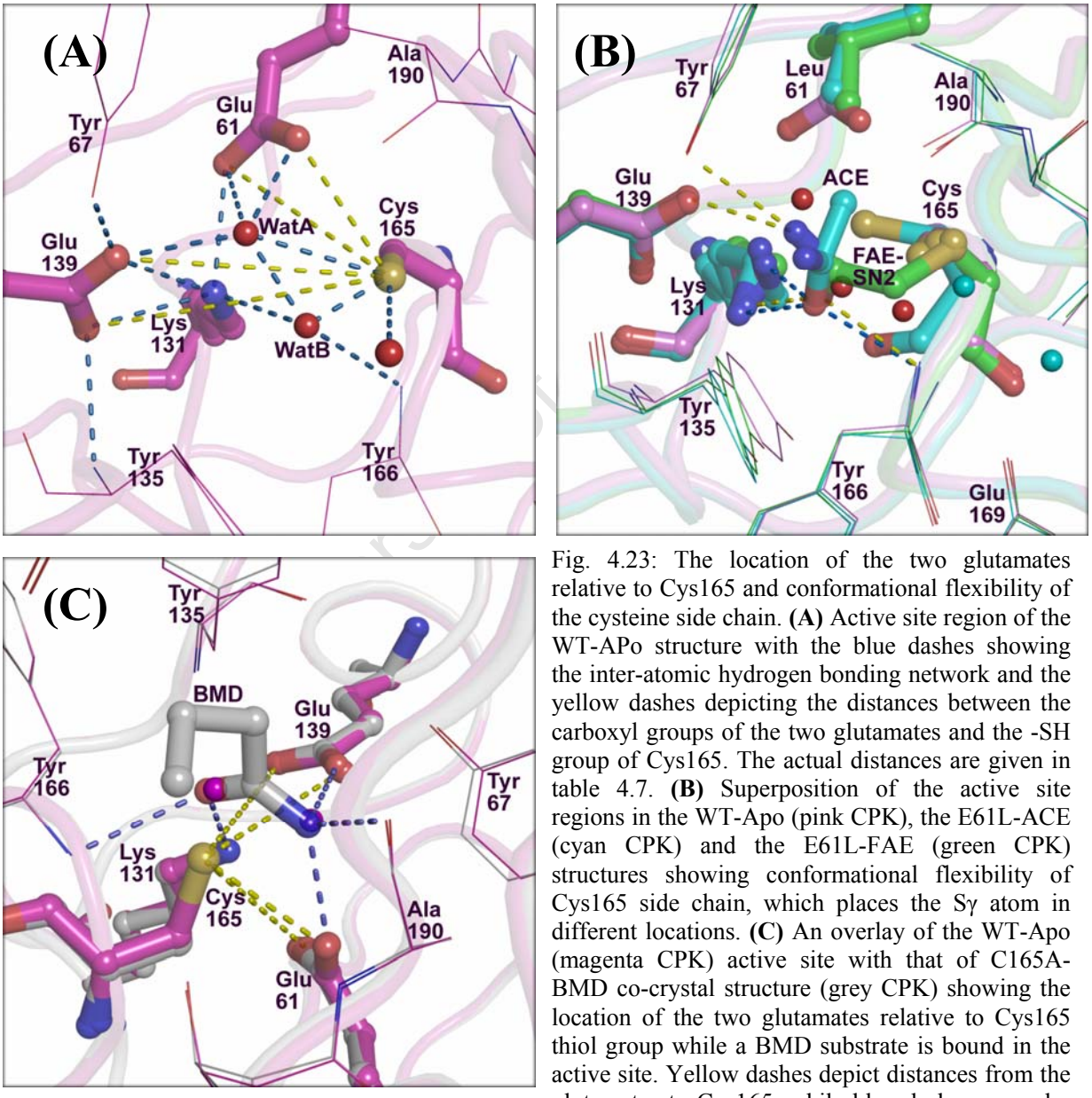


Fig. 4.23: The location of the two glutamates relative to Cys165 and conformational flexibility of the cysteine side chain. **(A)** Active site region of the WT-APo structure with the blue dashes showing the inter-atomic hydrogen bonding network and the yellow dashes depicting the distances between the carboxyl groups of the two glutamates and the -SH group of Cys165. The actual distances are given in table 4.7. **(B)** Superposition of the active site regions in the WT-Apo (pink CPK), the E61L-ACE (cyan CPK) and the E61L-FAE (green CPK) structures showing conformational flexibility of Cys165 side chain, which places the Sγ atom in different locations. **(C)** An overlay of the WT-Apo (magenta CPK) active site with that of C165A-BMD co-crystal structure (grey CPK) showing the location of the two glutamates relative to Cys165 thiol group while a BMD substrate is bound in the active site. Yellow dashes depict distances from the glutamates to Cys165, while blue dashes are polar contacts between BMD and the protein atoms.

Based on the reaction intermediates and cysteine adducts observed in this study when the E61Q/L NitN mutants were incubated with some amide substrates, the general base catalysis role of Glu61 (Glu1) is questionable. The E61L/Q mutants reacted with FAE, ACE and PMD to yield thioester acyl-enzyme intermediates that were positively identified by mass spectrometry (Fig. 4.7 and table 4.2), and they were also able to carry out an S_N2 nucleophilic substitution of fluorine in FAE as detected by mass spectrometry (Fig. 4.7 and table 4.2) and visualized through X-ray crystallography (section 4.4.8). These reactions indicate that Glu61 is not absolutely required to catalyze nucleophilic attack by the cysteine. The question is, how then is the catalytic cysteine activated in the absence of Glu1?

In the absence of Glu61, a probable option for a general base catalyst would be Glu139, although its ability to fulfill this role is highly unlikely given its location relative to Cys165 (Fig. 4.23 (A)). The carboxyl oxygen atoms of Glu139 are not only more than 5 Å away from the -SH group of Cys165 (Fig. 4.23 (A) and table 4.7), but their access would be obstructed when a substrate molecule is bound in the active site (Fig. 4.23 (C)). This heavily suggests that there could be an alternative mechanism for the cysteine thiol group activation in the E61Q/L NitN mutants. Although the mutation of Glu139 to glutamine resulted in an unstable protein that does not exhibit any biochemical reactivity, reaction intermediates or adduct formation, this could be entirely because of incorrect positioning of the substrates rather than its involvement in the activation of the thiol nucleophile. However, if Glu139 is indeed involved, then it must achieve this role through a bridging water molecule.

We propose two mechanisms (Fig. 4.24) for the catalytic cysteine activation during the S_N2 substitution reaction of the E61Q/L mutants with FAE. In figure 4.24 panel (A), the involvement of a base (X) is proposed. Since no obvious base candidates can be seen in the mutant protein, it is necessary to postulate that X is a water molecule, in which case HX would be a hydronium ion (H_3O^+). However, since there is no enough space for a water molecule that can fulfill this role in the substrate-complexed C165A-BMD and C165A-PMD co-crystal structures (see section 3.4.6.2 and figure 3.7), this mechanism is speculative. An alternative mechanism (Fig. 4.24 (B)) is suggested by the geometry of the situation; it is clearly possible for the cysteine S-H bond to align with the C-F bond of FAE. A nucleophilic attack on the α-carbon would result in an increase in electron density at the fluorine atom directly adjacent to and stereo-electronically aligned with the thiol hydrogen. Rather than postulate a mechanism involving water acting as a base, it seems much more attractive to

propose a single concerted mechanism, whereby, as the thiol group attacks at the α-carbon of FAE, the thiol proton is abstracted by the fluorine leaving group.

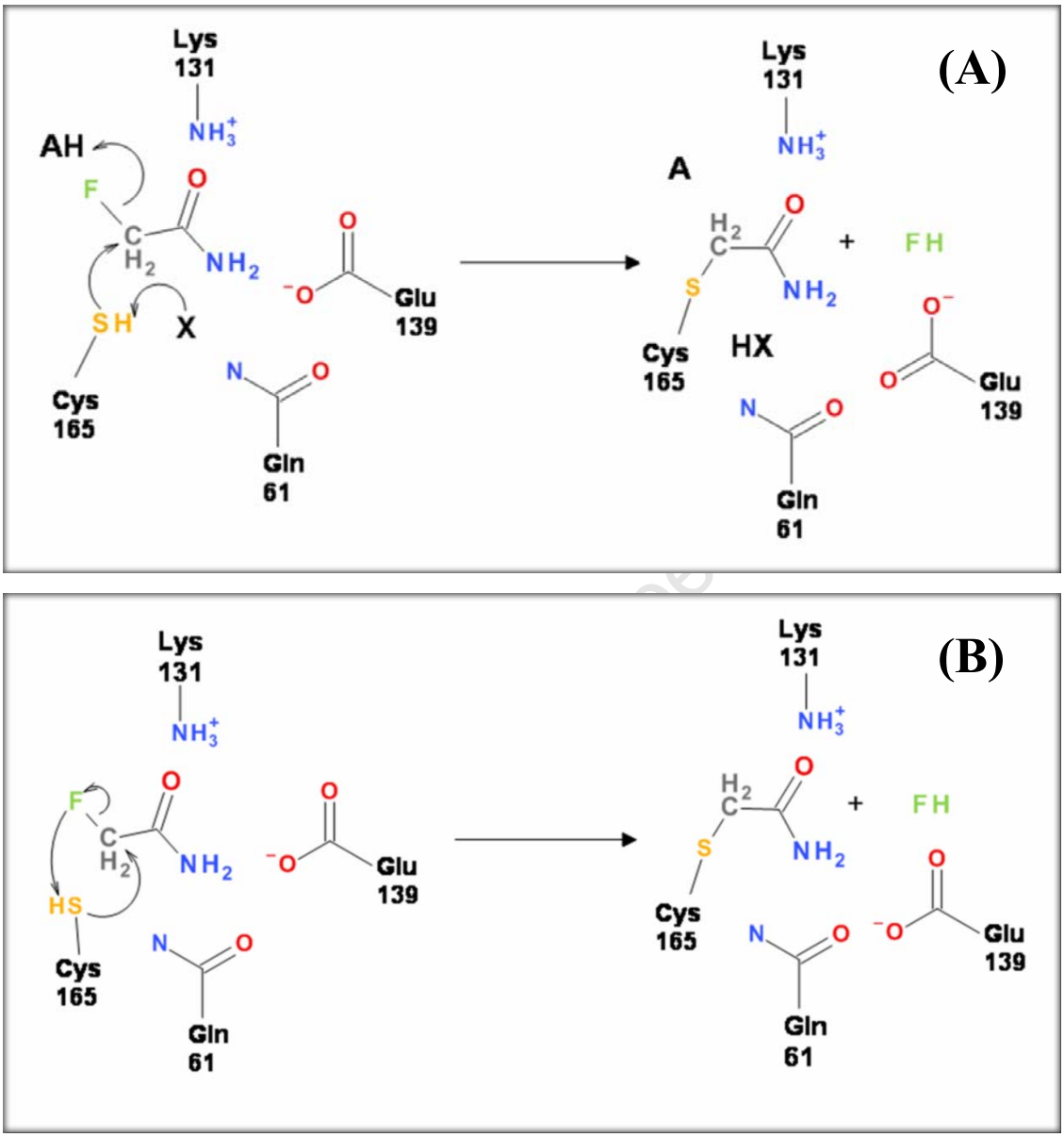


Fig. 4.24: The alternative mechanisms that are proposed in this work for the S_N2 substitution reaction involving E61Q/L NitN mutants and FAE. **(A)** A reaction scheme involving a possible base (X) for the activation of Cys165 nucleophile. We have proposed a water molecule to be the base, in which case HX would be a hydronium ion. **(B)** A concerted mechanism where the deprotonation of the catalytic cysteine is proposed to be substrate-assisted. The fluorine atom (F) abstracts the thiol atom during the attack, resulting in a one-step reaction where FH is eliminated.

The activation of the catalytic cysteine (Cys165) during the attack of the carbonyl carbon to yield a thioester intermediate in the E61Q/L mutants could also be substrate-assisted. The pi (π) orbital electrons from the double bond of the amide carbonyl group are capable of delocalizing to the amino group, which would create a negative charge on the nitrogen atom,

making it capable of abstracting a proton from the -SH group of the nucleophilic cysteine following optimal stereo-electronic alignment of the two groups. As the cysteine attacks at the carbonyl carbon, the amide group of the substrate would abstract the proton from the -SH group simultaneously, leading to the release of ammonia and the direct formation of an acyl-enzyme intermediate.

The possibility of a substrate-assisted cysteine activation mechanism in the absence of Glu61 as well as in the WT enzyme will need to be investigated.

4.4.11.2 The hydrolysis of the thioester intermediate

The deacylation phase of amide hydrolysis reaction is initiated by the attack of the thioester acyl-enzyme intermediate by an activated water molecule. Glu1 is proposed to be the general base catalyst of the deacylation water. In the E61L/Q mutants reacted with fluoroacetamide (FAE), acetamide (ACE) and propionamide (PMD), respective thioester intermediates that were trapped in the active site were detected by mass spectrometry (MS) (Fig. 4.7 and table 4.2) and that of E61L with PMD visualized in the crystal structure (section 4.4.7.1). The inability of the mutant enzymes to hydrolyze these thioester intermediates strongly suggests that the 'first' glutamate (Glu61) is required as a general base catalyst of the water substrate.

(A) The location of the deacylation water

Based on the superposition of the native NitN structure (WT-Apo) with that of the E61L-PMD structure having a PMD thioester intermediate in the active site (Fig. 4.25 (A)), Glu61 would be most suited and well positioned to deprotonate (activate) the deacylation water, which would be positioned between the two glutamates, that is, in a similar location to WatA in the WT-Apo structure (Fig. 4.25 (A)). This location is supported by quantum mechanical calculations of the butyramide (BMD) thioester intermediate, whose model is shown in figure 4.25 (B). The geometry of this intermediate also supports Glu61 to be the general base catalyst for WatA (Fig. 2.25 (B)) and also suggests an assisting role for Glu139 (Glu2) in positioning the water molecule and maintaining the active site configuration.

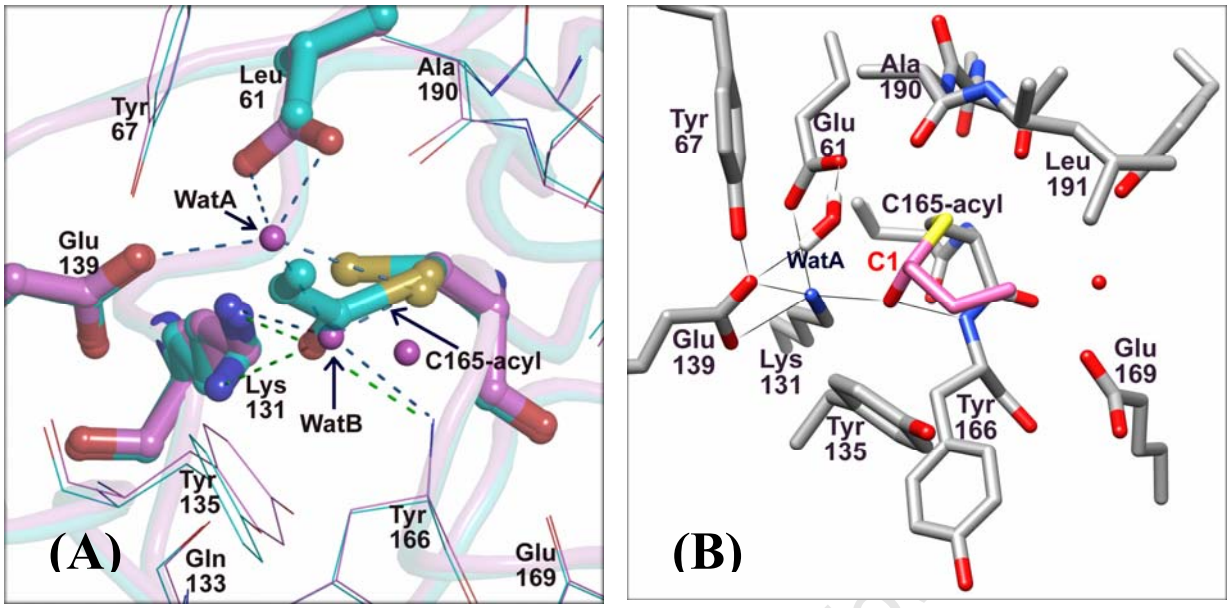


Fig. 4.25: The geometry of the thioester intermediate and the position of the deacylation water. **(A)** An overlay of the binding pocket of the WT-Apo structure (violet CPK) with that of the E61L-PMD structure (cyan CPK) showing the coordination of the PMD thioester intermediate (C165-acyl). Green dashes represent polar contacts between the carbonyl oxygen of the thioester and the oxyanion hole protein atoms, while violet non-bonded spheres are water molecules and their possible polar interactions (blue dashes) in the WT-Apo pocket. Glu61 is suitably positioned to fulfill the role of a general base catalyst of a water molecule in WatA position. **(B)** A lines representation of a fully converged quantum mechanical model of the BMD thioester intermediate (C165-acyl). The BMD acyl moiety is shown in pink CPK color and the grey lines represent hydrogen bonding network within the active site. The deacylation water (WatA) is coordinated between Glu61 and Glu139, and its location is optimal for a nucleophilic attack at the sp^2-hybridized carbonyl carbon center (C1). The geometry of the model suggests that Glu61 would be the general base catalyst and that Glu139 assists with positioning the water molecule. Figure (B) was made in UCSF Chimera (Pettersen et al. 2004).

A) The role of the catalytic lysine

As discussed in the literature review section (section 1.3.3), the crystal structure of the aliphatic amidase from *Pseudomonas aeruginosa* had a second tetrahedral intermediate of an acyl transfer reaction between acetamide (ACE) and hydroxylamine (Fig. 4.26 (A)) trapped in the active site (Andrade et al. 2007). In their analysis, Andrade and colleagues (2007) explored the mechanistic consequences of the binding of the amide carbonyl oxygen in the oxyanion hole that is formed by the positively-charged amino group (NH_3^+) of the catalytic lysine and the peptidic NH group of the residue next in sequence after the cysteine (Fig. 4.26). The positive character of the lysine amino group would polarize the delocalized electrons from the carbonyl group double bond, which would make the carbonyl carbon a better

electrophile, hence facilitating nucleophilic attack by both the cysteine and the water substrate. In addition, the oxyanion hole hydrogen bonding network involving the catalytic lysine (Fig. 4.26) would stabilize the 'oxyanion' that is formed as the lone pair of electrons from the amide carbonyl group double bond move to the carbonyl oxygen during the formation of the two tetrahedral intermediates.

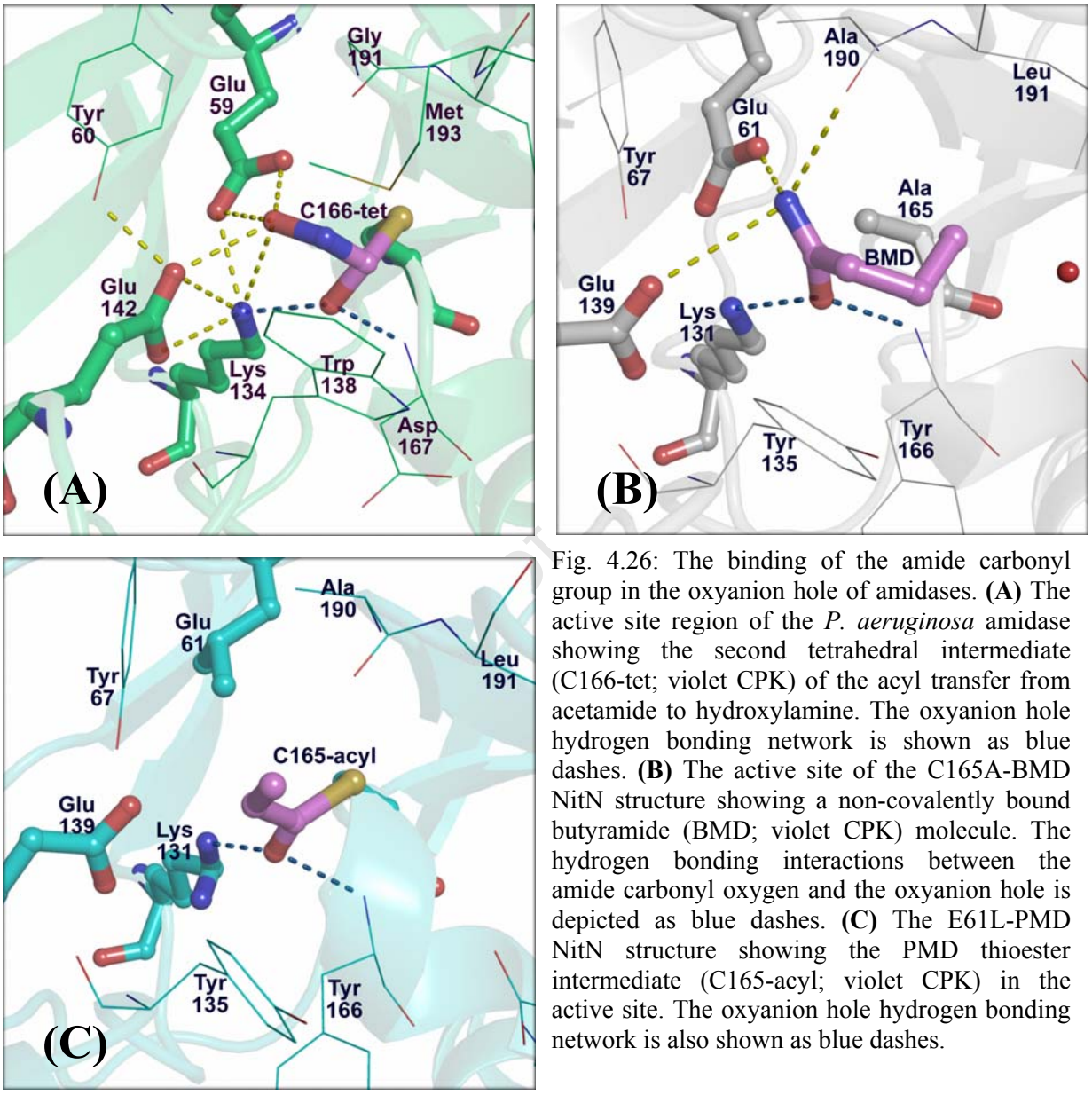


Fig. 4.26: The binding of the amide carbonyl group in the oxyanion hole of amidases. **(A)** The active site region of the *P. aeruginosa* amidase showing the second tetrahedral intermediate (C166-tet; violet CPK) of the acyl transfer from acetamide to hydroxylamine. The oxyanion hole hydrogen bonding network is shown as blue dashes. **(B)** The active site of the C165A-BMD NitN structure showing a non-covalently bound butyramide (BMD; violet CPK) molecule. The hydrogen bonding interactions between the amide carbonyl oxygen and the oxyanion hole is depicted as blue dashes. **(C)** The E61L-PMD NitN structure showing the PMD thioester intermediate (C165-acyl; violet CPK) in the active site. The oxyanion hole hydrogen bonding network is also shown as blue dashes.

The role of the oxyanion hole in ensuring the catalytic efficiency of the hydrolysis reaction has been extensively studied in serine proteases. Site-directed mutagenesis of an asparagine residue (N155A) that forms part of the oxyanion hole in the bacterial subtilisin protease resulted in the lowering of the catalytic constant (K_{cat}) by a factor of 200 - 300 (Bryan et al.

1986), indicating that the hydrogen bonding network in the oxyanion hole is responsible for lowering the activation energy of the reaction. The role of the oxyanion hole hydrogen bonding network in polarizing the planar carbonyl group of both the non-covalent enzyme-substrate complex and the acyl-enzyme intermediate has also been confirmed (Fodor et al. 2006;Whiting and Peticolas 1994;Wilmouth et al. 2001). The distortion of the electrons towards the carbonyl oxygen atom is also suggested to make the carbonyl carbon more electrophilic, which enhances the two nucleophilic attack steps in the catalytic process of serine proteases.

In this study, the crystal structure of the catalytic lysine NitN mutant (K131Q) (described in section 3.4.10) had a possible thioester intermediate of adipamide (ADM) or adipamic acid covalently attached to the catalytic cysteine (Fig. 4.27). While ADM has been confirmed to be a good substrate for NitN (see section 3.4.10 and figure 3.19), the trapping of the thioester intermediate in the K131Q mutant indicates that Lys131 is required to facilitate nucleophilic attack at the thioester by the water substrate, which confirms its role in polarizing the delocalized electrons from the carbonyl group double bond as suggested by Andrade and colleagues (2007). The change of Lys131 to a glutamine in the K131Q NitN mutant replaces an ionizable amino group with a non-ionizable amide group, which is unable to to participate directly in the oxyanion hole hydrogen bonding network (Fig. 4.27) and would therefore not be able to attract the electrons from the carbonyl group hence the deacylation water molecule would be incapable of attacking efficiently.

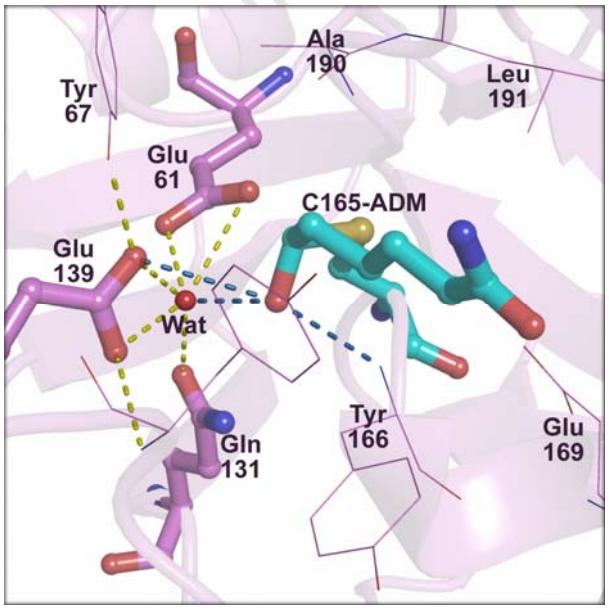


Fig. 4.27: The active site region of the K131Q-ADM structure showing the ADM thioester intermediate (C165-ADM; cyan CPK) that was trapped when Lys131 was changed to a glutamine in NitN. The blue dashes are polar contacts involving the carbonyl oxygen of the intermediate, while yellow dashes show polar interactions of other atoms in the active site.

4.4.12 A summary of the deduced roles of the two glutamates in catalysis

The mutational, mass spectrometric and X-ray crystallographic studies of the two active site glutamates in NitN, as presented here, suggest that the catalytic unit in amidases is composed of a sulfhydryl group (from the reactive cysteine, Cys165), a positively-charged amino group (from the catalytic lysine, Lys131), two negatively charged carboxylates (from the two active site glutamates, Glu61 and Glu139) and a water molecule. This is in agreement with previous suggestions of a CEEK catalytic tetrad by Kimani and colleagues (2007) and later by Soriano-Maldonado and coworkers (2011). The following sections will provide a detailed summary of the deduced, suggested or proposed roles of the two catalytic glutamates in NitN, as revealed by the findings presented in this thesis.

4.4.12.1 The importance of the two glutamates for activity and for maintaining the active site configuration

In the native WT NitN structure (WT-Apo), the two active site glutamates are involved in a highly ordered hydrogen bonding network (Fig. 4.6) that is important for maintaining the side chain configuration of the active site residues as well as the stability of the binding pocket. Mutation of either of the two glutamates resulted in unstable proteins that were largely expressed as inclusion bodies in *E. coli* (section 4.4.1); the soluble fractions were highly prone to precipitation during handling and the mutant proteins could not crystallize without stabilizing compounds or substrate molecules. This clearly indicates that in addition to maintaining the active site structure, the two glutamates are also necessary for maintaining the NitN fold. All mutant proteins also lacked enzymatic activity, implying that both glutamates are required for the hydrolysis of the amide substrates.

Estimations of the pKa values of the active site residues in the E61Q, E61L and E139Q NitN mutant structures using the H++ server (Gordon et al. 2005) showed that the pKa of Lys131 amino group was greatly reduced when either of the two glutamates was replaced relative to the estimated pKa value of Lys131 side chain in the WT structure (table 4.8). This is in agreement with the suggestions by Andrade and colleagues (2007) that a high pKa of the catalytic lysine amino group is maintained by the hydrogen bonds from the negatively-charged carboxylates of the two glutamates (Fig. 4.6), which would be important for ensuring

that the amino group remains protonated throughout the reaction. These observations highlight the importance of the two glutamates in maintaining an active site configuration that is optimal for substrate binding and hydrolysis.

Table 4.8 Estimated pKa values for some of the active site residues when either of the two active site glutamates is mutated.

NitN Structure	Estimated pKa value at pH 7.0				
	Glu61	Tyr67	Lys131	Glu139	Cys165
WT	7.1	34.4	**43.2**	-10.7	20.5
E61Q	-	27.9	**35.4**	-8.6	17.5
E61L	-	26.2	**17.6**	-6.0	20.6
E139Q	-4.8	19.4	**18.8**	-	23.0

4.4.12.2 The importance of the two glutamates in substrate positioning

The co-crystal structures of the C165A mutant with butyramide (C165A-BMD) and propionamide (C165A-PMD) revealed that short aliphatic amides bind in the active site of NitN through polar interactions between their carbonyl-amino group and the catalytic residues, as well as hydrophobic contacts of their side chain moiety with various active site atoms (see section 3.4.6.2 and figure 3.7 in chapter 3). While the carbonyl oxygen of the amide molecules would have a high affinity for the oxyanion hole (formed by the Lys131 amino group and the NH group of Tyr166 in NitN), accurate positioning of the substrate molecules for nucleophilic attack by the cysteine appears to be dependent on the positioning of the amide amino group through hydrogen bonding interactions with the carboxylates of the two glutamates. This is clearly evident from: (1), the lack of detectable activity or reaction adducts when the E139Q mutant was incubated with the short aliphatic amide substrates (section 4.4.3 and Fig. 4.7 (A)); (2), the 'artificial' reaction adducts (S_N2 reaction FAE-Cys165 adduct) formed due to the reaction of the cysteine with the α-carbon of incorrectly positioned FAE molecules in the E61Q/L NitN mutants (Fig. 4.7 (B) and section 4.4.8); and (3), the co-crystal structure of the incorrectly positioned ACR molecule in the E61Q/C165A NitN mutant (E61Q/C165A-ACR) (section 4.4.9).

4.4.12.3 The questionable role of Glu1 in activating the catalytic cysteine

Upon optimal binding of the amide substrates in the active site, the first nucleophilic attack is initiated by the activation of the catalytic cysteine through the removal of the thiol proton. The 'first' glutamate has previously been proposed to be responsible for the catalytic cysteine deprotonation (Nakai et al. 2000), a role that is generally accepted. The geometry and configuration of the active site as well as the location of Glu61 (Glu1) relative to Cys165 in NitN (see section 4.4.11.1 and figures 4.23 (A) and 4.25 (B)) are consistent with Glu61 being the general base catalyst for the cysteine as proposed.

However in this study, trapped thioester intermediates of FAE, ACE and PMD, as well as FAE S_N2 reaction adducts at the catalytic cysteine were observed in Glu61 mutants (E61Q/L) (see figure 4.7, table 4.2 and sections 4.4.7 and 4.4.8), indicating that Glu1 is not absolutely required for catalyzing nucleophilic attack by the cysteine. In the absence of Glu61, Glu139 might step-in as a general base catalyst, although this is highly unlikely given the location of its carboxylate relative to the thiol group of Cys165 (see figures 4.23 (A) and (C)). We propose that the nucleophilic attack at the amide carbonyl carbon might be substrate-assisted in the E61Q/L mutants, with the amide amino group serving to abstract the thiol proton concertedly as the cysteine Sγ atom attacks. Similarly, the attack at the α-carbon of FAE during the S_N2 reaction with the E61Q/L mutants could also be substrate-assisted, with the fluorine atom abstracting the thiol proton in a concerted manner as the Sγ atom of the nucleophile attacks (see section 4.4.11.1 and figure 4.24 (B)).

4.4.12.4 The role of the two glutamates in the deacylation phase of the reaction

The deacylation phase of the amide hydrolysis reaction is thought to be initiated when an activated water molecule attacks at the carbonyl carbon of the thioester intermediate; Glu1 is proposed to be the base catalyst for the deacylation water. The occurrence of trapped thioester intermediates of various aliphatic amides in the binding pocket of the E61Q/L NitN mutants (section 4.4.3 and table 4.2) indicate that Glu61 is required as a catalyst in the deacylation phase of the reaction. The geometry of a quantum mechanical model of the butyramide (BMD) thioester intermediate (Fig. 4.25 (B)) supports the role of Glu61 in activating the water substrate and also suggests a supporting role for Glu139 in positioning the water substrate and in maintaining the active site configuration.

Our data supports the suggested general acid catalyst role of the catalytic lysine in the deacylation step of the reaction - whereby the positively-charged amino group acts by polarizing the the electrons from the carbonyl group of the intermediate, resulting in increased electrophilicity at the reactive carbonyl carbon, which facilitates attack by the water substrate. The structure of the catalytic lysine mutant (K131Q) having a thioester intermediate of adipamide or adipamic acid trapped in the active site (section 3.4.10 and figures 3.18 and 4.27) provides satisfactory evidence for the involvement of Lys131 as a general acid catalyst. This role is also suggested for the acylation phase nucleophilic attack, in addition to the generally accepted role of stabilizing negatively-charged tetrahedral intermediates.

4.5 Concluding remarks

The work reported in this chapter has clearly demonstrated that the catalytic unit in the nitrilase superfamily amidases is composed of a Cys, Glu1, Glu2, Lys (CE_1E_2K) catalytic tetrad in which the two glutamates and the lysine function as a modular assembly, rather than the previously proposed CEK triad where each residue was implied as acting as a separate entity. The study has particularly emphasized the importance of the two glutamates in the positioning of the amide substrates for nucleophilic attack by the catalytic cysteine. The attack at other reactive carbon centers by the cysteine or its inability to react with the substrates are some of the observed consequences of incorrect substrate positioning following the mutation of either of the two glutamates. The proposed role of Glu1 being a general base catalyst of the deacylation water has been confirmed but its involvement in the activation of the catalytic cysteine is questionable. The mechanism for the nucleophilic attack by the cysteine will therefore need to be investigated further and the rates of the formation of the reaction intermediates and adducts in the absence of Glu1 will need to be characterized. Other important roles of the two glutamates include maintenance of both the NitN structural fold and the active site configuration.

Chapter 5

Activity and reactivity of NitN with acrylamide

5.1 Abstract

Although NitN exhibits activity on propionamide (PMD) and fluoroacetamide (FAE), it is unable to hydrolyze similar-sized acrylamide (ACR) molecules. Mass spectrometric and X-ray crystallographic studies of the wild-type (WT) and the two glutamate mutants (E61Q/L and E139Q) in the presence of ACR were carried out in order to understand why NitN lacked enzymatic activity on ACR and to investigate the effect of the mutations on ACR positioning within the active site. The conjugated structure of ACR was explored as one of the contributing factors towards the inability of NitN to hydrolyze ACR, as the amide carbonyl carbon atom has reduced electrophilicity due to polarization of the electrons from the alkene double bond towards the carbonyl group. Enzymatic assays of the wild-type NitN in the presence of both ACR and PMD showed that ACR has a mild inhibitory effect on NitN. Mass spectrometry revealed that the WT enzyme unfavorably reacted with both the carbonyl carbon and the β-alkene carbon atoms of ACR resulting in the trapping of small quantities of ACR thioester intermediates and Michael adducts at Cys165 respectively. These reaction intermediates accumulated with prolonged incubation and they irreversibly inactivated the enzyme, explaining the observed inhibitory effect. The active site pocket of NitN was incapable of adequately restraining ACR molecules, explaining the observed ability of the cysteine to react with the two carbon centers. The crystal structure of the WT NitN enzyme pre-incubated with ACR had electron density corresponding to an ACR Michael adduct at 0.3 atomic occupancy, mixed with a large proportion of unreacted protein. The inability of the wild-type enzyme to further hydrolyze the trapped ACR thioester intermediate was also attributed to the conjugated nature of the acyl moiety in the intermediate, which would make the carbonyl carbon a poor electrophile and hence the water substrate would not be able to attack readily. The E139Q mutant favorably reacted with the β-alkene carbon atom of ACR only as depicted by a high signal peak in the mass spectra, indicating a substrate orientation that favored stereo-electronic alignment of the cysteine thiol group and the β-carbon of ACR E61Q/L mutants on the other hand favorably reacted with both the carbonyl carbon and β-alkene carbon atoms of ACR to yield thioester intermediates and Michael adducts that were detected by mass spectrometry and visualized in the crystal structure of E61Q pre-incubated with ACR. These reactions emphasize the importance of two glutamates in substrates positioning for nucleophilic attack, and also reiterate the findings from the previous chapter that Glu1 is not required to catalyze the nucleophilic attack by the cysteine.

5.2 Introduction

Acrylamide (ACR) is an α,β-unsaturated carbonyl derivative belonging to the type-2 alkenes, which are characterized by conjugated structures formed when an electrophilic functional group (like a carbonyl or amide group) is linked to an alkene (Kemp and Vellaccio, 1980). The pi (π) valence electrons of conjugated alkenes are delocalized, and therefore these compounds constitute 'soft' (polarizable) electrophiles that are capable of reacting with 'soft' nucleophiles like cysteine thiol groups on proteins and glutathione (GSH), according to Pearson's "hard-soft acid-bases" (HSAB) model (Pearson and Songstad 1967). In ACR (Fig. 5.1), the pi (π) electrons from the alkene double bond are distorted (polarized) towards the electron-withdrawing carbonyl group, resulting in electron deficiency at the β-carbon (Cβ) atom (LoPachin et al. 2008). As a consequence, although ACR and other type-2 conjugated alkenes are bifunctional (they exhibit electrophilic reactivity at both the carbonyl carbon and the β-carbon atoms), the β-alkene carbon atom is a 'softer' electrophile and is therefore a favorable center for nucleophilic conjugate (Michael-type) addition reactions.

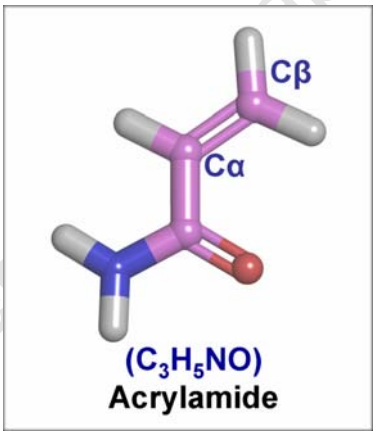


Fig. 5.1: A ball and stick representation of the chemical structure of ACR.
The amide carbonyl-amino group is linked to the Cα=Cβ alkene.

ACR is prepared from the hydrolysis of acrylonitrile by nitrile hydratases, and is extensively used to make polyacrylamide, which is used in various industrial settings as a water-resistant thickener in wastewater treatment, ore processing, and in the manufacture of dyes, paper, cosmetics, glues, permanent press fabrics *et cetera*. In scientific laboratories, ACR is used to prepare polyacrylamide gels for electrophoretic separation of proteins. ACR is also a food toxicant formed in the course of the Maillard reaction (Mottram et al. 2002), which occurs between the carbonyl group of a reducing sugar (such as glucose) and an amine (amino acid, peptide or protein) during high-temperature preparation of some starchy foods like french

fries, chips, or bread, *et cetera*. Although the polymeric form is non-toxic, exposure to monomeric ACR results in neurotoxicity in both humans and animal models with symptoms including ataxia, skeletal muscle weakness and numbness of hands and feet (LoPachin et al. 2002;LoPachin et al. 2003;LoPachin 2004;Smith and Oehme 1991). Other possible harmful effects of ACR include carcinogenicity (Friedman 2005;Rice 2005), genotoxicity (Besaratinia and Pfeifer 2005) and reproductive and developmental toxicities (Dearfield et al. 1988). ACR toxicity has been a major health concern since humans consume a relatively high dose of ACR on daily basis (Friedman 2005) from cooked and processed foods, in addition to exposure through occupation and pollution, all of which have been linked to toxicity of various organs (Bergmark et al. 1993;Perez et al. 1999;Tucek et al. 2002).

Due to the soft electrophilic property of the β-carbon, ACR and other conjugated type-2 alkenes preferentially form Michael-adducts with sulphydryl groups on cysteine residues of numerous cellular proteins, which are the principal soft nucleophilic targets in biological systems (Friedman 2003;LoPachin and Decaprio 2005;Pearson and Songstad 1967). The mechanism for ACR neurotoxicity has been confirmed to involve synaptic nerve terminals through selective formation of Michael adducts with the sulphydryl groups of regulatory synaptic proteins (e.g. N-ethylmaleimide sensitive factor (NSF)), which results in the inhibition of neurotransmitter release and eventual nerve terminal damage in the peripheral and central nervous systems (LoPachin et al. 2007;LoPachin 2004;LoPachin and Barber 2006). Protein cross-linking by α,β-unsaturated carbonyls has also been identified as a possible mechanism for synapsomal toxicity *in vitro* (LoPachin et al. 2007), and it is facilitated by the bifunctional electrophilic property of type-2 alkenes. Initial adduct formation at the β-carbon might be followed by attack from a neighboring nucleophile at the reactive carbonyl carbon atom, resulting in the formation of protein cross-links (Kurtz and Lloyd 2003).

The general mechanism for a Michael addition reaction between the cysteine thiol group and ACR (or its derivatives) proceeds via a tetrahedral intermediate, as shown in the schematic diagram in figure 5.2. A general base catalyst, which could be a glutamate or aspartate residue in enzymes, is often involved. The base deprotonates the thiol group of the cysteine residue, enhancing its nucleophilic ability to attack the β-carbon of a conjugated type-2 alkene. Upon the attack, the double bond opens, resulting in a tetrahedral intermediate that has a negative charge at the β-carbon. Protonation of the intermediate by the base results in a Michael

adduct, which in the case of ACR is a thioether Cys-S-beta-propionamide (Cys-S-PMD). It is worth noting that the *in vitro* alkylation (Michael addition) reactions between cysteine and ACR are usually carried out at alkaline pH conditions, in which case most of the cysteines in the protein or reaction medium are deprotonated (Bordini et al. 2000). In such instances, activation of the cysteine nucleophile by a base would not be necessary.

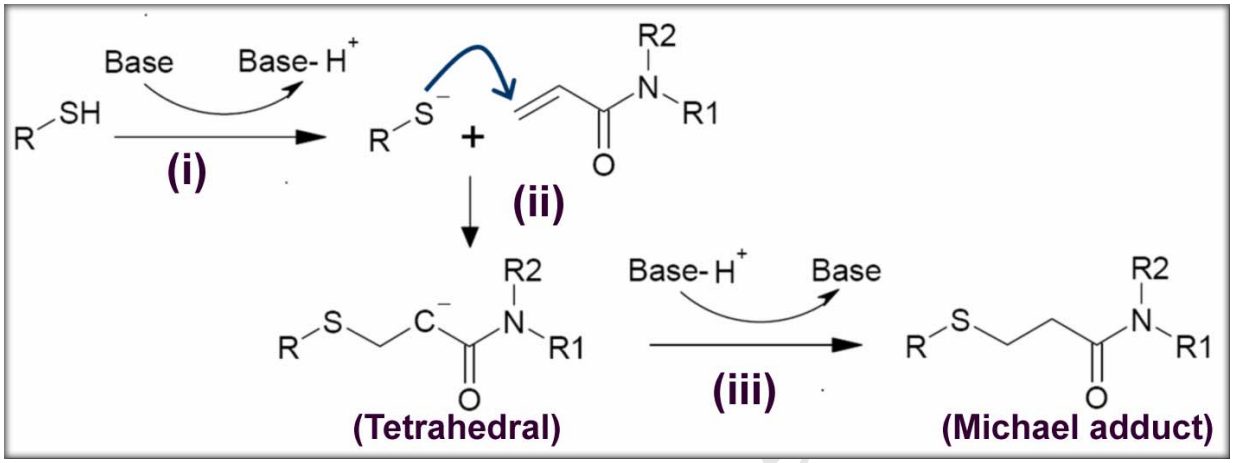


Fig. 5.2: General mechanism of Michael addition of type-2-alkenes to the thiol group of cysteine residues. The sulphydryl group (SH) is activated through deprotonation by a base to a thiolate **(i)**, which then attacks the β-carbon **(ii)** resulting in a tetrahedral intermediate. Protonation of the intermediate by the base **(iii)** results in a thioether Cys-S-PMD Michael adduct.

Other nucleophilic centers targeted by α,β-unsaturated carbonyl derivatives include the amino group of lysine side chains, the hydroxyl group of tyrosines and the amino-terminal groups of polypeptide chains (Bergmark et al. 1993;Bordini et al. 2000;Kaminskas et al. 2005;Perez et al. 1999;Uchida 1999). Michael addition reactions with these other groups are much less prevalent than thiol group addition, and this is mainly because of nucleophilicity incompatibilities, as the amino groups are relatively 'hard' nucleophiles (LoPachin et al. 2007); Michael addition between soft electrophiles and hard nucleophiles is not favored, although it can still occur to some degree (Coles 1984;Hinson and Roberts 1992).

Alkylation of cysteine residues with ACR is one of the many amino acid tagging strategies employed in mass spectrometry-based (MS-based) proteomics for detection, separation, enrichment, quantification, sequence analysis and sequencing of peptides and proteins (Friedman et al. 1965;Giron et al. 2010). It is also popular for identification and quantification of cysteinyl residues using 1-dimensional (1D) and 2D SDS-PAGE (Mineki et al. 2002;Yan et al. 1998). Cysteine modification with ACR is particularly used for 'protection' of cysteine residues during preparation of samples for 2D-PAGE in gel-based proteomics approaches

(Brune 1992), which is a mandatory step as it prevents random regeneration of disulphide bonds, as well as reactions of cysteines with residual non-polymerized ACR monomers in the gels, which might otherwise cause undesirable migration patterns (Bordini et al. 1999;Herbert et al. 2001).

Cysteine bioconjugation with ACR is also an attractive amino acid tagging procedure as it is highly selective under certain reaction conditions, it provides high reaction yields and the thioether derivative adduct formed, Cys-S-beta-propionamide (Cys-S-PMD), is stable during various reactions such as automated Edman degradation and can easily be recovered from protein electrophoresis gels and other aqueous solutions (Brune 1992;Giron et al. 2010). In addition to protein characterization, alkylation of cysteines by ACR has been proposed to be an ideal method for monitoring folding and unfolding of proteins by mass spectrometry (Hamdan et al. 2001). In their *in vitro* folding-refolding experiments with various proteins, Hamdan and colleagues (2001) observed a progressive increase in the ACR alkylation signal by mass spectrometry as the test proteins unfolded under increasingly denaturing conditions.

Aliphatic amidases from branch 2 of the nitrilase enzyme superfamily enzyme exhibit hydrolytic activity on short-chain aliphatic amides only, with propionamide > acrylamide > acetamide being the most rapidly hydrolyzed substrates in decreasing order of preference (Clarke 1970;Fournand et al. 1998a;Makhongela et al. 2007;Skouloubris et al. 1997). These enzymes are attractive candidates for bioremediation of ACR in effluent water from ACR-processing industries and for acrylonitrile detoxification using nitrile hydratase-amidase coupled systems. The active site pockets of branch 2 aliphatic amidases are small, highly constrained and incapable of binding amide substrates that are larger than four carbon atoms. ACR Michael addition at the catalytic cysteine has not been observed in these amidases; mostly probably because the catalytic residues would be highly selective for the amide carbonyl-amino group in addition to the active pocket of branch 2 amidases being highly constrained (Kimani et al. 2007). This would ensure that ACR is positioned in such a way that the catalytic cysteine only has access to the carbonyl carbon, would mostly favor the ACR hydrolysis reaction.

Nesterenkonia sp. amidase (NitN), a unique amidase that was been classified as an aliphatic amidase based on its *in vitro* specificity for short aliphatic amides, has surprisingly negligible activity on ACR (Nel et al. 2011). This is in spite of the fact that PMD, a structural analog of

ACR (Fig. 5.3), is the best-known short aliphatic amide substrate for this enzyme. NitN also exhibits around 20% activity on fluoroacetamide (FAE) relative to its activity on PMD at 100% (Nel et al. 2011). These enzymatic activity data raise a few questions such as: Why is NitN able to hydrolyze PMD and FAE, but not ACR? How does ACR interact with the NitN protein and what would be the factors that prevent nucleophilic attack at the carbonyl carbon atom of ACR by the wild-type (WT) NitN enzyme? The aim of the work reported in this chapter was to use mass spectrometric and X-ray crystallographic methods to understand interactions and reactivity of ACR with the WT NitN enzyme, as well as with the mutants of the two active site glutamate residues (E61Q/L and E139Q) described previously.

Although the substrate binding pocket of NitN is structurally different from the active site pockets of the branch 2 aliphatic amidases described in the previous paragraph, ACR Michael addition at the catalytic cysteine would also not be expected in the NitN WT enzyme. As explained previously, the high affinity of the catalytic residues for the amide carbonyl-amino group would ensure that the β-carbon atom of ACR is not accessible by the catalytic cysteine of NitN.

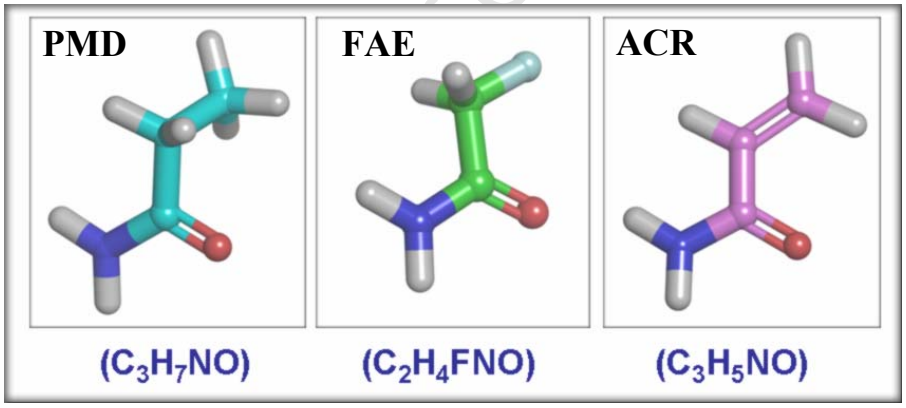
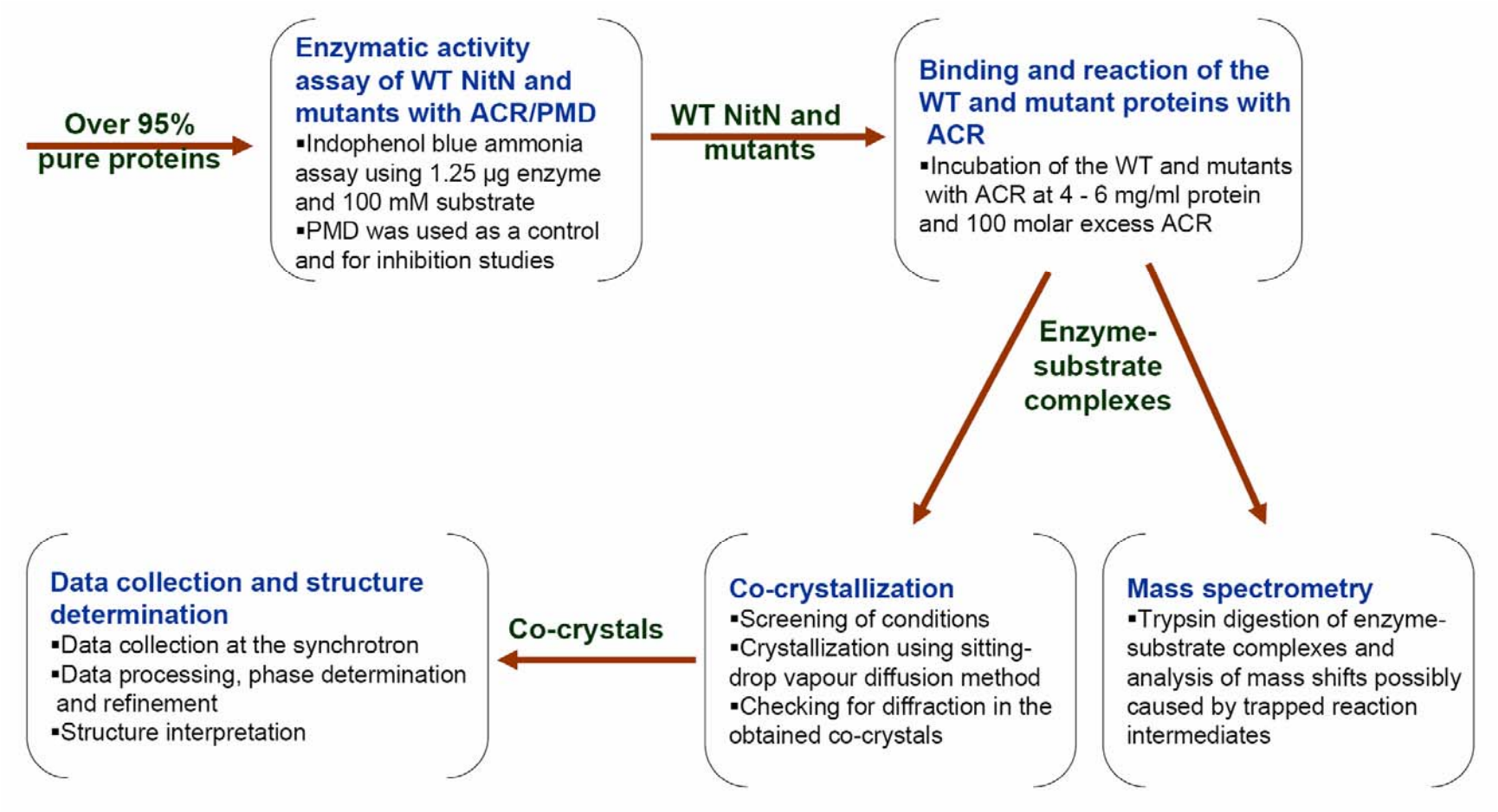


Fig. 5.3: Ball and stick rendition of the chemical structures of propionamide (PMD), fluoroacetamide (FAE) and acrylamide (ACR). The three amides are very similar in terms of the atom numbers, with the only differences being in the Cα-Cβ region and the replacement of the β-carbon atom with fluorine in FAE. PMD is the best-known short aliphatic amide substrate for NitN, and FAE the second best, while NitN has hardly any activity on ACR.

Since the preparation of the two active site glutamate mutants (E139Q and E61Q/L) and their purification has been described in the previous chapter, this chapter will only cover the analytical findings, ranging from the biochemical enzymatic assay, mass spectrometry, to crystallization, structure determination and interpretation of the WT NitN and active site mutant proteins in the presence of ACR.

5.3 Methods - A summary

An outline of the steps followed in the biochemical and mass spectrometric analyses of the WT, E61Q/L, and E139Q NitN mutants with ACR, crystallization and structure determination of NitN co-crystal structures. Preparation of NitN mutant proteins, their expression in *E. coli* cells and purification has been reported previously in chapter 4.0. Details of all experimental procedures are found in the materials and methods section (Chapter 2.0).



5.4 Results and discussion

5.4.1 Enzymatic activity of the WT NitN with ACR

The reaction of the WT NitN with short aliphatic amides was monitored over a 60 minute period, with the amount of ammonia (μM) produced after every 10 minute interval being determined using the indophenol blue ammonia detection method as described in the materials and methods section (chapter 2.0, see section 2.4.3). As expected, there was no ammonia produced when the WT enzyme was incubated with ACR (Fig. 5.4), implying that the WT NitN was incapable of hydrolyzing ACR. To determine whether ACR does bind to the NitN active site and whether it has inhibitory effect on the NitN enzyme, the reaction of the WT enzyme with PMD was monitored both in the presence and absence of ACR. As can be seen in the reaction chart in figure 5.4 below, the presence of ACR in the reaction mixture affected the reaction of NitN with PMD, with the amount of ammonia produced under the same reaction conditions being reduced considerably if ACR was added at the same time as PMD, and reduced even further when the enzyme was pre-incubated with ACR for 30 minutes before PMD was introduced into the reaction mixture. These findings indicate that the WT NitN is indeed inhibited by ACR, and the inhibitory mechanism is likely to be mainly through covalent modification of the catalytic cysteine by ACR or its reaction intermediates. The formation of the potential covalent adduct is however a slow process, as the NitN WT enzyme still retained activity even after 60 minutes of the reaction period. The two substrates (PMD and ACR), when present at the same, are likely to compete for the binding site, which further reduces the chances of NitN binding and hydrolyzing PMD, hence the reduced activity.

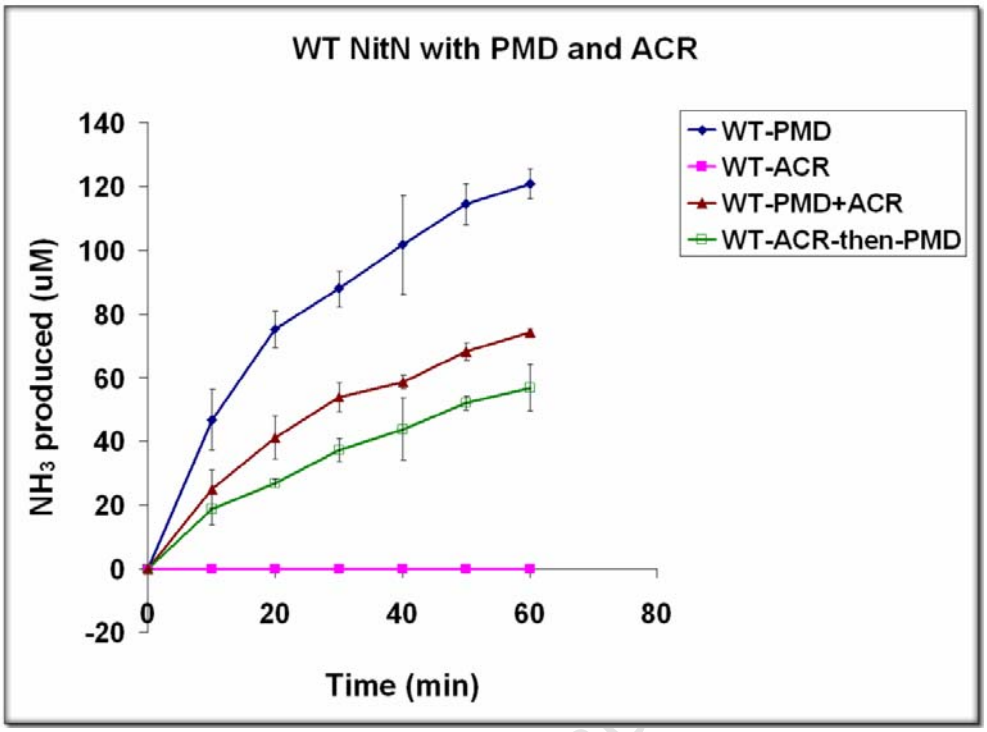


Fig. 5.4: Enzymatic activity plots for WT NitN with PMD and ACR. The error bars indicate standard deviations from the mean of triplicate readings. NitN has no activity on ACR as no ammonia was produced within the 60 min incubation period. The activity of the WT with PMD (WT-PMD) was reduced when ACR was added at the same time as PMD (WT-PMD+ACR), and even further reduced when the enzyme is pre-incubated with ACR for 30 min before PMD

5.4.2 Mass spectrometric characterization

Mass spectrometry experiments of the WT NitN with ACR were conducted in order to check if NitN does react with ACR to form covalent reaction intermediates or adducts that would be trapped in the active site. Although ACR alkylation of nucleophilic cysteine residues is common in thiol proteins, it was not presumed to be a prevalent reaction in NitN, given that ACR would be expected to bind to the NitN active site pocket in the same manner as PMD and BMD in the C165A-PMD and C165A-BMD co-crystal structures (see section 3.4.6.2 and figure 3.7 in chapter 3), so that only the carbonyl carbon would accessible to the catalytic cysteine.

Mass spectrometric analyses of the E139Q and E61Q NitN mutants reacted with ACR were also performed in order to investigate the effect of the mutations to the binding and positioning of ACR in the binding pocket and to probe any reaction adducts that may be formed due to the altered active site environment in the mutant enzymes. While the E139Q NitN mutant had no detectable activity or reaction with any of the short aliphatic amide substrates tested previously (PMD, FAE, BMD and ACE), a change of Glu61 to glutamine or leucine resulted in 'abnormal' S_N2 reaction adducts of FAE at the catalytic cysteine, caused by altered positioning of the substrates in the binding pocket (see figures 4.7 (A) and (B) and section 4.4.8). Incorrect positioning and orientation of ACR in the NitN mutants' binding pockets was speculated to be likely to result in artefactual adducts at the catalytic cysteine, with a high possibility of Michael addition due to the soft electrophilicity property of the β-carbon atom of ACR, as discussed previously.

The formation of possible reaction intermediates and adducts at the reactive cysteine was followed by monitoring the change (increase) in the mass of a 17 amino-acid tryptic fragment bearing the catalytic cysteine (QLSLLV**C**YDVEFPEMVR; 2041.0133 Da). The interpretation of the MS data was achieved through manual comparison of the mass spectra from the reacted proteins with those of the unreacted (apo) mutants and the WT NitN protein. The expected mass shifts and total masses of the tryptic fragments corresponding to the thioester acyl-enzyme intermediate and the Michael adduct of ACR at the catalytic cysteine are presented in table 5.1.

Table 5.1 Expected mass changes in the tryptic fragment bearing the catalytic cysteine (Cys165) if covalent acyl-enzyme intermediates and Michael adducts are trapped in the active sites of the WT, E139Q and E61Q/L NitN mutants.

Reacted or unreacted WT or mutant protein	**Fragment mass for acyl-enzyme intermediate (Da)** [a]		**Fragment mass indicating Michael-type adduct (Da)** [a]	
	Expected mass shift	**Total mass of the tryptic fragment**	**Expected mass shift**	**Total mass of the tryptic fragment**
Apo protein	-	2041	-	2041
Protein + ACR (71 Da)	+54	2095	+71	2112

[a]Masses are rounded to the nearest whole number.

5.4.2.1 Mass spectrometry of the WT NitN with ACR

After incubation of 1 mg/ml (0.0333 mM) WT NitN enzyme with 3.33 mM (100 molar excess) ACR monomers for 2 hours at 25 degrees, the resulting solution yielded the MALDI-TOF mass spectrum in figure 5.5 (A). The spectrum is dominated by the tryptic fragment peak representing unmodified Cys165 at *m/z* 2041, but also has two additional weak signal (minor) peaks of interest: a 2112 Da peak corresponding to an ACR-Cys165 Michael adduct and a 2095 Da peak which represents a possible ACR acyl-enzyme intermediate-containing fragment. The presence of both the ACR thiol adduct and the acyl-enzyme intermediate peaks clearly confirms that: (1) ACR is capable of accessing and binding to the active site pocket of the WT NitN, and (2) the NitN catalytic cysteine is able to carry out a nucleophilic attack at both the carbonyl carbon and β-alkene carbon atoms of the ACR molecule, although at a very low reaction rates, judging from the percentage intensity of the 2112 Da and 2095 Da peaks in relation to the signal of the unreacted Cys165 fragment peak (Fig. 5.5 (A)). The positioning and interaction of ACR with the protein atoms is expected to be similar to the observed positioning of PMD1 and BMD in the C165A-PMD and C165A-BMD co-crystal structures (see section 3.4.6.2, chapter 3), as the carbonyl oxygen has a high affinity for the oxyanion hole and the amino group would interact favorably with the side chain carboxylates of the two active site glutamate residues. Assuming optimal positioning of ACR in the NitN active site and based on these mass spectrometric findings, a number of questions arise: (1) how is the cysteine thiol group able to attack at both the carbonyl carbon and the β-carbon atoms?; (2) since the positioning of ACR in the pocket is expected to place the carbonyl carbon in alignment with the cysteine sulfhydryl group, why is the rate of reaction at this carbon center so slow?; and (3) since the 2095 Da peak in the mass spectra provides evidence that small amounts of ACR acyl-enzyme intermediates are formed when the WT NitN is incubated with ACR, why is the reaction unable to progress beyond this point?

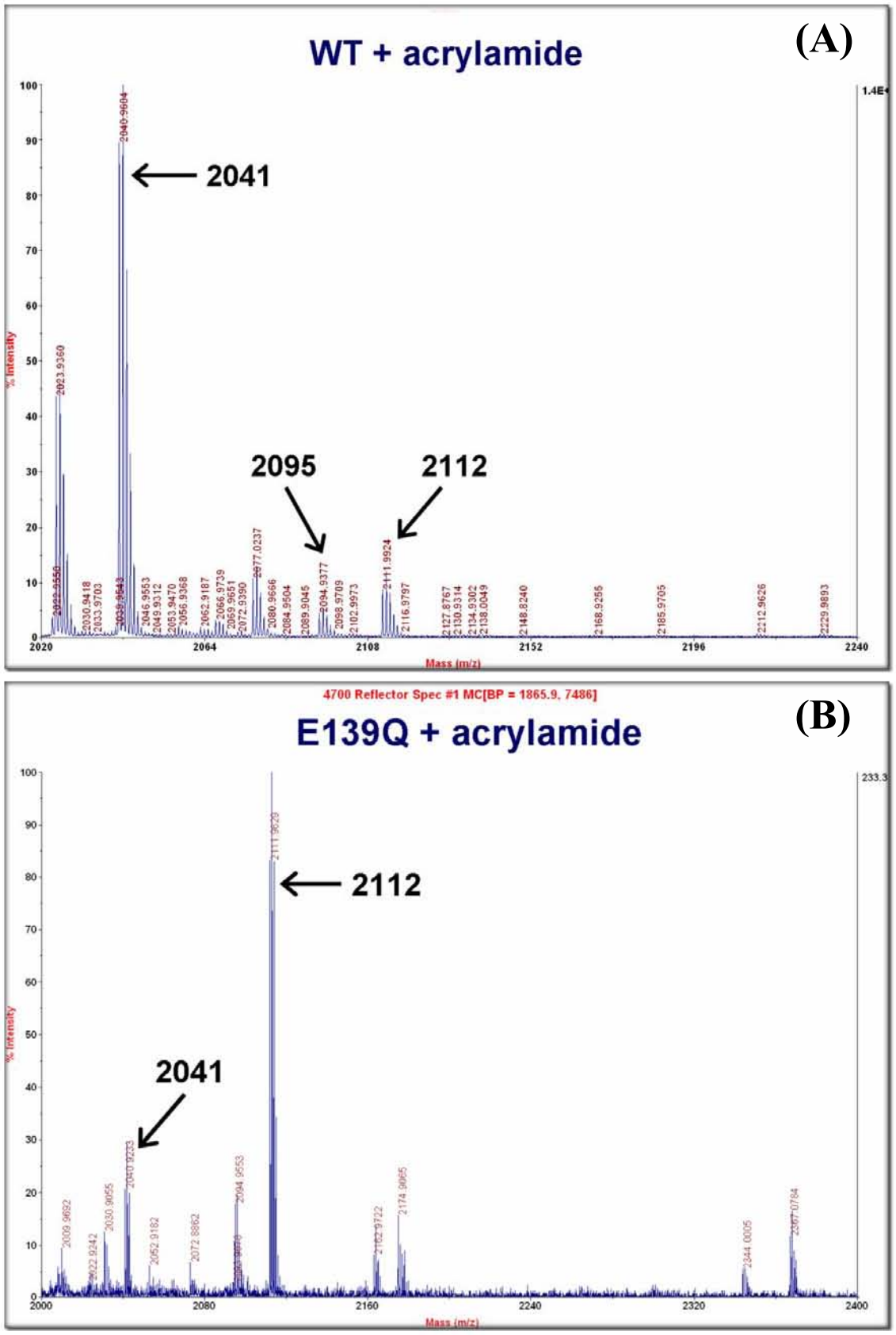


Fig. 5.5: Mass spectra of the WT NitN **(A)** and the E139Q NitN **(B)** after incubation with ACR.

According to the Frontier Molecular Orbital Interactions model (Pearson and Songstad 1967), the lowest unoccupied molecular orbital (LUMO) of the targeted electrophile and the highest occupied molecular orbital (HOMO) of the nucleophile must interact in order for chemical reactions to take place. The ability of the nucleophilic cysteine sulfhydryl group to react with both the carbonyl and β-alkene carbon atoms of ACR indicates that ACR molecules are not adequately restrained in the active site, which would be a consequence of the large size of the NitN binding pocket (see section 1.4.2). As the bound ACR molecules 'wobble' in the active site, the HOMO of the cysteine sulfhydryl group would occasionally coincide with the LUMO of the β-carbon to result in small amounts of ACR-Cys165 Michael adducts, and occasionally with the carbonyl carbon atom to yield small amounts of acyl-enzyme intermediates. The inability of the NitN pocket to adequately restrain short aliphatic amides (PMD, FAE, BMD, ACE and ACR) may explain why the catalytic rates and NitN affinity for these substrates are lower than those observed in other branch 2 aliphatic amidases (Nel et al. 2011).

Based on both the lack of enzymatic activity and the percentage intensity of the ACR acyl-enzyme intermediate peak (*m/z* 2095) relative to that of the unreacted Cys165 fragment peak (*m/z* 2041), it would be reasonable to conclude that unlike branch 2 aliphatic amidases, the WT NitN enzyme has difficulty reacting with the carbonyl carbon of ACR. The reduced potency of the nucleophilic attack at the carbonyl carbon could be attributed partially to the inability of NitN pocket to adequately restrain bound ACR molecules and partly to the weak electrophilic character of the carbonyl carbon center. As discussed previously (section 5.2), the reduced electrophilicity at the carbonyl carbon of ACR (a type-2 conjugated alkene) is due to the ability of the carbonyl group to withdraw electrons from the Cα=Cβ alkene double bond, which makes the β-carbon atom a better ('softer') electrophile instead.

The mass spectrometric detection of trapped ACR thioester intermediate in the WT NitN enzyme (Fig. 5.5 (A)) indicates that the reaction was unable to proceed through to the deacylation phase, or that the deacylation reaction was very slow. This is also thought to be due to the conjugated structure of the acyl moiety of the intermediate, which implies that the electrons at the alkene double bond would be distorted towards the electron-withdrawing carbonyl group, which would make the carbonyl carbon a poorer electrophile and therefore the deacylation water would not be able to attack efficiently.

Given that the reaction of the NitN catalytic cysteine with the carbonyl carbon of ACR is unfavorable, the amount of ammonia released during the formation of the observed acyl-enzyme intermediate would be too small to be detected by the ammonia indophenol blue enzymatic assay employed in this study. While prolonged incubation times might result in more ACR acyl-enzyme intermediates and ACR-thiol adducts being formed, the fact that the active enzyme is not being regenerated may mean that no substantial amount of ammonia would be released or detected even after several hours of incubation. At some stage during the incubation, all the catalytic cysteines in the reaction mixture would become covalently adducted, resulting in inactive enzymes.

The mass spectrometric evidence for the formation of small amounts of ACR-Cys165 Michael adducts and ACR thioester intermediate with the WT NitN (Fig. 5.5) provide a satisfactory explanation for the reduced enzymatic activity of the native NitN enzyme with PMD in the presence of ACR (section 5.4.1 and figure 5.4). The slow (unfavorable) rates of the formation of these covalent adducts and intermediates explain why during the enzymatic activity assay of NitN WT using both ACR and PMD in the reaction mixture, the enzyme still retained some activity regardless of whether the two substrates were added at the same time or the enzyme was pre-incubated with ACR for 30 minutes before PMD was added into the reaction mixture. However since the formation of these cysteine adducts is irreversible, the enzymatic activity would eventually be abolished following prolonged incubation periods with ACR.

5.4.2.2 Mass spectrometry of the E139Q mutant with ACR

The E139Q NitN mutant incubated with ACR under the same reaction conditions as the WT enzyme resulted in the MALDI-TOF spectrum in figure 5.5 (B). The tryptic fragment peak (*m/z* 2112) of the ACR Michael addition to the catalytic cysteine dominates the spectrum, but a weak peak of a fragment corresponding to the free enzyme at *m/z* 2041 is also present. The presence of a more intense Michael adduct peak clearly indicates that the reactive cysteine in the E139Q mutant is preferentially capable of reacting with ACR at the β-carbon and also confirms that the E139Q NitN binding pocket is accessible to the substrates, despite the instability of the loop region where E139Q mutation is located, as observed in the crystal structure (see section 4.4.6.1 and figure 4.12, chapter 4).

The ease of ACR thiol adduct formation in the E139Q protein is most likely a consequence of ACR positioning in the active site pocket of the mutant. Since Glu139 is critical in positioning short aliphatic amide substrates for hydrolysis in the NitN pocket, its replacement would result in incorrect positioning and possible instability of bound substrates. In the case of ACR this places the β-carbon in alignment with the sulfhydryl group of the catalytic cysteine, facilitating the Michael addition reaction.

5.4.2.3 Mass spectrometry of the E61Q/L NitN mutants with ACR

Incubation of the E61Q/L NitN mutants with ACR following the same conditions as the WT enzyme resulted in the mass spectra in figures 5.6 (A) and (B) below. E61Q reacted with ACR to yield 3 fragment peaks of interest (Fig. 5.6 (A)): a dominant ACR-Cys165 Michael adduct peak at *m/z* 2112, a minor peak corresponding to the unbound Cys165 tryptic fragment at 2041 Da and a medium intensity peak that coincides with an ACR acyl-enzyme (thioester) intermediate at *m/z* 2095. E61L NitN also reacted with ACR, yielding the same three peaks of interest (Fig. 5.6 (B)), but in this case the unbound Cys165 peak (*m/z* 2041) was more intense than the ACR-Cys165 adduct peak (*m/z* 2112).

The occurrence of both the ACR thioester intermediate and ACR-Cys165 Michael adduct peaks is a clear indication that the nucleophilic cysteine of the mutant enzymes is able to attack at both the carbonyl carbon and the β-carbon atoms, but the reaction at the β-carbon is more favorable based on the percentage intensities of the Michael adduct peaks in both mutants. This preferential reaction could be attributed to altered positioning of ACR molecules within the pocket of the E61Q/L mutants, which favorably allows stereo-electronic alignment of the β-carbon and the thiol of the catalytic cysteine. Indeed, the co-crystal structure of E61Q/C165A NitN with a non-covalently bound ACR molecule (E61Q/C165A-ACR) revealed twisting and tilting of ACR relative to the BMD molecule in the active site of the C165A mutant (Fig. 5.7; also see section 4.4.9 for details), and it highlighted the critical role of Glu61 in positioning the amide substrates for catalysis. Accompanying these substrate positioning changes is possible ‘unsteadiness’ of the bound substrates, such that as the ACR ‘wobbles’ in the active site, reaction of the cysteine sulfhydryl group with either the carbonyl or the β-alkene carbon atoms takes place whenever the nucleophile is in alignment with the respective electrophile. In addition to inaccurate positioning of ACR, the reaction at the β-alkene carbon is enhanced by the ‘softer’ electrophilic property at this center.

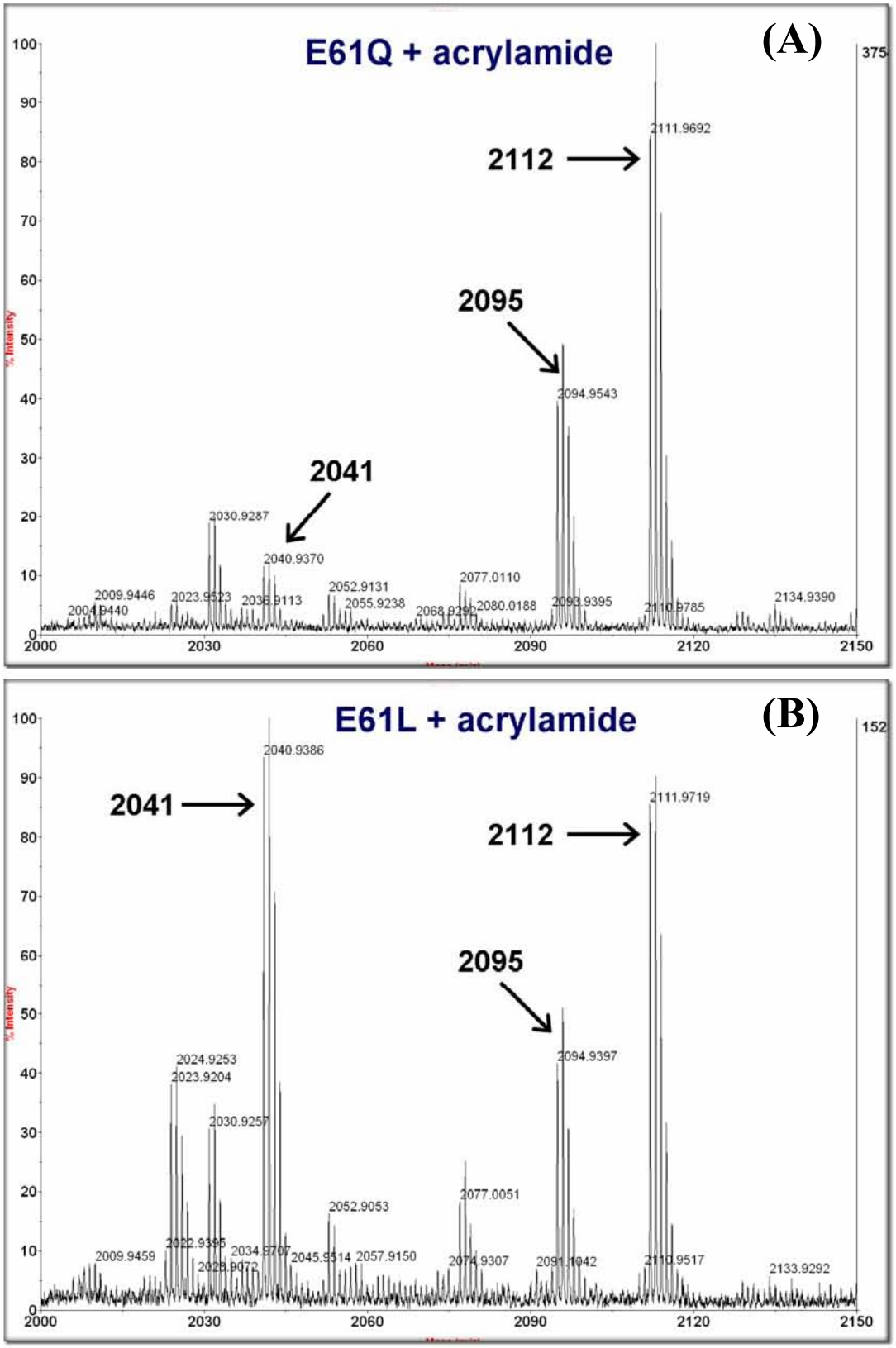


Fig. 5.6: Mass spectra of the E61Q NitN **(A)** and the E61L NitN **(B)** after incubation with ACR.

The E61L NitN mutant appears to be less potent in carrying out the Michael addition reaction as the associated 2112 Da peak is less intense than the corresponding peak in the E61Q mass spectrum (Fig. 5.6). These reaction differences could be attributed to the expected differences in the active site environments of the mutants, as leucine and glutamine side chains have different polarity characteristics. The occurrence of ACR thioester intermediates and covalent Michael adducts at the catalytic cysteine is a clear indication that Glu61 mutants (E61Q/L) are able to react with ACR monomers at pH conditions close to neutral (pH 7.5). This reiterates previous observations that Glu61 (Glu1) is not absolutely required to catalyze nucleophilic attack by the cysteine (section 4.4.11.1) and it is indicative of possible alternative means for the reactive cysteine activation.

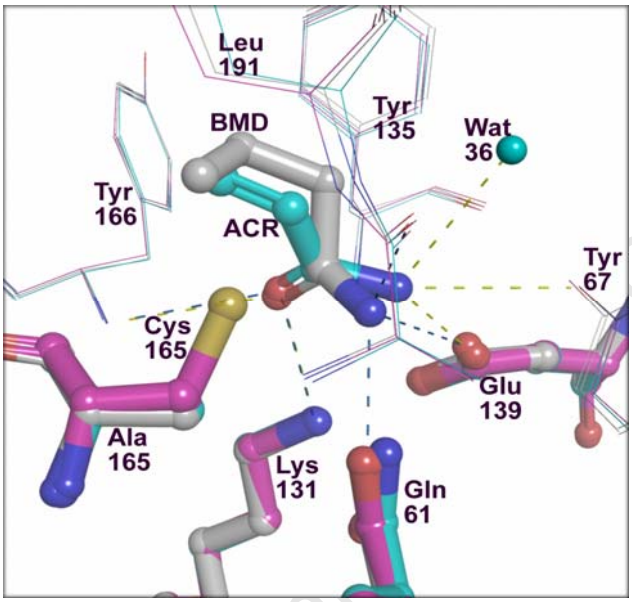


Fig. 5.7: An overlay of the active site region of the E61Q/C165A-ACR (cyan CPK), C165A-BMD (grey CPK) and the WT-apo (magenta CPK) NitN crystal structures demonstrating the change in ACR positioning and orientation as a result of E61Q mutation. ACR is twisted and tilted relative to the BMD, whose amino group is positioned optimally through polar contacts with Glu61 in the C165A NitN structure. The tilting of the ACR molecule allows the cysteine to favorably react with the β-carbon atom resulting in ACR alkylation.

5.4.3 Crystallization of the proteins reacted with ACR

The WT NitN and the E61Q/L and E139Q mutants (at 5 mg/ml) were pre-incubated with ACR monomers (33.33 mM) for 2 hours at room temperature, before crystallization trials were setup using previously identified conditions for the NitN WT and the glutamate mutant proteins. Although the reacted proteins were found to precipitate heavily within the crystallization drops, a few diffraction quality crystals for the WT and the E61Q mutant proteins reacted with ACR (WT-ACR and E61Q-ACR respectively) appeared after 1-2 weeks of incubation in reservoir buffers 39 (0.1 M HEPES sodium pH 7.5, 2% PEG400, 2.0 M

ammonium sulfate) and 38P (0.2 M ammonium tartrate dibasic, 20% PEG3350) respectively. E139Q mutant protein pre-incubated with ACR precipitated heavily, with no success in obtaining crystals even after numerous optimization strategies (including variation of protein concentration and pH and use of additives) and screening of new conditions. E61L NitN mutant protein pre-incubated with ACR (E61L-ACR) was also unstable and no crystals formed in all crystallization conditions that were screened. Co-crystallization of ACR with C165A was also attempted, with the aim of visualizing a non-covalently bound ACR molecule in the NitN binding pocket. A number of crystals were obtained within precipitated protein in two different crystallization conditions, but upon data collection and solution of the structures, none had density corresponding to the ACR in the active site; this implies that only uncomplexed protein was stable enough to pack in the crystals. These datasets were therefore abandoned and are not reported in this chapter.

5.4.4 Determination of structures, model refinement and validation

The diffraction data were collected at the synchrotron facilities in Grenoble France. The datasets for the WT-ACR and E61Q-ACR crystals were processed to the highest possible resolution based on the redundancy and completeness of the data, as well as reasonable R_{merge} values (less than 50%) and (I/σ(I)) values of about 2 in the highest resolution shell. Data collection statistics are presented in table 5.2. Both datasets were of high quality as indicated by high (I/σ(I)) values, low overall merging R-factor R_{merge}) values, data completeness levels approaching 100% and high data redundancy values. Analysis of the merged datasets with the *Xtriage* program (Adams et al. 2010) found no possible outliers among acentric and centric reflections (Read 1999), and no ice rings-related problems were detected in the ice ring-sensitive resolution ranges.The space group and unit cell parameters of the two datasets were the same as those of the WT NitN structure (PDB id, 3hkx) (Nel et al. 2011).

The structures were solved by molecular replacement using the 3-D coordinates of the WT NitN structure as the starting model, and refined using maximum likelihood strategies implemented in both *PHENIX* and *REFMAC5*. The catalytic cysteine (Cys165) was examined in order to identify any continuous difference density corresponding to covalently attached reaction intermediates and/or Michael adducts. Alternating model rebuilding and refinement

cycles were performed until all highlighted model errors had been addressed and when the cross-validation free R-factor (R_{free}) ceased to decrease.

Table 5.2: Data collection statistics

Data collection statistics		
Dataset	**WT-ACR**	**E61Q-ACR**
Crystallization buffer	39[a]	38P[b]
Synchrotron source	BM14[c]	Proxima I[d]
Space group	$C222_1$	$C222_1$
Unit cell parameters	a = 74.820 (Å) b = 115.08 (Å) c = 65.390 (Å) $\alpha=\beta=\gamma$= 90.0°	a = 76.100 (Å) b = 115.45 Å) c = 65.930 (Å) $\alpha=\beta=\gamma$= 90.0°
Cell content analysis		
Unit cell volume (Å^3)	563026.563	579296.625
Vm (Å^3/Da)[e]	2.35	2.41
Molecules in ASU[f]	1	1
Solvent content (%)	47.60	49.07
Resolution range (Å)[g]	17.98 - 1.20 (1.26 - 1.20)	38.05 - 1.40 (1.45 - 1.40)
No. of observed reflections	639597	240562
No. of unique reflections	88066	57302
Completeness (%)[g]	99.9 (100.0)	99.5 (98.6)
Redundancy[g]	7.3 (7.1)	4.20 (4.10)
Mean $(I) /\sigma(I)$[g]	19.4 (5.3)	9.8 (3.2)
R_{merge} (%)[g, h]	5.40 (38.2)	6.6 (38.4)

[a]Crystallization buffer 39 - 0.1 M HEPES sodium pH 7.5, 2% PEG400, 2.0 M ammonium sulfate.
[b]Crystallization buffer 38P - 0.2 M ammonium tartrate dibasic, 20% PEG3350.
[c]BM14 - Macromolecular beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble.
[d]Proxima I - Macromolecular beamline at the Synchrotron Soleil, Paris.
[e]V_m is the Matthew's coefficient, which allows estimation of the number of protein subunits in the asymmetric unit. $V_m = V/(MW*n)$, where V is the volume of the unit cell, MW is the molecular mass of the molecule and n is the number of asymmetric units in the unit cell.
[f]ASU - asymmetric unit.
[g]Values in parentheses are for the highest resolution shell.
[h]$R_{merge} = [\Sigma_h\Sigma_j |I_j(h) - \langle I(h)\rangle|] / \Sigma_h\Sigma_j |I_j(h)|$, where $I_j(h)$ and $\langle I(h)\rangle$ are the jth and the mean measurements of the intensity of reflection h, respectively.

Similar to all the other NitN structures presented in this report, some of the N-terminal amino acids corresponding to the hexa-histidine tag and the linker region were missing in the two structures, and there was also no visible density for the helix region between residues 258 and

268, which may be due to autocleavage or atomic disorder. Aside from the missing residues, the refined structures were of high quality as depicted by the refinement and validation statistics in table 5.3. The final models had low crystallographic R-factor values and small differences between R_{cryst} and R_{free} values, a figure of merit of greater than 0.9 in both cases, Ramachandran plots having all residues falling in the favored and allowed regions, reasonable bond lengths and angles and the model resolution extending beyond 1.5 Å in both cases. The Ramachandran plots for all residues case in the final models are presented in Appendix III.

Table 5.3: Model refinement and validation statistics

Dataset	**WT-ACR**	**E61Q-ACR**
Refinement statistics		
Resolution range	17.98 - 1.20	38.05 - 1.45
No. of reflections in working set (95%)	79439	46483
No. of reflections in test set (5%)	4414	2640
No. of atoms		
Protein[a]	2088	1973
Ligand/ion	22	15
Water	308	333
Missing residues	A1-12 A259-267	A1-19 A258-267
R_{cryst} (%)[b]	12.88	14.22
R_{free} (%)[c]	14.97	18.03
Overall figure of merit	0.940	0.912
Average B value (Å^2)	13.68	17.69
Model validation		
Rms[d] deviations from ideality		
Bond lengths (Å)	0.03	0.03
Bond angles (°)	2.25	2.26
Bad rotamers (%)	1.00	0.50
Ramachandran outliers (%)	0.00	0.00
All-atom clashscore	5.10	3.97

[a]Number of protein atoms include duplicated atoms in alternate conformations.
[b]$R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$ where F_o and F_c are observed and calculated structure factor amplitudes of reflection h of the working set of reflections, respectively.
[c]R_{free} is equal to R_{cryst} for h belonging to the test set of reflections.
[d]Rms - root mean square.

5.4.5 The WT-ACR crystal structure

The mass spectrum of the WT NitN reacted with ACR had three fragment peaks of interest, representing three reaction states of the catalytic cysteine (Cys165): a large portion of unreacted (free) cysteine, as well as small amounts of the ACR thioester intermediate and the ACR-Cys165 Michael adduct. On the determination of the crystal structure of the WT reacted with ACR (WT-ACR), it was immediately clear that the thiol group of the catalytic cysteine adopted at least two different conformations, and continuous positive difference density that was attached to the Sγ atom was visible (Fig. 5.8), indicating covalent modification. The extra density was however difficult to interpret as it did not correspond to a specific modification state of the catalytic cysteine, but instead appeared to be a mixture of states, and therefore required a systematic modeling of various combinations of possible reaction states and monitoring the fit to the resultant electron density maps after refinement.

In first scenario, the two conformers of the Cys165 thiol group were assumed to be modified to an ACR Michael adduct and a Cys-sulfinic acid respectively. Upon refinement, a large blob of negative difference density (Fig. 5.9 (A)), which gives an indication of atoms that have been modeled in the wrong place, ruled out the possibility of a Cys-sulfinic acid modification. The ACR Michael adduct however seemed to refine relatively well at low (0.3) atomic occupancy. The possibility of visualizing all three reaction states as depicted by mass spectrometry (MS), that is, an ACR thioester acyl-enzyme intermediate, an ACR-Cys Michael adduct and a free cysteine (with three closely associated water molecules) was explored next and the resultant electron density map after refinement is shown in figure 5.9 (B). This combination of the reaction states was also ruled out due to the occurrence of negative difference density, which indicated that the thiol of the free cysteine had been placed in the wrong place. This could imply that all the three reaction states of the catalytic cysteine were not present in the crystal, or one of the states had very low atomic occupancy. Modeling of a combination of the ACR thioester acyl-enzyme intermediate and the ACR-Cys165 Michael adduct was also attempted, although it was not a realistic interpretation since the MS data had clearly showed that the unbound (free) protein constituted the bulk of the WT NitN and ACR reaction solution. The ACR-Cys165 Michael adduct model agreed well with the measured electron density, but a blob of negative difference density around the Sγ-carbonyl carbon bond of the other adduct ruled out the possibility of having an ACR thioester intermediate in

the crystal (Fig. 5.9 (C)). Finally, a more realistic model was obtained when an ACR Michael adduct at 0.3 atomic occupancy and a free cysteine together with associated water molecules having 0.7 atomic occupancy were modeled (Fig. 5.9 (D)). It is likely that the proportion of the thioester acyl-enzyme intermediate component of the ACR reaction mixture was either too low to appear in the density or that it may have precipitated out before or during crystallization.

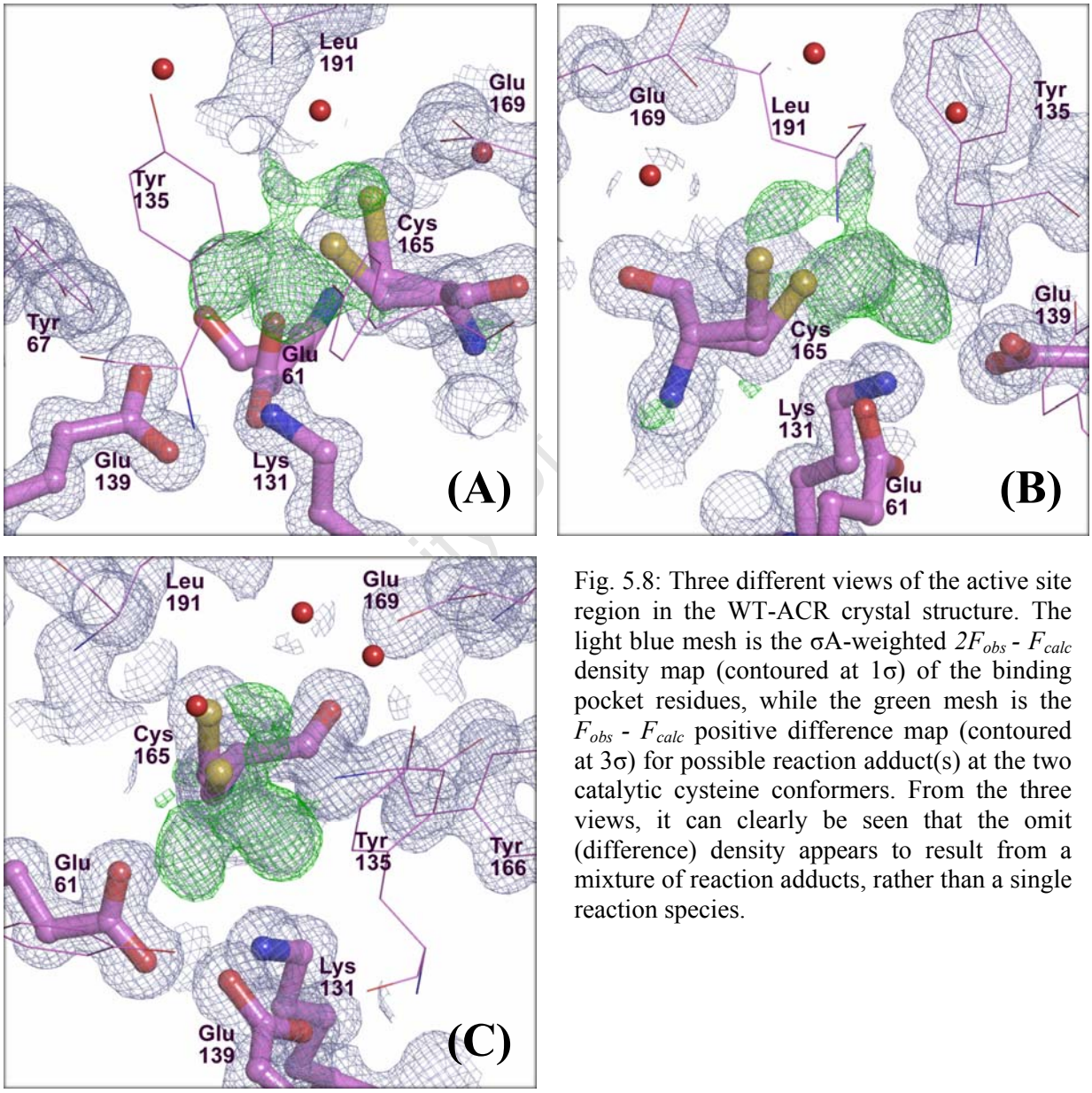


Fig. 5.8: Three different views of the active site region in the WT-ACR crystal structure. The light blue mesh is the σA-weighted $2F_{obs}$ - F_{calc} density map (contoured at 1σ) of the binding pocket residues, while the green mesh is the F_{obs} - F_{calc} positive difference map (contoured at 3σ) for possible reaction adduct(s) at the two catalytic cysteine conformers. From the three views, it can clearly be seen that the omit (difference) density appears to result from a mixture of reaction adducts, rather than a single reaction species.

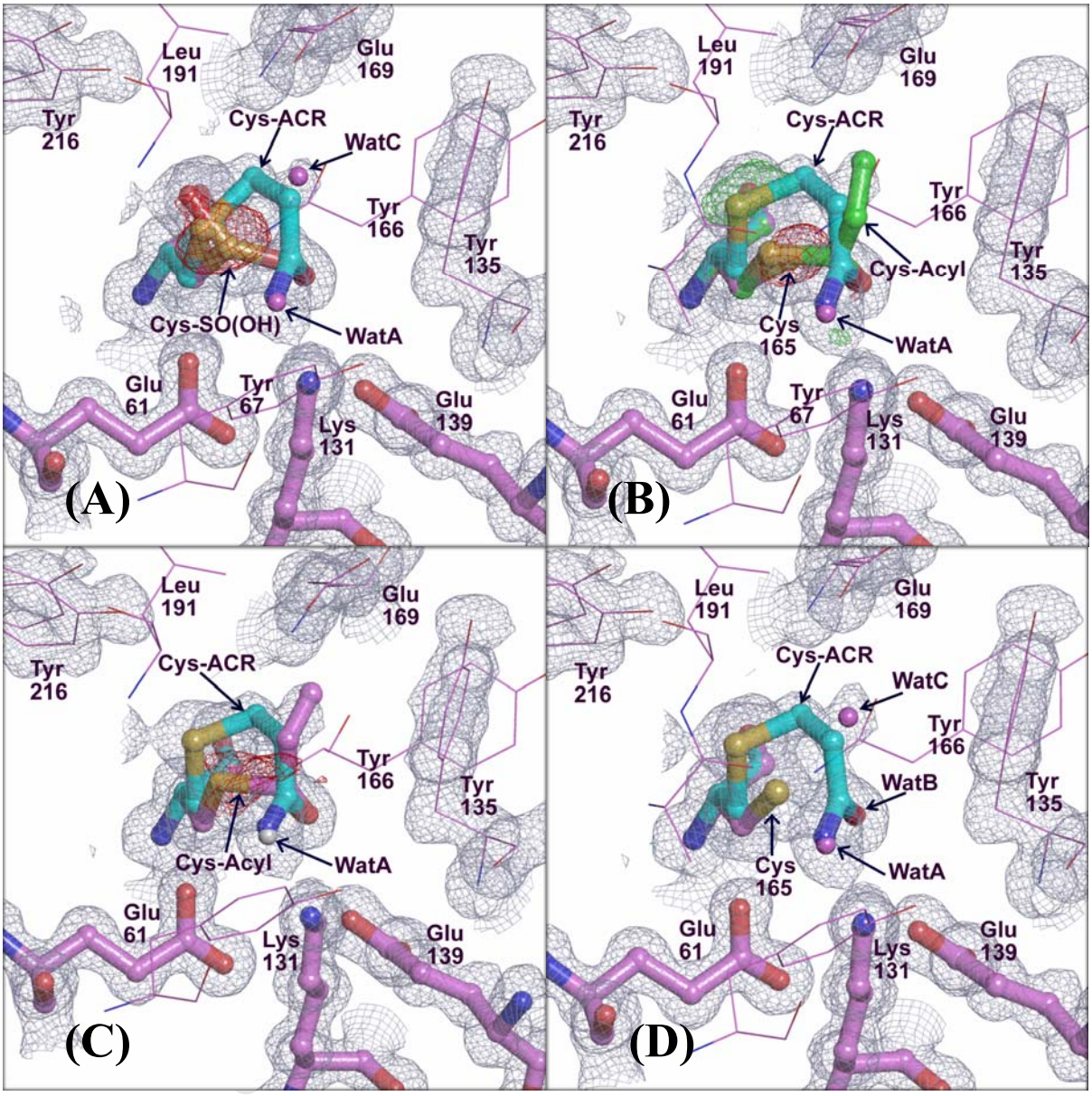


Fig. 5.9: Possible reaction adducts modeled at the catalytic cysteine of the WT-ACR NitN structure. σA-weighted $2F_{obs}$ - F_{calc} density maps are rendered as grey mesh (contoured at 1σ), F_{obs} - F_{calc} positive difference maps are shown as green mesh (contoured at 3σ) while negative difference maps are rendered as red mesh (contoured at -3σ). **(A)** A combination of Cys165-Sulfinic acid (Cys-SO(OH); violet CPK) and Cys-ACR Michael adduct (cyan CPK) modifications at the two Cys165 conformers. **(B)** Modeling three reaction states of Cys165 as free cysteine (Cys165; violet CPK), ACR thioester intermediate (Cys-Acyl; green CPK) and Cys-ACR Michael adduct (cyan CPK). **(C)** The Cys165 conformers are modeled as a Cys-ACR Michael adduct (cyan CPK) and an ACR thioester intermediate (Cys-Acyl; violet CPK). **(D)** The modification at Cys165 is interpreted as a small percentage of Cys-ACR adduct (Cyan CPK) and a free cysteine (violet CPK) that is associated with three water molecules, which are rendered as non-bonded spheres and labeled accordingly. This was taken as the best option.

Thus following various attempts to fit all possible combinations of the reaction states in the WT-ACR electron density map, the modification at the catalytic cysteine was deduced to comprised small percentage of Cys165-ACR Michael adduct and a free cysteine that was associated with three coordinating water molecules. In this interpretation, the carbonyl oxygen atom of the Cys165-ACR Michael adduct is located in the oxyanion hole, where it hydrogen bonds with the backbone NH group of Tyr166 and the side chain amino group of Lys131. The amide amino group of the adduct makes polar contacts with the OE_2 atom of Glu61 and the backbone carboxyl oxygen of Ala190 (Fig. 5.10 (A) and (B)). In the unreacted (free) cysteine protein species, the water molecule in position B (WatB) is in an identical location to the carbonyl oxygen of the Cys165-ACR adduct, where it is involved in the oxyanion hole polar interactions network, while the water molecule in position A (WatA) is coordinated between the side chains of the two active site glutamates and the thiol group of the free cysteine conformer. The water molecule in position C (WatC) interacts with both WatA and WatB, as well as the -SH group of Cys165. Details of these interactions are shown in figures 5.10 (A) and (C). The active site region of the unbound (free) protein species in the WT-ACR structure is very similar to that of the WT-Apo NitN crystal structure, with the hydrogen bonding network as well as the positions of the three active site water molecules being conserved in the two structures (Fig. 5.10 (D)). This is a confirmation that the interpretation of the WT-ACR active site electron density map is correct.

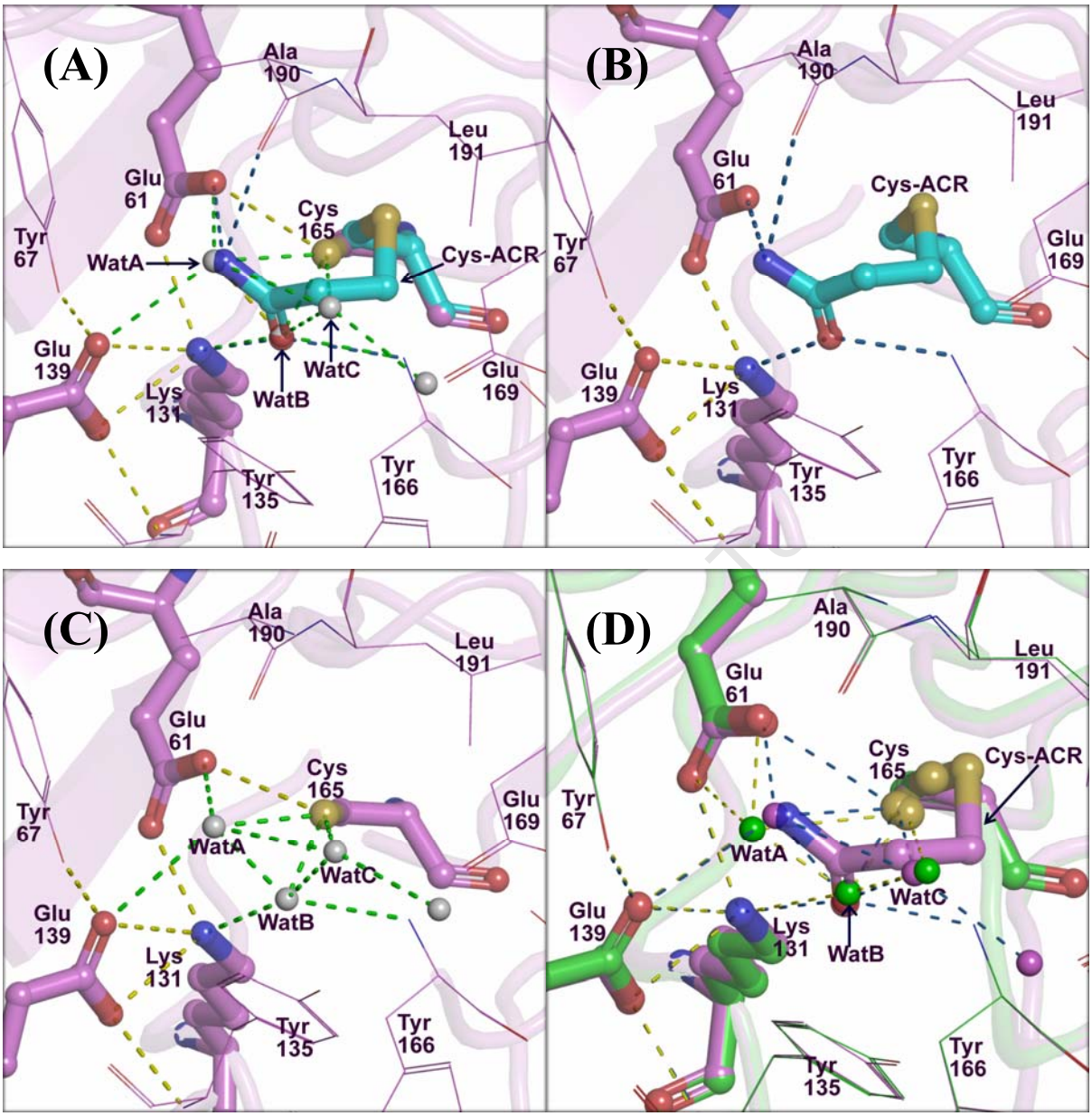


Fig. 5.10: Coordination and polar interactions in the active site region of the WT-ACR NitN crystal structure. **(A)** The active site region showing the two protein species in the crystal; the free cysteine conformer (Cys165) with associated water molecules (grey non-bonded spheres) and the Cys-ACR Michael adduct (cyan CPK ball and sticks). Dashes are possible polar interactions involving the Michael adduct (blue dashes), water molecules (green) and protein atoms. **(B)** Rendition of the Cys-ACR adduct and the active site residues showing the interactions involving the adduct and the protein atoms (blue dashes) and those involving protein atoms only (yellow dashes). **(C)** Rendition of the free cysteine protein species active site. Interactions involving the waters are shown as green dashes. **(D)** An overlay of the WT-ACR (violet CPK) and the WT-Apo (green CPK) NitN active sites. Green non-bonded spheres and yellow dashes are water molecules and contacts in the WT-Apo structure, while violet non-bonded spheres and blue dashes are water molecules and contacts respectively in the WT-ACR active site.

5.4.5.1 ACR alkylation of Cys73

During the refinement of all NitN structures that were crystallized in the presence of ACR, it was observed that Cys73 was also alkylated by ACR to form a Cys73-ACR Michael adduct (Fig. 5.11 (A)) that was also detected by mass spectrometry. In addition to the fragment peaks corresponding to the reactions at the catalytic cysteine (Cys165), the WT NitN protein incubated with ACR yielded two extra peaks of interest (Fig. 5.12): a 3894 Da peak corresponding to an unmodified 37 amino-acid fragment where Cys73 is located (Table 4.1 (A) in chapter 4) and a 3965 Da peak, which is equivalent to 3894 + 71 Da and represents Michael addition of ACR (71 Da) to Cys73. Cys73 is located on a small helix that is close to the surface of the NitN monomeric structure, with its side chain docking into a solvent-accessible space between symmetry-related molecules (Fig. 5.11 (B)). The sulfhydryl group of Cys73 has also been observed to be highly susceptible to modification through oxidation (see the E61L-ACE structure in section 4.4.10.2 and figure 4.22 in chapter 4).

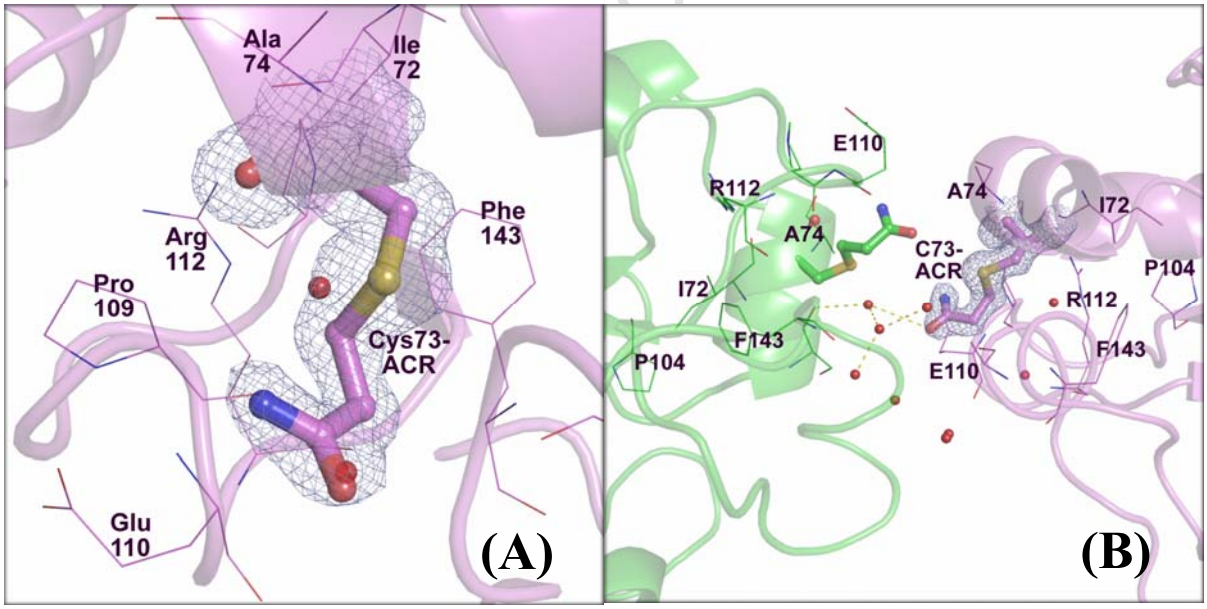


Fig. 5.11: Alkylation of Cys73 by ACR in the WT-ACR NitN structure and the location of the modified Cys73 in the monomeric structure. **(A)** Rendition of the Cys73-ACR adduct and the neighboring residues. The modeled Cys73-ACR adduct fits well in the σA-weighted $2F_{obs}$ - F_{calc} density map, rendered as light purple mesh and contoured at 1σ. **(B)** Parts of the two symmetry-related WT-ACR NitN molecules (green and violet CPK) showing the location of the Cys73-ACR adducts from each monomer, in the solvent-filled space between the two symmetry-related monomers.

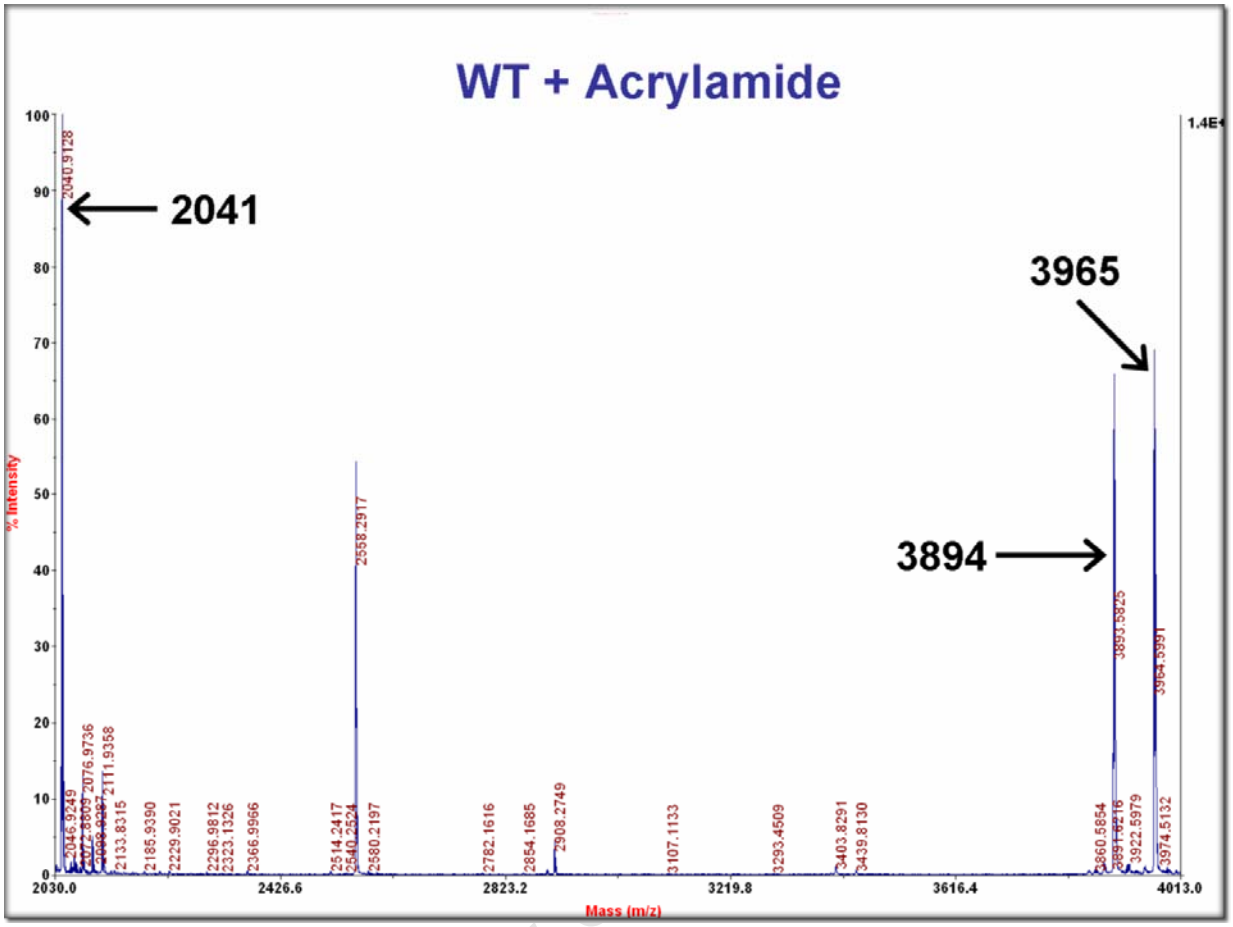


Fig. 5.12: A section of the mass spectra of the WT NitN after incubation with ACR, showing the 3894 Da and 3965 Da peaks corresponding to the unmodified and ACR alkylated tryptic fragments bearing Cys73 respectively.

The occurrence of ACR alkylation at Cys73 raises questions about the mechanism of the Michael addition reaction, particularly of Cys73 thiol group activation in an environment that lacks the active site infrastructure. At the pH of around 7.5 where all the reactions and most crystallization experiments of NitN with ACR were carried out, the sulfhydryl group of cysteine residues in the protein would be protonated. The ability of Cys73 to react with the β-carbon of ACR supports the involvement of a Michael addition mechanism that does not require a base, or one that utilizes a readily available catalyst such as water.

5.4.6 The E61Q-ACR crystal structure

Similar to the mass spectrum of the WT NitN reacted with ACR, the mass spectra of the E61Q/L NitN mutants incubated with ACR had three fragment peaks representing three reaction states of the catalytic cysteine: the free (cysteine) protein, ACR thioester intermediate and ACR-Cys165 Michael adduct-bearing protein. However, unlike the WT-ACR spectrum that was dominated by the unbound protein peak, the E61Q-ACR spectrum was dominated by the Cys165-ACR alkylation adduct fragment peak (*m/z* 2112; Fig. 5.6), indicating an increased potency of the catalytic cysteine Michael addition reaction in the E61Q mutant. Crystallization of the E61Q NitN protein pre-incubated with ACR yielded a few diffraction-quality crystals, one of which resulted in a structure that had clear density corresponding to a Cys165-ACR Michael adduct (Fig. 5.13 (A) and (B)). The other reaction states observed by MS (the free protein and the ACR thioester intermediate) were not present in the crystal, probably due to the instability of these protein forms, resulting in their precipitation during crystallization.

The Cys165-ACR adduct in the E61Q-ACR structure is stabilized by four key interactions to protein atoms and an active site water molecule. As expected, the carbonyl oxygen atom is located in the oxyanion hole where it interacts with the side chain amino group of Lys131 and the backbone NH group of Tyr166, while the amide amino group of the adduct hydrogen bonds to the OE_1 carboxyl oxygen atom of Glu139 and a water molecule (Wat327) (Fig. 5.13 (C)). Due to the mutation of Glu61 to glutamine, the Cys165-ACR Michael adduct in the E61Q-ACR structure adopts a different conformation relative to the Michael adduct at Cys165 in the WT-ACR structure (Fig. 5.13 (D)), as a result of the differences in the positioning of the amide amino group in the two structures. In the WT-ACR structure, the amide amino group makes polar contacts with the OE_2 carboxyl oxygen of Glu61 and the backbone carboxyl oxygen of Ala190.

The crystal structure of E61L-ACR could not be determined as the reacted protein was unstable and only formed thick precipitates in the crystallization drops with no crystals growing even after prolonged incubation. Unlike the case of the E61Q NitN mutant, the Michael addition reaction between Cys165 and ACR is less favorable in the E61L NitN mutant, as most of the protein in the reaction mixture was unreacted according to the MS data

(Fig. 5.6). This may explain the observed instability and the inability of the E61L-ACR NitN protein to form crystals.

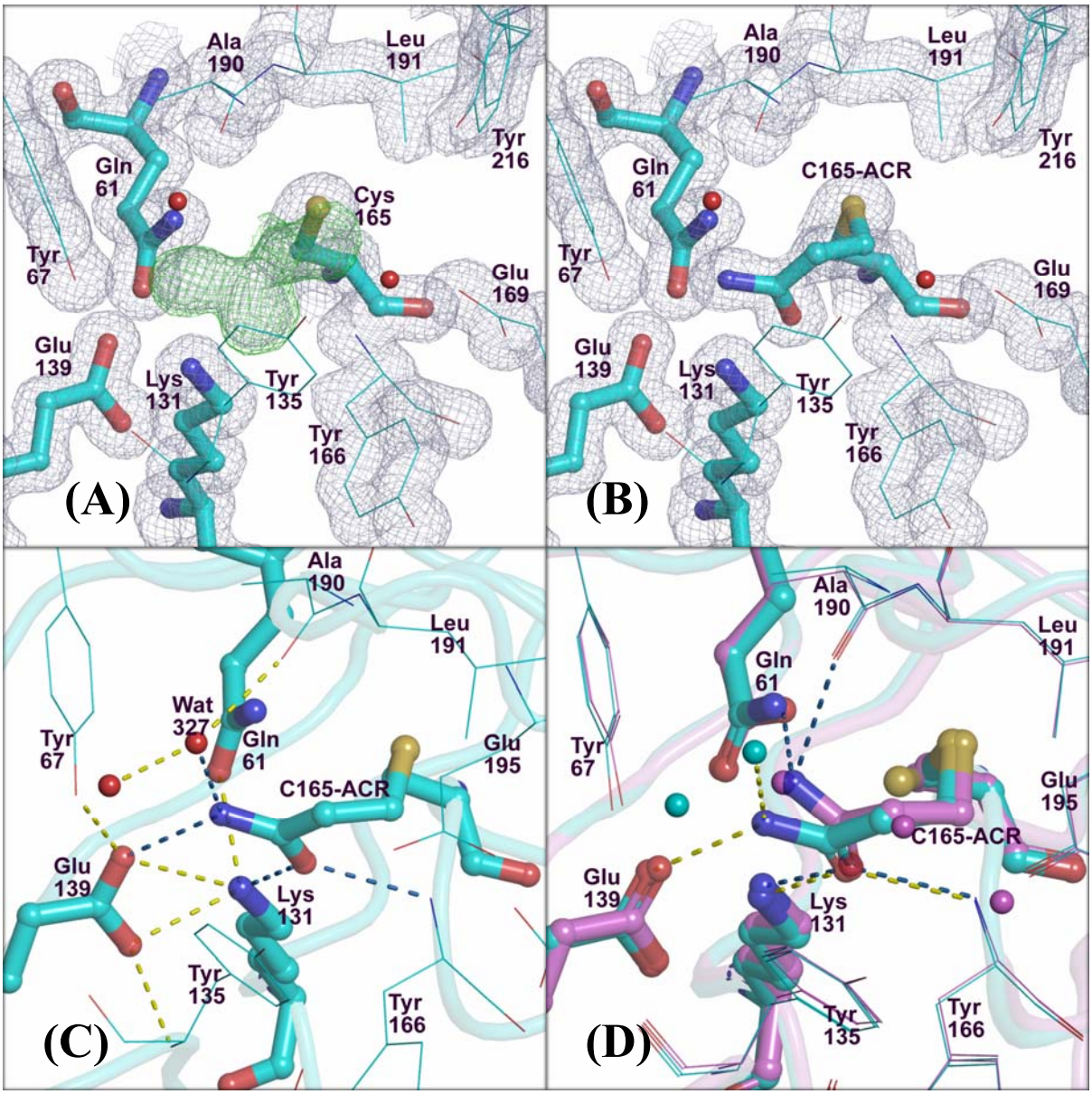


Fig. 5.13: The Cys165-ACR adduct in the E61Q-ACR NitN structure. **(A)** The σA-weighted $2F_{obs}$ - F_{calc} density map (grey, contoured at 1σ) for the active site residues and the σA-weighted F_{obs} - F_{calc} positive difference map (green mesh, contoured at 3σ) showing clear density for the omitted ACR adduct at Cys165. **(B)** The final σA-weighted $2F_{obs}$ - F_{calc} density map (grey, contoured at 1σ) for the active site residues and the refined Cys165-ACR adduct. **(C)** Active site configuration and polar interactions in the E61Q-ACR active site. Red non-bonded spheres are water molecules, blue dashes are interactions that stabilize the Cys165-ACR adduct and yellow dashes are potential hydrogen bonds involving active site residues and waters. **(D)** An overlay of the active site regions in the WT-ACR (violet CPK) and the E61Q-ACR (cyan CPK) structures. Violet non-bonded spheres are water molecules and blue dashes are interactions involving the Cys165-ACR in the WT-ACR structure. Cyan non-bonded spheres are water molecules and yellow dashes are polar contacts involving Cys165-ACR in the E61Q-ACR structure.

5.4.7 The proposed Michael addition mechanism

The ability of the E61Q/L mutants to form the observed ACR acyl-enzyme intermediates and Michael adducts is indicative of possible alternative means for the reactive cysteine activation. As explored previously (section 4.4.11.1), the involvement of Glu139 in the catalyzing the nucleophilic attack by Cys165 is doubtful, based on the location of its carboxylate relative to the thiol group of Cys165 (See figure 4.23 in chapter 4). A water molecule might be a likely candidate for a general base catalyst, although this is also unlikely based on the limited space in the active site pocket upon substrate binding.

Similar to the mechanisms that we proposed previously for the S_N2 elimination reaction of the E61Q/L mutants with FAE (see figure 2.24 in chapter 4), we have proposed two alternative mechanisms for ACR Michael addition at the catalytic cysteine (Cys165) in the E61Q/L mutants (Fig. 5.14) as well as at Cys73. Figure 5.14 panel (A) shows a mechanism proposing the involvement of a base (X). Since there is no readily available base candidates in the protein both in the active site and in the environment of Cys73, we have postulated that a water molecule could function as a base catalyst (X), in which case HX in figure 5.14 (A) would be a hydronium ion. In an alternative mechanism (Fig. 5.14 (B)), we have proposed a concerted reaction scheme that involves deprotonation of the thiol group by the α-carbon atom of ACR and simultaneous nucleophilic attack at the β-carbon atom by the cysteine Sγ atom, resulting in the formation of the Cys-ACR adduct in a single step. Based on the geometry of the E61Q mutant active site and the positioning of the bound ACR molecules (see figure 5.7 above), it is possible for the cysteine S-H bond to align parallel to the double bond of the bound ACR. A nucleophilic attack on the terminal β-carbon would result in an increase in electron density at the α-carbon, directly adjacent to and stereo-electronically aligned with the thiol hydrogen. This would facilitate abstraction of the thiol proton by the α-carbon atom of ACR in concert with the nucleophilic attack, and as the double bond opens, a covalent bond linking the β-carbon of ACR and the Sγ atom of the cysteine is formed. Similarly, we propose a substrate-assisted mechanism for the nucleophilic attack at the carbonyl carbon leading to the formation of ACR thioester intermediates in the E61Q/L mutants. Following optimal stereo-electronic alignment, the amide amino group would abstract the thiol proton simultaneously as the cysteine attacks at the carbonyl carbon. These mechanisms will need to be verified.

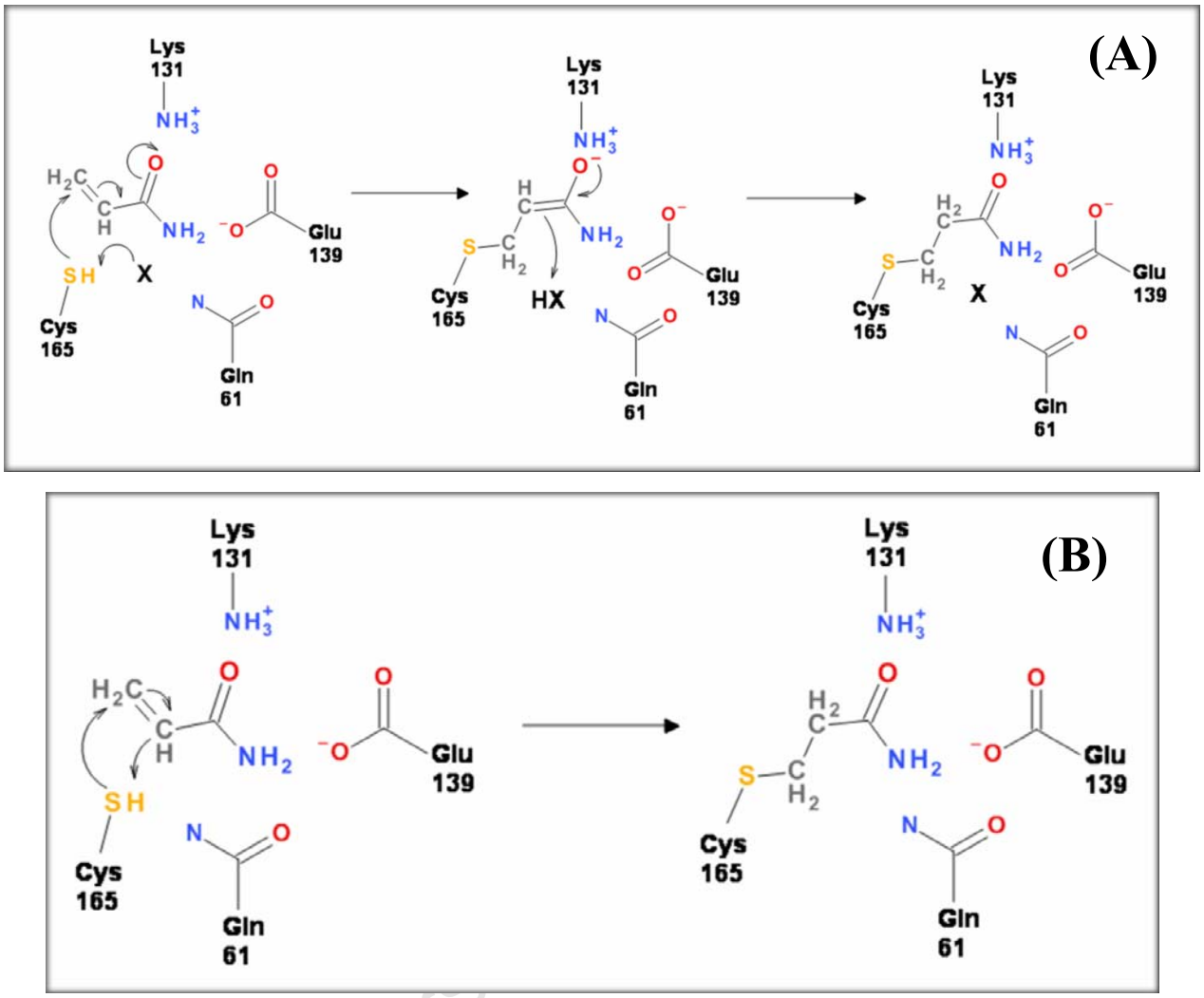


Fig. 5.14: The alternative mechanisms that are proposed in this work for the ACR Michael addition reaction at the active site cysteine (Cys165) in the E61Q/L NitN mutants and Cys73 that is located on the surface of the molecule. **(A)** A Michael addition reaction scheme involving a possible base (X) for the activation of the cysteine nucleophile. We have proposed a water molecule to be the base; hence HX would be a hydronium ion. **(B)** A concerted mechanism where the deprotonation of the cysteine is proposed to be substrate-assisted. The α-carbon atom would abstract the thiol hydrogen in concert with the nucleophilic attack, resulting in the formation of the Cys-ACR adduct in one step.

Since the ACR alkylation reaction is observed in WT NitN as well as the E61Q/L and E139Q NitN mutants, the potency of this reaction is suggested to be determined by the positioning of the ACR molecules in the binding pocket of the respective protein. Cys165-ACR adduction is least favorable in the WT NitN enzyme, where the presence of the two active site glutamates positions the ACR molecules in a manner that discourages nucleophilic attack at the β-alkene carbon atom. This reiterates the importance of the catalytic glutamates in the positioning the amide substrates for catalysis.

Chapter 6

Conclusions

6.1 Study objectives revisited

The overall aim of this work was to understand the catalytic process and substrate specificity in the recently characterized aliphatic amidase from the *Nesterenkonia* species (NitN). Despite the lack of knowledge of the natural substrates for NitN, this research was hoped to provide a detailed understanding of the enzyme's activity, which would enhance its development as an industrial biocatalyst. The study sought to address three specific objectives as outlined below:

- **To probe the structural determinants of NitN specificity on short aliphatic amides** - this was meant to provide a detailed understanding of the binding and interactions of short aliphatic amides with NitN, thereby allowing evaluation of the potential for improving NitN activity on these substrates or identifying and designing better substrates. It was also aimed at providing insights into the positioning of aliphatic amides with more than one carbon atoms within the active site pocket, since the only enzyme-substrate co-crystal structure of an aliphatic amidase that is available is that of a formamidase with formamide (with one carbon atom).

- **To evaluate the catalytic role of the two conserved active site glutamate residues, one of which has been neglected until recently** - this was aimed at uncovering and clarifying the details of the catalytic mechanism not only in NitN, but also in other nitrilase superfamily amidases. It was hoped that the detailed understanding of the catalytic role of the two glutamates, and hence the catalytic process, would generate information that would benefit design of future inhibitors directed at medically-relevant amidases.

- **To investigate the activity and reactivity of NitN with ACR** - this was aimed at understanding why NitN lacks enzymatic activity on ACR, even though it is capable of hydrolyzing similarly-sized short aliphatic amides.

Site-directed mutagenesis, combined with mass spectrometric and X-ray crystallographic techniques were chosen to achieve these goals. Co-crystallization of the catalytic cysteine (Cys165) NitN mutants with a range of short aliphatic amides sought to trap non-covalently

bound substrate molecules in the active site pocket, while mass spectrometric analysis of the two active site glutamates (Glu61 and Glu139) NitN mutants in the presence of short aliphatic amide substrates sought to detect any covalently bound reaction intermediates trapped within the active site as a result of the mutation, hence enabling inferences to be made about the catalytic roles of the two glutamate residues. X-ray crystallography was used as a tool for visualizing the reaction adducts identified by mass spectrometry.

The achievements presented in this work can be summarized as follows:

6.2 Insights into substrate binding and NitN specificity profile

The catalytic cysteine mutant models for substrate binding studies

Two catalytic cysteine NitN mutants, C165S and C165A, have successfully been generated using site-directed mutagenesis, over-expressed in *E. coli* BL21 (DE3) cells and purified to over 95% purity and homogeneity using a two-step purification protocol. The mutants lacked enzymatic activity as expected, but they had similar stability properties to the WT NitN protein, and were therefore suitable for co-crystallization studies.

The C165A NitN mutant has been found to be an excellent model for substrate binding studies as the active site environment of the unbound C165A (C165A-Apo) structure is preserved relative to that of the WT NitN enzyme. The C165S NitN mutant on the other hand is a poor model for substrate binding studies as the orientation of the Ser165 side chain alters the active site configuration, and it partially blocks the binding pocket thereby preventing binding of the short aliphatic amides at the bottom of the pocket (catalytic sub-site).

The binding of short aliphatic amides to NitN

The co-crystal structures of C165A NitN in complex with BMD and PMD (C165A-BMD and C165A-PMD) have been determined. Analyses of these crystal structures have revealed a large unconstrained active site pocket that binds two short aliphatic amide molecules back to back, with one molecule occupying the catalytic sub-site at the bottom of the pocket and the second molecule binding to a second sub-site close to the mouth of the pocket. This suggests that NitN would be capable of binding and hydrolyzing larger amide substrate molecules. The ability to bind two aliphatic amides has been suggested to be a mechanism that enhances

catalytic efficiency of NitN on these substrates, owing to the accompanying stabilization of both the pocket and the substrate molecule at the catalytic sub-site of the pocket.

The two PMD molecules were bound to the NitN binding pocket in a stable orientation, while the BMD molecule at the second sub-site close to the mouth of the pocket was unstable in the co-crystal structures. The preference for PMD (with three carbon atoms) over BMD (with four carbon atoms) by NitN has been suggested to be partly due to the overall stability that is conferred by stable binding of two PMD molecules, and partly due to the speculated delayed release of the butyric acid product as a result of the BMD conformation and the extra hydrophobic contacts involving its side chain moiety.

Comparisons of the two substrate-bound co-crystal structures with the unbound C165A-Apo and the WT-Apo crystal structures have also revealed that no major conformational changes are perceivable upon binding of the substrates. Slightly larger conformational changes were however observed in the C165A-BMD structure, which might contribute towards the reduced NitN activity on BMD.

Short aliphatic amides positioning within the NitN pocket

The C165A-PMD and C165A-BMD co-crystal structures have highlighted the importance of the two active site glutamates in substrate binding and positioning, and also revealed for the first time, optimal binding, unbiased positioning and orientation of short aliphatic amides for catalysis. The PMD and BMD molecules at the catalytic sub-site have the amide carbonyl oxygen atom occupying the two-pronged oxyanion hole that is formed by the positively charged amino group of the catalytic lysine and the peptidic NH group of Tyr166. The amide amino group on the other hand interacts with the backbone O of Ala190, as well as the carboxylates of the two active site glutamates, while the side chain moieties of the substrates dock into the available space of the binding pocket where they are stabilized by hydrophobic contacts with the pocket residues.

Larger and better substrates for NitN

In addition to the confirmed ability of NitN to accommodate two short aliphatic amide molecules in the active site, electron density of unknown larger molecules has been observed in the active site pockets of the catalytic cysteine NitN mutants, further supporting our earlier

speculations that NitN would be capable of binding molecules that are larger and structurally different from short aliphatic amides. The crystal structure of the catalytic lysine NitN mutant (K131Q) has been determined, and it has been observed to have a reaction adduct of either adipamide (ADM) or related molecules covalently attached to the catalytic cysteine. Enzymatic activity assays have confirmed that NitN is able to hydrolyze ADM, and that ADM is a better substrate than the known short aliphatic amide substrates.

Structural basis for lack of activity on carbamyl and nitrile compounds by NitN

Comparisons of the C165A-PMD and C165A-BMD NitN co-crystal structures with the structures of substrate-complexed N-carbamyl-D-amino acid amidohydrolase (DCase) have revealed that the inability of NitN to hydrolyze carbamyl compounds is due to the improper geometry and limited size (space) of the active site pocket, which would result in steric clashes between various pocket residues and the carbamyl compounds.

The co-crystal structure of C165A NitN with propionitrile (PNT) has been determined, and two PNT molecules have been observed in the binding pocket. The geometry of the active site pocket and the positioning of the linear PNT molecule at the catalytic sub-site (PNT1) of NitN has revealed that the reactive cyano carbon would not be accessible by the –SH group of the catalytic cysteine. This may explain why NitN is unable to hydrolyze nitrile compounds. The nitrile nitrogen of the PNT1 molecule is positioned by interactions with the amino group of Lys131 and the backbone NH group of Tyr166 in the oxyanion hole, while the rest of the molecule is stabilized through hydrophobic contacts with various pocket residues. We have proposed that the binding of the nitrile nitrogen in the oxyanion hole constitutes a general acid catalysis mechanism that is employed by nitrile-degrading enzymes to enhance the efficiency of the nucleophilic attack. This is through polarization of the electrons at the C-N triple bond by the positively charged amino group of the catalytic lysine, which would increase the electrophilicity of the reactive cyano carbon.

6.3 Insights into the catalytic role of the two active site glutamates

The role of the two glutamates in maintaining the stability of the NitN active site and the 3D structure

Five NitN mutants of the two active site glutamates, E139Q, E61Q, E61L, E61Q-C165A and E61L-C165S, have been successfully been generated using site-directed mutagenesis, over-expressed in *E. coli* BL21 (DE3) cells and purified to over 95% purity and homogeneity using affinity chromatography and gel filtration. All the mutants were only partially soluble during expression, they were highly unstable throughout purification and crystallization procedures, and they lacked enzymatic activity. This clearly demonstrates the importance of the two glutamates in the proper folding of the enzyme, in maintaining the three-dimensional (3D) structure and in the catalytic process.

The importance of the two glutamates in the positioning of short aliphatic amides

Mass spectrometric and X-ray crystallographic characterization of the two glutamate mutants in the presence of four short aliphatic amide NitN substrates (PMD, FAE, BMD and ACE) have reiterated the critical role of the two glutamates in positioning the substrate molecules for catalysis, since the mutant enzymes either form artificial substrate adducts at the catalytic cysteine, or they are completely incapable of reacting with any carbon center in the amide substrates. The E61Q/L NitN mutants were able to carry out an S_N2 substitution reaction through the attack of the α-carbon atom (instead of the carbonyl carbon) of FAE, resulting in the elimination of fluorine and the covalent attachment of the rest of the substrate to the catalytic cysteine. The FAE S_N2 reaction adducts have been detected by mass spectrometry and visualized in the high resolution structures of the E61Q/L mutants (E61L-FAE and E61Q-FAE). The S_N2 reaction was found to be less favorable in the E61L mutant than in the E61Q NitN protein.

The co-crystal structure of E61Q-C165A double mutant in complex with ACR (E61Q/C165A-ACR) has been determined, and it has confirmed that the positioning and orientation of the amide substrates within the binding pocket is altered in the absence of Glu61. Comparison of the E161Q/C165A-ACR structure with the WT NitN structure has

revealed that in the altered position, the α-carbon of the substrates would be in alignment with the -SH group of the catalytic cysteine, explaining the occurrence of the S_N2 reactions.

The E139Q mutant was unable to react with any of the four short aliphatic amide substrates as no enzymatic activity or reaction adducts were detected. This was attributed to incorrect positioning of substrates in the pocket, which discourages nucleophilic attack at any carbon center in the amide substrates. Despite the instability of the E139Q NitN protein and the associated difficulties in obtaining diffraction quality crystals, a crystallization condition containing sodium cacodylate as a buffering agent has been identified, and one crystal structure of the E139Q mutant having a dimethylarsenate adduct at the catalytic cysteine (E13Q-Cac) has been determined. The replacement of Glu139 with Gln was found to have a destabilizing effect on the loop where the mutation is located, which emphasizes the importance of Glu139 in maintaining the structure of the active site as well as that of the entire 3D NitN structure.

The first glutamate residue (Glu61) as a general base catalyst

Due to its location relative to the catalytic cysteine, Glu61 has been suggested to be responsible for fulfilling the role of a general base catalyst during the acylation phase of the reaction, either directly or through a coordinated water molecule. However, it has been established that the E61Q/L mutants are capable of reacting with the carbonyl carbon atoms of some aliphatic amide substrates, leading to the formation of thioester acyl-enzyme intermediates that have been detected by mass spectrometry. The crystallization of the thioester-adducted E61L/Q proteins has proved to be challenging due to protein instability, hence most of these have not been visualized apart from the thioester intermediate of PMD and E61L mutant (E61L-PMD).

The ability of the mutants to react with both the α-carbon (in S_N2 reaction with FAE) and the carbonyl carbon (to form trapped thioester intermediates) of the short aliphatic amides in the absence of Glu61 implies that the first glutamate (Glu61) is not always required for the activation of the catalytic cysteine. This has raised interesting questions surrounding the exact mechanism for the reactive cysteine deprotonation. We have proposed that the activation could be substrate-assisted, and further work will need to be carried out to clarify these speculations.

The occurrence of trapped acyl-enzyme intermediates in the E61L/Q mutants suggests that Glu61 is required as a general base catalyst in the deacylation step of the reaction. This step is initiated by the attack of the carbonyl carbon of the intermediate by an activated water molecule. The involvement of Glu139 (second glutamate) in the catalytic cysteine activation is unlikely based on its location relative to the catalytic cysteine and its inaccessibility upon binding of the substrate molecules and the formation of the reaction intermediates.

The role of the two glutamates in maintaining the configuration of the active site

Comparisons of the glutamate mutant structures with the WT Nit structure, as well as calculations of the theoretical pKa values of the active site residues, have revealed that the two glutamate residues are important for maintaining the configuration of the active site environment. The amino group of the catalytic lysine has been observed to adopt a different conformation in the E61L mutant structures (E61L-FAE for instance), and the change of either of the two glutamates has been observed to greatly reduce the pKa of the catalytic lysine (Lys131) amino group. This confirms that the two glutamates are required for maintaining a high pKa for the lysine, ensuring that it remains protonated throughout the reaction.

The crystal structure of the Lys131 mutant (K131Q) having a thioester intermediate of adipamide or adipamic acid has revealed that Lys131 is required for enhancing the nucleophilic attack at the thioester intermediate by the water substrate, which confirms the suggested role of the lysine in polarizing the electrons at the double bond of the amide carbonyl group, which increases the electrophilicity at the carbonyl carbon center.

The location of the deacylation water

Quantum mechanical calculations have suggested that the nucleophilic water substrate that attacks the thioester intermediate in the deacylation phase of the reaction would be positioned between the carboxylates of the two active site glutamate residues. The geometry of the active site suggests that the ‘first’ glutamate (Glu61) plays the role of a general base catalyst, while the ‘second’ glutamate (Glu139) assists with the positioning of the deacylation water and maintaining the active site configuration in this step.

The evidence for a Cys, Glu, Glu and Lys (CEEK) catalytic tetrad

Based on the deduced or suggested catalytic functions of the two active site glutamates including their roles in maintaining the stability of both the active site and the overall 3D structure, in maintaining the configuration and the charge environment of the active site, in optimal positioning of the amide substrates for nucleophilic attack, as well as the general base catalyst role of the first glutamate, it is evident that the catalytic unit is composed of a CEEK tetrad in which the two glutamates and the lysine function as a modular assembly, rather than the previously proposed CEK triad where each residue was implied as acting as a separate entity.

6.4 Insights into the activity and reactivity of NitN with ACR

The effect of the ACR conjugated structure on NitN activity

The characterization of the NitN substrate binding pocket and the exploration of the conjugated structure of ACR have provided an explanation as to why NitN lacks hydrolytic activity on ACR. The reduced electrophilicity at the carbonyl carbon of ACR, and the large unconstrained NitN binding pocket that is incapable of adequately restraining short aliphatic amides have been identified as the two combined factors that discourage nucleophilic attack at the carbonyl carbon center by the catalytic cysteine.

The inhibitory effect of ACR on NitN

Enzymatic activity assays of the WT NitN with PMD in the presence of ACR have revealed that ACR has an inhibitory effect on the enzyme. Mass spectrometric and X-ray crystallographic characterization of the WT NitN in the presence of ACR have revealed that the enzyme unfavorably reacts with both the carbonyl and the β-alkene carbon atoms, resulting in a thioester acyl-enzyme intermediate and an ACR Michael adduct at the catalytic cysteine, respectively. These covalent adducts permanently inactivate the enzyme, explaining the inhibitory effect observed in the presence of ACR. The ability of the enzyme to react at both carbon atoms has been attributed to the instability of ACR molecules within the large and unconstrained NitN binding pocket, allowing the catalytic cysteine to attack at both carbon centers.

The crystal structure of the WT NitN protein pre-incubated with ACR has been solved, and was found to be comprised of a mixed species of a large proportion of the free protein (unbound cysteine) and a small proportion of the ACR-adducted protein, confirming the mass spectrometric findings. The thioester acyl-enzyme intermediate proportion was not observed in the crystal.

Favorable Michael addition reaction with the two active site glutamate mutants

The E139Q NitN mutant was found to favorably react with the β-alkene carbon of ACR to form an ACR Michael adduct at the catalytic cysteine that was detected by mass spectrometry. This confirmed that the E139Q binding pocket was accessible by the amide substrates. It also indicated that the substrate positioning within the mutant's binding pocket was altered as a result of the Glu139 (second glutamate) replacement. This favorably leads to stereo-electronic alignment of the β-carbon center with the thiol group of the catalytic cysteine, allowing the Michael addition reaction. E139Q mutant protein pre-incubated with ACR was highly unstable and incapable of forming crystals.

Both E61Q and E61L NitN mutants favorably react with both the carbonyl and β-alkene carbon atoms resulting in ACR thioester intermediates and ACR Michael addition at the catalytic cysteine, respectively. The ability of the E61Q/L mutants to favorably react with the β-carbon atom of ACR has reiterated the importance of Glu61 in the positioning of short aliphatic amide substrates in the NitN binding pocket for catalysis. The Michael addition reaction was more favorable with the E61Q mutant than with the E61L counterpart, which highlighted possible substrate positioning differences in the active site pocket of the two mutants. The ability of the mutants to react with the two carbon centers (α-carbon and carbonyl carbon) has been attributed to both the altered substrate positioning and the unstable binding of ACR molecules within the large and unconstrained NitN pocket, which allows the catalytic cysteine to attack at both carbon atoms. The E61Q-ACR crystal structure having an ACR Michael adduct at the catalytic cysteine has been solved, but the E61L protein pre-incubated with ACR was unstable and unable to form diffraction-quality crystals.

The possible mechanism for the catalytic cysteine activation

The crystal structures of the WT-ACR and the E61Q-ACR have revealed that Cys73, which is located in the interface between two symmetry-related molecules, was also able to carry out a Michael addition reaction with ACR resulting in an akylation adduct that was also detected by mass spectrometry. This observation has raised questions surrounding the activation mechanism of the cysteine nucleophile. We have proposed that the deprotonation of the cysteine thiol group is substrate-assisted for Michael addition reactions in and outside of the active site environment. The proposed mechanism would be facilitated by optimal alignment of the reacting groups, which would allow the proton to be directly abstracted by the α-carbon of ACR during the concerted attack of the β-carbon center by the nucleophilic cysteine. This speculation will need to be investigated further.

6.5 Future work

Further work on substrate profiling of the NitN enzyme

The kinetic parameters of NitN with ADM and the related compounds should be determined. Co-crystallization studies with the C165A NitN mutant should also be carried out in order to visualize and characterize non-covalently bound ADM molecules in the NitN binding pocket. This will inform future design of better substrates for NitN. In silico metabolite database screening and substrate docking should also be utilized for identifying better substrates for NitN, which will be important for developing NitN as a biocatalyst for industrial use.

Further work on the characterization of the amidases catalytic mechanism

Further work in the line of understanding the catalytic mechanism of amidases should focus on quantum mechanical studies of the WT NitN and the two glutamate mutants (E61Q/L and E139Q) in the presence of substrate molecules, in order to understand the activation mechanism of the reactive cysteine during the acylation phase of the reaction. Biophysical studies aimed at characterizing the rates of the formation of the reaction intermediates and adducts observed in this study in the absence of Glu61 (Glu1) should also be performed. Attempts should be made to co-crystallize the WT NitN with inert substrates or with the reactive substrates under unfavorable reaction conditions in order to visualize non-covalently bound substrate molecules in the native enzyme. This would be important for determining the location of the catalytic cysteine thiol group relative to the carboxyl group of Glu61, which

would provide insights into whether the thiol group is activated by Glu61, and whether this activation occurs through direct hydrogen bonding or via a water molecule. Experiments aimed at trapping thioester and tetrahedral intermediates in the native NitN enzyme should also be considered, so as to gain insights into the process of their formation. This will require careful manipulation of the reaction conditions in order to capture each intermediate state. Mass spectrometric and X-ray crystallographic characterization of the two glutamate mutants in the presence of ADM and other larger amide substrates should be carried out in order to assess their reactivity and positioning within the binding pockets of the mutants.

References

Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W., McCoy,A.J., Moriarty,N.W., Oeffner,R., Read,R.J., Richardson,D.C., Richardson,J.S., Terwilliger,T.C. and Zwart,P.H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol Crystallogr.* **66**, 213-221.

Aminot,A., Kirkwood,D.S. and Kθrouel,R. (1997) Determination of ammonia in seawater by the indophenol-blue method: Evaluation of the ICES NUTS I/C 5 questionnaire. *Marine Chemistry* **56**, 59-75.

Amerik,A.Y., Antonov,V.K., Gorbalenya,A.E., Kotova,S.A., Rotanova,T.V. and Shimbarevich,E.V. (1991) Site-directed mutagenesis of La protease: A catalytically active serine residue. *FEBS Letters* **287**, 211-214.

Andrade,J., Karmali,A., Carrondo,M.A. and Frazao,C. (2007) Structure of amidase from Pseudomonas aeruginosa showing a trapped acyl transfer reaction intermediate state. *J. Biol. Chem.* **282**, 19598-19605.

Asano,Y., Tachibana,M., TANI,Y. and Yamada,H. (1982) Microbial-Degradation of Nitrile Compounds .6. Purification and Characterization of Amidase Which Participates in Nitrile Degradation. *Agricultural and Biological Chemistry* **46**, 1175-1181.

Bachmair,A. and Varshavsky,A. (1989) The degradation signal in a short-lived protein. *Cell* **56**, 1019-1032.

Badie-McFarland,G., Cohen,L.W., Riggs,A.D., Morin,C., Itakura,K. and Richards,J.H. (1982) Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function. *Proc. Natl. Acad. Sci. U. S. A* **79**, 6409-6413.

Baker,R.T. and Varshavsky,A. (1995) Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. *J. Biol. Chem.* **270**, 12065-12074.

Banerjee,A., Sharma,R. and Banerjee,U.C. (2002) The nitrile-degrading enzymes: current status and future prospects. *Appl. Microbiol. Biotechnol.* **60**, 33-44.

Barglow,K.T., Saikatendu,K.S., Bracey,M.H., Huey,R., Morris,G.M., Olson,A.J., Stevens,R.C. and Cravatt,B.F. (2008) Functional proteomic and structural insights into molecular recognition in the nitrilase family enzymes. *Biochemistry* **47**, 13514-13523.

Battye,T.G., Kontogiannis,L., Johnson,O., Powell,H.R. and Leslie,A.G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol Crystallogr.* **67**, 271-281.

Belenky,P., Bogan,K.L. and Brenner,C. (2007) NAD+ metabolism in health and disease. *Trends Biochem. Sci.* **32**, 12-19.

Bellinzoni,M., Buroni,S., Pasca,M.R., Guglierame,P., Arcesi,F., De,R.E. and Riccardi,G. (2005) Glutamine amidotransferase activity of NAD+ synthetase from Mycobacterium tuberculosis depends on an amino-terminal nitrilase domain. *Res. Microbiol.* **156**, 173-177.

Bellinzoni,M., De,R.E., Branzoni,M., Milano,A., Peverali,F.A., Rizzi,M. and Riccardi,G. (2002) Heterologous expression, purification, and enzymatic activity of Mycobacterium tuberculosis NAD(+) synthetase. *Protein Expr. Purif.* **25**, 547-557.

Bergmark,E., Calleman,C.J., He,F. and Costa,L.G. (1993) Determination of hemoglobin adducts in humans occupationally exposed to acrylamide. *Toxicol. Appl. Pharmacol.* **120**, 45-54.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242.

Besaratinia,A. and Pfeifer,G.P. (2005) DNA adduction and mutagenic properties of acrylamide. *Mutat. Res.* **580**, 31-40.

Betz,J.L. and Clarke,P.H. (1972) Selective evolution of phenylacetamide-utilizing strains of Pseudomonas aeruginosa. *J. Gen. Microbiol.* **73**, 161-174.

Binkowski,T.A., Naghibzadeh,S. and Liang,J. (2003) CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res.* **31**, 3352-3355.

Bordini,E., Hamdan,M. and Righetti,P.G. (2000) Probing acrylamide alkylation sites in cysteine-free proteins by matrix-assisted laser desorption/ionisation time-of-flight. *Rapid Communications in Mass Spectrometry* **14**, 840-848.

Bordini,E., Hamdan,M. and Righetti,P.G. (1999) Probing the reactivity of S-S bridges to acrylamide in some proteins under high pH conditions by matrix-assisted laser desorption/ ionisation. *Rapid Communications in Mass Spectrometry* **13**, 1818-1827.

Bork,P. and Koonin,E.V. (1994) A new family of carbon-nitrogen hydrolases. *Protein Sci.* **3**, 1344-1346.

Bradford,M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254.

Brenner,C. (2002) Catalysis in the nitrilase superfamily. *Curr. Opin. Struct. Biol.* **12**, 775-782.

Brenner,C., Bieganowski,P., Pace,H.C. and Huebner,K. (1999) The histidine triad superfamily of nucleotide-binding proteins. *J. Cell Physiol* **181**, 179-187.

Brown,J.E., Brown,P.R. and Clarke,P.H. (1969) Butyramide-utilizing mutants of Pseudomonas aeruginosa 8602 which produce an amidase with altered substrate specificity. *J. Gen. Microbiol.* **57**, 273-285.

Brown,J.E. and Clarke,P.H. (1970) Mutations in a regulator gene allowing Pseudomonas aeruginosa 8602 to grow on butyramide. *J. Gen. Microbiol.* **64**, 329-342.

Brown,P.R. and Clarke,P.H. (1972) Amino acid substitution in an amidase produced by an acetanilide-utilizing mutant of Pseudomonas aeruginosa. *J. Gen. Microbiol.* **70**, 287-288.

Brune,D.C. (1992) Alkylation of cysteine with acrylamide for protein sequence analysis. *Anal. Biochem.* **207**, 285-290.

Brunger,A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.

Bryan,P., Pantoliano,M.W., Quill,S.G., Hsiao,H.Y. and Poulos,T. (1986) Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proc. Natl. Acad. Sci. U. S. A* **83**, 3743-3745.

Buddelmeijer,N. and Young,R. (2010) The essential Escherichia coli apolipoprotein N-acyltransferase (Lnt) exists as an extracytoplasmic thioester acyl-enzyme intermediate. *Biochemistry* **49**, 341-346.

Bury-Mone,S., Skouloubris,S., Dauga,C., Thiberge,J.M., Dailidiene,D., Berg,D.E., Labigne,A. and De Reuse,H. (2003) Presence of Active Aliphatic Amidases in Helicobacter Species Able To Colonize the Stomach. *Infection and Immunity* **71**, 5613-5622.

Caprioli,R.M., Malorni,A. and Sidona,G. (1996) Mass Spectrometry in Biomolecular Sciences, *NATO ASI Ser. C*, Vol. **475**; Kluwer: Dordrecht, The Netherlands.

Carter,P. and Wells,J.A. (1988) Dissecting the catalytic triad of a serine protease. *Nature* **332**, 564-568.

CCP4 (Collaborative Computational Project Number 4) (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**,760-763.

Chaney,A.L. and Marbach,E.P. (1962) Modified Reagents for Determination of Urea and Ammonia. *Clinical Chemistry* **8**, 130-132.

Chapman,J.R. (1996) Protein and Peptide Analysis by Mass Spectrometry. *Methods Mol. Biol.* Vol. **61**, Humana Press: Totowa, NJ.

Chebrou,H., Bigey,F., Arnaud,A. and Galzy,P. (1996) Study of the amidase signature group. *Biochim. Biophys. Acta* **1298**, 285-293.

Chen,C.Y., Chiu,W.C., Liu,J.S., Hsu,W.H. and Wang,W.C. (2003) Structural basis for catalysis and substrate specificity of Agrobacterium radiobacter N-carbamoyl-D-amino acid amidohydrolase. *J. Biol. Chem.* **278**, 26194-26201.

Cheong,T.K. and Oriel,P.J. (2000) Cloning of a wide-spectrum amidase from Bacillus stearothermophilus BR388 in Escherichia coli and marked enhancement of amidase expression using directed evolution*. *Enzyme Microb. Technol.* **26**, 152-158.

Chin,K.H., Tsai,Y.D., Chan,N.L., Huang,K.F., Wang,A.H. and Chou,S.H. (2007) The crystal structure of XC1258 from Xanthomonas campestris: a putative procaryotic Nit protein with an arsenic adduct in the active site. *Proteins: Structure, Function, and Bioinformatics* **69**, 665-671.

Clarke,P.H. (1970) The aliphatic amidase from Pseudomonas aeruginosa. *Advances in Microbial Physiol.* **4**, 179-222.

Cobzaru,C., Ganas,P., Mihasan,M., Schleberger,P. and Brandsch,R. (2011) Homologous gene clusters of nicotine catabolism, including a new omega-amidase for alpha-ketoglutaramate, in species of three genera of Gram-positive bacteria. *Research in Microbiology* **162**, 285-291.

Cohen,G. (1997) ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *Journal of Applied Crystallography* **30**, 1160-1161.

Coles,B. (1984) Effects of modifying structure on electrophilic reactions with biological nucleophiles. *Drug Metab Rev.* **15**, 1307-1334.

Cooper,A.J. (2004) The role of glutamine transaminase K (GTK) in sulfur and alpha-keto acid metabolism in the brain, and in the possible bioactivation of neurotoxicants. *Neurochem. Int.* **44**, 557-577.

Cooper,A.J.L. and Meister,A. (1981) Comparative studies of glutamine transaminases from rat tissues. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **69**, 137-145.

Craik,C.S., Roczniak,S., Largman,C. and Rutter,W.J. (1987) The catalytic role of the active site aspartic acid in serine proteases. *Science* **237**, 909-913.

Davis,R.W., Botstein,D. and Roth,J.R. (1980) A manual for genetic engineering: Advanced Bacterial Genetics. *Cold Spring Harbour Laboratory*, New York (QH434 D38).

Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B., Snoeyink,J., Richardson,J.S. and Richardson,D.C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* **35**, W375-W383.

Dearfield,K.L., Abernathy,C.O., Ottley,M.S., Brantner,J.H. and Hayes,P.F. (1988) Acrylamide: its metabolism, developmental and reproductive effects, genotoxicity, and carcinogenicity. *Mutat. Res.* **195**, 45-77.

DeLano, W.L. (2004) The PyMOL Molecular Graphics System. *DeLano Scientific*, San Carlos CA, USA. http://www.pymol.org.

Drahl,C. (2008) Surprise from S_N2 snapshots; ion velocity measurements unveil additional unforeseen mechanism. *Chemical and Engineering news* **86 (2)**, 9.

Dupre,S. and Cavallini,D. (1979) Purification and properties of pantetheinase from horse kidney. *Methods Enzymol.* **62**, 262-267.

Dupre,S., Graziani,M.T., Rosei,M.A., Fabi,A. and Del,G.E. (1970) The enzymatic breakdown of pantethine to pantothenic acid and cystamine. *Eur. J. Biochem.* **16**, 571-578.

Eisele,F. and Wolf,D.H. (2008) Degradation of misfolded protein in the cytoplasm is mediated by the ubiquitin ligase Ubr1. *FEBS Lett.* **582**, 4143-4146.

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

Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 72-82.

Fodor,K., Harmat,V., Neutze,R., Szilagyi,L., Graf,L. and Katona,G. (2006) Enzyme:substrate hydrogen bond shortening during the acylation phase of serine protease catalysis. *Biochemistry* **45**, 2114-2121.

Fournand,D. and Arnaud,A. (2001) Aliphatic and enantioselective amidases: from hydrolysis to acyl transfer activity. *J. Appl. Microbiol.* **91**, 381-393.

Fournand,D., Arnaud,A. and Galzy,P. (1998a) Study of the acyl transfer activity of a recombinant amidase overproduced in an Escherichia coli strain. Application for short-chain hydroxamic acid and acid hydrazide synthesis. *Journal of Molecular Catalysis B-Enzymatic* **4**, 77-90.

Fournand,D., Bigey,F. and Arnaud,A. (1998b) Acyl transfer activity of an amidase from Rhodococcus sp. strain R312: formation of a wide range of hydroxamic acids. *Appl. Environ. Microbiol.* **64**, 2844-2852.

Friedman,M. (2003) Chemistry, biochemistry, and safety of acrylamide. A review. *J. Agric. Food Chem.* **51**, 4504-4526.

Friedman,M. (2005) Biological effects of Maillard browning products that may affect acrylamide safety in food: biological effects of Maillard products. *Adv. Exp. Med. Biol* **561**, 135-156.

Friedman,M., Cavins,J.F. and Wall,J.S. (1965) Relative Nucleophilic Reactivities of Amino Groups and Mercaptide Ions in Addition Reactions with Alpha Beta-Unsaturated Compounds. *Journal of the American Chemical Society* **87**, 3672-&.

Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Scalmani,G., Barone,V., Mennucci,B., Petersson,G.A., Nakatsuji,H., Caricato,M., Li,X., Hratchian,H.P., Izmaylov,A.F., Bloino,J., Zheng,G., Sonnenberg,J.L., Hada,M., Ehara,M., Toyota,K., Fukuda,R., Hasegawa,J., Ishida,M., Nakajima,T., Honda,Y., Kitao,O., Nakai,H., Vreven,T., Montgomery,Jr.,J.A., Peralta,J.E., Ogliaro,F., Bearpark,M., Heyd,J.J., Brothers,E., Kudin,K.N., Staroverov,V.N., Kobayashi,R., Normand,J., Raghavachari,K., Rendell,A., Burant,J.C., Iyengar,S.S., Tomasi,J., Cossi,M., Rega,N., Millam,N.J., Klene,M., Knox,J.E., Cross,J.B., Bakken,V., Adamo,C., Jaramillo,J., Gomperts,R., Stratmann,R.E., Yazyev,O., Austin,A.J., Cammi,R., Pomelli,C., Ochterski,J.W., Martin,R.L., Morokuma,K., Zakrzewski,V.G., Voth,G.A., Salvador,P., Dannenberg,J.J., Dapprich,S., Daniels,A.D., Farkas,Ö., Foresman,J.B., Ortiz,J.V., Cioslowski,J. and Fox,D.J. (2009) Gaussian 09, Revision A.1, *Gaussian, Inc.*, Wallingford C.

Freyssinet,G., Peleissier,B., Freyssinet,M. and Delon,R. (1996) Crops resistant to oxynils: from the laboratory to the market. *Field Crops Res.* **45**:125-133.

Giron,P., Dayon,L. and Sanchez,J.C. (2010) Cysteine tagging for MS-based proteomics. *Mass Spectrometry Reviews*.

Gonda,D.K., Bachmair,A., Wunning,I., Tobias,J.W., Lane,W.S. and Varshavsky,A. (1989) Universality and structure of the N-end rule. *J. Biol. Chem.* **264**, 16700-16712.

Gordon,J.C., Myers,J.B., Folta,T., Shoja,V., Heath,L.S. and Onufriev,A. (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33**, W368-W371.

Gregoriou,M. and Brown,P.R. (1979) Inhibition of the aliphatic amidase from Pseudomonas aeruginosa by urea and related compounds. *Eur. J. Biochem.* **96**, 101-108.

Grifantini,R., Pratesi,C., Galli,G. and Grandi,G. (1996) Topological mapping of the cysteine residues of N-carbamyl-D-amino-acid amidohydrolase and their role in enzymatic activity. *J. Biol. Chem.* **271**, 9326-9331.

Grossowicz,N., Wainfan,E., Borek,E. and Waelsch,H. (1950) The enzymatic formation of hydroxamic acids from glutamine and asparagine. *J. Biol. Chem.* **187**, 111-125.

Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723.

Hamdan,M., Bordini,E., Galvani,M. and Righetti,P.G. (2001) Protein alkylation by acrylamide, its N-substituted derivatives and cross-linkers and its relevance to proteomics: a matrix assisted laser desorption/ionization-time of flight-mass spectrometry study. *Electrophoresis* **22**, 1633-1644.

Hanahan,D., Jessee,J. and Bloom,F.R. (1991) Plasmid transformation of Escherichia coli and other bacteria. *Methods Enzymol.* **204**, 63-113.

Hahn,C.S. and Strauss,J.H. (1990) Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease. *J Virol.* **64**, 3069-3073.

Hara,N., Yamada,K., Terashima,M., Osago,H., Shimoyama,M. and Tsuchiya,M. (2003) Molecular identification of human glutamine- and ammonia-dependent NAD^+ synthetase. *J. Biol. Chem.* **278 (13)**, 10914-10921.

Harper,D.B. (1977a) Fungal degradation of aromatic nitriles. Enzymology of C-N cleavage by Fusarium solani. *Biochem. J.* **167**, 685-692.

Harper,D.B. (1977b) Microbial metabolism of aromatic nitriles. Enzymology of C-N cleavage by Nocardia sp. (Rhodochrous group) N.C.I.B. 11216. *Biochem. J.* **165**, 309-319.

Hashimoto,H., Aoki,M., Shimizu,T., Nakai,T., Morikawa,H., Ikenaka,Y., Takahashi,S. and Sato,M. (2003) The crystal structures of the C171A/V236A mutant of N-carbamyl-D-amino acid amidohydrolase complexed with N-carbamyl-D-amino acid substrates. *Protein Data Bank*, PDB ids 1UF5, 1UF7 and 1UF8.

Herbert,B., Galvani,M., Hamdan,M., Olivieri,E., MacCarthy,J., Pedersen,S. and Righetti,P.G. (2001) Reduction and alkylation of proteins in preparation of two-dimensional map analysis: why, when, and how? *Electrophoresis* **22**, 2046-2057.

Hinson,J.A. and Roberts,D.W. (1992) Role of covalent and noncovalent interactions in cell toxicity: effects on proteins. *Annu. Rev. Pharmacol. Toxicol.* **32**, 471-510.

Huang,W., Jia,J., Cummings,J., Nelson,M., Schneider,G. and Lindqvist,Y. (1997) Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold. *Structure.* **5**, 691-699.

Huang,X., Holden,H.M. and Raushel,F.M. (2001) Channeling of substrates and intermediates in enzyme-catalyzed reactions. *Annu. Rev. Biochem.* **70**, 149-180.

Hung,C.L., Liu,J.H., Chiu,W.C., Huang,S.W., Hwang,J.K. and Wang,W.C. (2007) Crystal structure of Helicobacter pylori formamidase AmiF reveals a cysteine-glutamate-lysine catalytic triad. *J. Biol. Chem.* **282**, 12220-12229.

Hymes,J., Fleischhauer,K. and Wolf,B. (1995) Biotinylation of histones by human serum biotinidase: assessment of biotinyl-transferase activity in sera from normal individuals and children with biotinidase deficiency. *Biochem. Mol. Med.* **56**, 76-83.

Hymes,J. and Wolf,B. (1996) Biotinidase and its roles in biotin metabolism. *Clin. Chim. Acta* **255**, 1-11.

Ingvorsen,K., Hojer-Pedersen,B. and Godtfredsen,S.E. (1991) Novel cyanide-hydrolyzing enzyme from Alcaligenes xylosoxidans subsp. denitrificans. *Appl. Environ. Microbiol.* **57**, 1783-1789.

Jacobs,A. (1997) Understanding organic reaction mechanisms. *Cambridge University Press* Vol. **1**, 179-181.

Kabsch,W. (2010) XDS. *Acta Crystallographica Section D* **66**, 125-132.

Kaminskas,L.M., Pyke,S.M. and Burcham,P.C. (2005) Differences in lysine adduction by acrolein and methyl vinyl ketone: implications for cytotoxicity in cultured hepatocytes. *Chem. Res. Toxicol.* **18**, 1627-1633.

Karmali,A., Pacheco,R., Tata,R. and Brown,P. (2001) Substitutions of Thr-103-Ile and Trp-138-Gly in amidase from Pseudomonas aeruginosa are responsible for altered kinetic properties and enzyme instability. *Mol. Biotechnol.* **17**, 201-212.

Karmali,A., Tata,R. and Brown,P.R. (2000) Substitution of Glu-59 by Val in amidase from Pseudomonas aeruginosa results in a catalytically inactive enzyme. *Mol. Biotechnol.* **16**, 5-16.

Kato,Y., Ooi,R. and Asano,Y. (1998) Isolation and characterization of a bacterium possessing a novel aldoxime-dehydration activity and nitrile-degrading enzymes. *Arch. Microbiol.* **170**, 85-90.

Kato,Y., Ooi,R. and Asano,Y. (2000) Distribution of aldoxime dehydratase in microorganisms. *Appl. Environ. Microbiol.* **66**, 2290-2296.

Katona,G., Wilmouth,R.C., Wright,P.A., Berglund,G.I., Hajdu,J., Neutze,R. and Schofield,C.J. (2002) X-ray structure of a serine protease acyl-enzyme complex at 0.95-A resolution. *The Journal of biological chemistry* **277**, 21962-21970.

Kemp DS and Vellaccio F (1980) Conjugated alkenes. *In Organic chemistry, Worth Publishers, New York.* **Chapter 19**, pp. 650-684.

Kimani,S.W., Agarkar,V.B., Cowan,D.A., Sayed,M.F. and Sewell,B.T. (2007) Structure of an aliphatic amidase from Geobacillus pallidus RAPc8. *Acta Crystallogr. D. Biol. Crystallogr.* **63**, 1048-1058.

Kobayashi,M., Goda,M. and Shimizu,S. (1998) Nitrilase catalyzes amide hydrolysis as well as nitrile hydrolysis. *Biochem. Biophys. Res. Commun.* **253**, 662-666.

Kobayashi,M., Izui,H., Nagasawa,T. and Yamada,H. (1993) Nitrilase in Biosynthesis of the Plant Hormone Indole-3-Acetic Acid from Indole-3-Acetonitrile: Cloning of the Alcaligenes Gene and Site-Directed Mutagenesis of Cysteine Residues. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 247-251.

Kobayashi,M., Nagasawa,T. and Yamada,H. (1989) Nitrilase of Rhodococcus rhodochrous J1. *European Journal of Biochemistry* **182**, 349-356.

Kobayashi,M. and Shimizu,S. (1998) Metalloenzyme nitrile hydratase: Structure, regulation, and application to biotechnology. *Nat Biotech* **16**, 733-736.

Kobayashi,M., Komeda,H., Yanaka,N., Nagasawa,T. and Yamada,H. (1992a) Nitrilase from Rhodococcus rhodochrous J1. Sequencing and overexpression of the gene and identification of an essential cysteine residue. *J. Biol. Chem.* **267**, 20746-20751.

Kobayashi,M., Yanaka,N., Nagasawa,T. and Yamada,H. (1992b) Primary structure of an aliphatic nitrile-degrading enzyme, aliphatic nitrilase, from Rhodococcus rhodochrous K22 and expression of its gene and identification of its active site residue. *Biochemistry* **31**, 9000-9007.

Krasnikov,B.F., Chien,C.H., Nostramo,R., Pinto,J.T., Nieves,E., Callaway,M., Sun,J., Huebner,K. and Cooper,A.J. (2009a) Identification of the putative tumor suppressor Nit2 as omega-amidase, an enzyme metabolically linked to glutamine and asparagine transamination. *Biochimie* **91**, 1072-1080.

Krasnikov,B.F., Nostramo,R., Pinto,J.T. and Cooper,A.J. (2009b) Assay and purification of omega-amidase/Nit2, a ubiquitously expressed putative tumor suppressor, that catalyzes the deamidation of the alpha-keto acid analogues of glutamine and asparagine. *Anal. Biochem.* **391**, 144-150.

Kumaran,D., Eswaramoorthy,S., Gerchman,S.E., Kycia,H., Studier,F.W. and Swaminathan,S. (2003) Crystal structure of a putative CN hydrolase from yeast. *Proteins: Structure, Function, and Bioinformatics* **52**, 283-291.

Kurtz,A.J. and Lloyd,R.S. (2003) 1,N 2-Deoxyguanosine Adducts of Acrolein, Crotonaldehyde, and trans-4-Hydroxynonenal Cross-link to Peptides via Schiff Base Linkage. *Journal of Biological Chemistry* **278**, 5970-5976.

Kvalnes-Krick,K.L. and Traut,T.W. (1993) Cloning, sequencing, and expression of a cDNA encoding beta-alanine synthase from rat liver. *J. Biol. Chem.* **268**, 5686-5693.

Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

LaRonde-LeBlanc,N., Resto,M. and Gerratana,B. (2009) Regulation of active site coupling in glutamine-dependent NAD(+) synthetase. *Nat. Struct. Mol. Biol.* **16**, 421-429.

Levy,F., Johnsson,N., Rumenapf,T. and Varshavsky,A. (1996) Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl. Acad. Sci. U. S. A* **93**, 4907-4912.

Lin,C.H., Chung,M.Y., Chen,W.B. and Chien,C.H. (2007) Growth inhibitory effect of the human NIT2 gene and its allelic imbalance in cancers. *FEBS J* **274**, 2946-2956.

Liu,J.K., Liu,C.H. and Lin,C.S. (1997) The role of nitrogenase in a cyanide-degrading Klebsiella oxytoca strain. *Proc. Natl. Sci. Counc. Repub. China B* **21**, 37-42.

LoPachin,R.M. (2004) The changing view of acrylamide neurotoxicity. *Neurotoxicology* **25**, 617-630.

LoPachin,R.M., Balaban,C.D. and Ross,J.F. (2003) Acrylamide axonopathy revisited. *Toxicol. Appl. Pharmacol.* **188**, 135-153.

LoPachin,R.M. and Barber,D.S. (2006) Synaptic cysteine sulfhydryl groups as targets of electrophilic neurotoxicants. *Toxicol. Sci.* **94**, 240-255.

LoPachin,R.M., Barber,D.S. and Gavin,T. (2008) Molecular mechanisms of the conjugated alpha,beta-unsaturated carbonyl derivatives: relevance to neurotoxicity and neurodegenerative diseases. *Toxicol. Sci.* **104**, 235-249.

LoPachin,R.M., Barber,D.S., Geohagen,B.C., Gavin,T., He,D. and Das,S. (2007) Structure-toxicity analysis of type-2 alkenes: in vitro neurotoxicity. *Toxicol. Sci.* **95**, 136-146.

LoPachin,R.M. and Decaprio,A.P. (2005) Protein adduct formation as a molecular mechanism in neurotoxicity. *Toxicol. Sci.* **86**, 214-225.

LoPachin,R.M., Ross,J.F., Reid,M.L., Das,S., Mansukhani,S. and Lehning,E.J. (2002) Neurological evaluation of toxic axonopathies in rats: acrylamide and 2,5-hexanedione. *Neurotoxicology* **23**, 95-110.

Lovell,S.C., Davis,I.W., Arendall,W.B., III, de Bakker,P.I., Word,J.M., Prisant,M.G., Richardson,J.S. and Richardson,D.C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins: Structure, Function, and Bioinformatics* **50**, 437-450.

Lucher,L.A., Chen,Y.M. and Walker,J.B. (1989) Reactions catalyzed by purified L-glutamine: keto-scyllo-inositol aminotransferase, an enzyme required for biosynthesis of aminocyclitol antibiotics. *Antimicrob. Agents Chemother.* **33**, 452-459.

Lundgren,S., Lohkamp,B., Andersen,B., Piskur,J. and Dobritzsch,D. (2008) The crystal structure of beta-alanine synthase from Drosophila melanogaster reveals a homooctameric helical turn-like assembly. *J. Mol. Biol.* **377**, 1544-1559.

Maestracci,M., Thiery,A., Arnaud,A. and Galzy,P. (1984) Activity and regulation of an amide (acylamide amidohydrolase, EC 3.5.1.4) with a wide spectrum from Brevibacterium sp. *Archives of Microbiol.* **138**, 315-320.

Maestracci,M., Thiery,A., Arnaud,A. and Galzy,P. (1986) A Study of the Mechanism of the Reactions Catalyzed by the Amidase Brevibacterium Sp R312. *Agricultural and Biological Chemistry* **50**, 2237-2241.

Mahadevan,S. and Thimann,K.V. (1964) Nitrilase II. Substrate specificity and possible mode of action. *Arch. Biochem. Biophys.* **107**, 62-68.

Makhongela,H.S., Glowacka,A.E., Agarkar,V.B., Sewell,B.T., Weber,B., Cameron,R.A., Cowan,D.A. and Burton,S.G. (2007) A novel thermostable nitrilase superfamily amidase from Geobacillus pallidus showing acyl transfer activity. *Appl. Microbiol. Biotechnol.* **75**, 801-811.

McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J Appl. Crystallogr.* **40**, 658-674.

Meister,A., Levintow,L., Greenfield,R.E. and Abendschein,P.A. (1955) Hydrolysis and transfer reactions catalyzed by omega-amidase preparations. *The Journal of biological chemistry* **215**, 441-460.

Meister,A., Sober,H.A., Tice,S.V. and Fraser,P.E. (1952) Transamination and associated deamidation of asparagine and glutamine. *The Journal of biological chemistry* **197**, 319-330.

Meng,E.C., Pettersen,E.F., Couch,G.S., Huang,C.C. and Ferrin,T.E. (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC. Bioinformatics.* **7**, 339.

Mikosch,J., Trippel,S., Eichhorn,C., Otto,R., Lourderaj,U., Zhang,J.X., Hase,W.L., Weidem++ller,M. and Wester,R. (2008) Imaging Nucleophilic Substitution Dynamics. *Science* **319**, 183-186.

Mineki,R., Taka,H., Fujimura,T., Kikkawa,M., Shindo,N. and Murayama,K. (2002) In situ alkylation with acrylamide for identification of cysteinyl residues in proteins during one- and two-dimensional sodium dodecyl sulphate-polyacrylamide gel electrophoresis. *Proteomics.* **2**, 1672-1681.

Mogk,A., Schmidt,R. and Bukau,B. (2007) The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends Cell Biol.* **17**, 165-172.

Moreau,J.L., Bernet,N., Arnaud,A. and Galzy,P. (1993) isolation of Brevibacterium sp r312 mutants potentially useful for the enzymatic production of adipic acid. *Canadian Journal of Microbiology* **39**, 524-528.

Mottram,D.S., Wedzicha,B.L. and Dodson,A.T. (2002) Acrylamide is formed in the Maillard reaction. *Nature* **419**, 448-449.

Muller,D. and Gabriel,J. (1999) Bacterial degradation of the herbicide bromoxynil by Agrobacterium radiobacter in biofilm. *Folia Microbiol. (Praha)* **44**, 377-379.

Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 240-255.

Nagano,N., Noguchi,T. and Akiyama,Y. (2007) Systematic comparison of catalytic mechanisms of hydrolysis and transfer reactions classified in the EzCatDB database. *Proteins: Structure, Function, and Bioinformatics* **66**, 147-159.

Nakada,Y., Jiang,Y., Nishijyo,T., Itoh,Y. and Lu,C.D. (2001) Molecular characterization and regulation of the aguBA operon, responsible for agmatine utilization in Pseudomonas aeruginosa PAO1. *J. Bacteriol.* **183**, 6517-6524.

Nakada,Y. and Itoh,Y. (2003) Identification of the putrescine biosynthetic genes in Pseudomonas aeruginosa and characterization of agmatine deiminase and N-carbamoylputrescine amidohydrolase of the arginine decarboxylase pathway. *Microbiology* **149**, 707-714.

Nakai,T., Hasegawa,T., Yamashita,E., Yamamoto,M., Kumasaka,T., Ueki,T., Nanba,H., Ikenaka,Y., Takahashi,S., Sato,M. and Tsukihara,T. (2000) Crystal structure of N-carbamyl-D-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases. *Structure.* **8**, 729-737.

Nakamura,A., Yao,M., Chimnaronk,S., Sakai,N. and Tanaka,I. (2006) Ammonia channel couples glutaminase with transamidase reactions in GatCAB. *Science* **312**, 1954-1958.

Nel,A.J., Tuffin,I.M., Sewell,B.T. and Cowan,D.A. (2011) Unique aliphatic amidase from a psychrotrophic and haloalkaliphilic nesterenkonia isolate. *Appl. Environ. Microbiol.* **77**, 3696-3702.

Nishihara,K., Kanemori,M., Kitagawa,M., Yanagi,H. and Yura,T. (1998) Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli. *Appl. Environ. Microbiol.* **64**, 1694-1699.

Nishihara,K., Kanemori,M., Yanagi,H. and Yura,T. (2000) Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Appl. Environ. Microbiol.* **66**, 884-889.

Novo,C., Farnaud,S., Tata,R., Clemente,A. and Brown,P.R. (2002) Support for a three-dimensional structure predicting a Cys-Glu-Lys catalytic triad for Pseudomonas aeruginosa amidase comes from site-directed mutagenesis and mutations altering substrate specificity. *Biochem. J.* **365**, 731-738.

Novo,C., Tata,R., Clemente,A. and Brown,P.R. (1995) Pseudomonas aeruginosa aliphatic amidase is related to the nitrilase/cyanide hydratase enzyme family and Cys166 is predicted to be the active site nucleophile of the catalytic mechanism. *FEBS Lett.* **367**, 275-279.

O'Reilly,C. and Turner,P.D. (2003) The nitrilase family of CN hydrolysing enzymes - a comparative study. *J. Appl. Microbiol.* **95**, 1161-1174.

Ogawa,J., Chung,M.C., Hida,S., Yamada,H. and Shimizu,S. (1994) Thermostable N-carbamoyl-D-amino acid amidohydrolase: screening, purification and characterization. *J. Biotechnol.* **38**, 11-19.

Ogawa,J., Shimizu,S. and Yamada,H. (1993) N-carbamoyl-D-amino acid amidohydrolase from Comamonas sp. E222c purification and characterization. *Eur. J. Biochem.* **212**, 685-691.

Olivieri,R., Fascetti,E., Angelini,L. and Degen,L. (1981) Microbial transformation of racemic hydantoins to D-amino acids. *Biotechnology and Bioengineering* **23**, 2173-2183.

Otwinowski, Z. and Minor, W. (1997) [20] Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology Macromolecular Crystallography Part A* ed. Charles,W.C. pp. 307-326. Academic Press.

Pace,H.C. and Brenner,C. (2001) The nitrilase superfamily: classification, structure and function. *Genome Biol.* **2**, REVIEWS0001.

Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J., Bieganowski,P., Pekarsky,Y., Croce,C.M. and Brenner,C. (2000) Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers. *Curr. Biol.* **10**, 907-917.

Palani,K., Burley,S.K. and Swaminathan,S. (2009) The crystal structure of a glutamine-dependent NAD^+ synthetase from Cytophaga hutchinsonii. *Protein Data Bank*, PDB id 3ILV.

Patricelli,M.P. and Cravatt,B.F. (2000) Clarifying the catalytic roles of conserved residues in the amidase signature family. *J. Biol. Chem.* **275**, 19177-19184.

Pearson,R.G. and Songstad,J. (1967) Application of Principle of Hard and Soft Acids and Bases to Organic Chemistry. *Journal of the American Chemical Society* **89**, 1827-&.

Perez,H.L., Cheong,H.K., Yang,J.S. and Osterman-Golkar,S. (1999) Simultaneous analysis of hemoglobin adducts of acrylamide and glycidamide by gas chromatography-mass spectrometry. *Anal. Biochem.* **274**, 59-68.

Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612.

Pflugrath,J.W. (1999) The finer things in X-ray diffraction data collection. *Acta Crystallogr. D Biol Crystallogr.* **55**, 1718-1725.

Piotrowski,M., Janowitz,T. and Kneifel,H. (2003) Plant C-N Hydrolases and the Identification of a Plant N-Carbamoylputrescine Amidohydrolase Involved in Polyamine Biosynthesis. *Journal of Biological Chemistry* **278**, 1708-1712.

Piotrowski,M., Schonfelder,S. and Weiler,E.W. (2001) The Arabidopsis thaliana isogene NIT4 and its orthologs in tobacco encode beta-cyano-L-alanine hydratase/nitrilase. *J. Biol. Chem.* **276**, 2616-2621.

Piskur,J., Schnackerz,K.D., Andersen,G. and Bjornberg,O. (2007) Comparative genomics reveals novel biochemical pathways. *Trends Genet.* **23**, 369-372.

Pollak,P., Romender,G., Hagedorn,F. and Gelbke,H-P. (1991) In: Elvers B, Hawkins S, Schulz G (eds) Ullman's encyclopedia of industrial chemistry, 5th edn. *Wiley-VCH, Weinheim*. **vol A17**, pp 363-376.

Raczynska,J.E., Vorgias,C.E., Antranikian,G. and Rypniewski,W. (2010) Crystallographic analysis of a thermoactive nitrilase. *J. Struct. Biol.*

Ramachandran,G.N., Ramakrishnan,C. and Sasisekharan,V. (1963) Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7**, 95-99.

Raushel,F.M., Thoden,J.B. and Holden,H.M. (2003) Enzymes with molecular tunnels. *Acc. Chem. Res.* **36**, 539-548.

Read,R.J. (1999) Detecting outliers in non-redundant diffraction data. *Acta Crystallogr. D. Biol. Crystallogr.* **55**, 1759-1764.

Rice,J.M. (2005) The carcinogenicity of acrylamide. *Mutat. Res.* **580**, 3-20.

Sakai,N., Tajika,Y., Yao,M., Watanabe,N. and Tanaka,I. (2004) Crystal structure of hypothetical protein PH0642 from Pyrococcus horikoshii at 1.6A resolution. *Proteins: Structure, Function, and Bioinformatics* **57**, 869-873.

Sassetti,C.M., Boyd,D.H. and Rubin,E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48**, 77-84.

Sawyer,D.T., Sugimoto,H. and Calderwood,T.S. (1984) Base (O.-2, e-, or OH-)-induced autoxygenation of organic substrates: a model chemical system for cytochrome P-450-catalyzed monoxygenation and dehydrogenation by dioxygen. *Proc. Natl. Acad. Sci. U. S. A* **81**, 8025-8027.

Semba,S., Han,S.Y., Qin,H.R., McCorkell,K.A., Iliopoulos,D., Pekarsky,Y., Druck,T., Trapasso,F., Croce,C.M. and Huebner,K. (2006) Biological functions of mammalian Nit1, the counterpart of the invertebrate NitFhit Rosetta stone protein, a possible tumor suppressor. *The Journal of biological chemistry* **281**, 28244-28253.

Sewell,B.T., Thuku,R.N., Zhang,X. and Benedik,M.J. (2005) Oligomeric structure of nitrilases: effect of mutating interfacial residues on activity. *Ann. N. Y. Acad. Sci.* **1056**, 153-159.

Sharma,M., Sharma,N. and Bhalla,T. (2009) Amidases: versatile enzymes in nature. *Reviews in Environmental Science and Biotechnology* **8**, 343-366.

Sienczyk,M. and Oleksyszyn,J. (2009) Irreversible inhibition of serine proteases - design and in vivo activity of diaryl alpha-aminophosphonate derivatives. *Curr. Med. Chem* **16**, 1673-1687.

Skouloubris,S., Labigne,A. and De Reuse,H. (2001) The AmiE aliphatic amidase and AmiF formamidase of Helicobacter pylori: natural evolution of two enzyme paralogues. *Mol. Microbiol.* **40**, 596-609.

Skouloubris,S., Labigne,A. and De Reuse,H. (1997) Identification and characterization of an aliphatic amidase in Helicobacter pylori. *Mol. Microbiol.* **25**, 989-998.

Smith,E.A. and Oehme,F.W. (1991) Acrylamide and polyacrylamide: a review of production, use, environmental fate and neurotoxicity. *Rev. Environ. Health* **9**, 215-228.

Soriano-Maldonado,P., Martinez-Gomez,A.I., ndujar-Sanchez,M., Neira,J.L., Clemente-Jimenez,J.M., Las Heras-Vazquez,F.J., Rodriguez-Vico,F. and Martinez-Rodriguez,S. (2011) Biochemical and mutational studies of formamidase from Bacillus cereus CECT 5050T support a C-E-E-K tetrad in several members of the nitrilase superfamily. *Appl. Environ. Microbiol.*

Soubrier,F., Levy-Schil,S., Mayaux,J.F., Petre,D., Arnaud,A. and Crouzet,J. (1992) Cloning and primary structure of the wide-spectrum amidase from Brevibacterium sp. R312: high homology to the amiE product from Pseudomonas aeruginosa. *Gene* **116**, 99-104.

Stevenson,D.E., Feng,R. and Storer,A.C. (1990) Detection of covalent enzyme-substrate complexes of nitrilase by ion-spray mass spectroscopy. *FEBS Lett.* **277**, 112-114.

Ta,T.C., Joy,K.W. and Ireland,R.J. (1985) Role of asparagine in the photorespiratory nitrogen metabolism of pea leaves. *Plant Physiol* **78**, 334-337.

Tata,R., Marsh,P. and Brown,P.R. (1994) Arg-188 and Trp-144 are implicated in the binding of urea and acetamide to the active site of the amidase from Pseudomonas aeruginosa. *Biochim. Biophys. Acta* **1205**, 139-145.

Tempel,W., Rabeh,W.M., Bogan,K.L., Belenky,P., Wojcik,M., Seidle,H.F., Nedyalkova,L., Yang,T., Sauve,A.A., Park,H.W. and Brenner,C. (2007) Nicotinamide riboside kinase structures reveal new pathways to NAD+. *PLoS. Biol.* **5**, e263.

Terwilliger,T.C. (2004) Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2144-2149.

Thiery,A., Maestracci,M., Arnaud,A. and Galzy,P. (1986) acyltransferase activity of the wide spectrum amidase of Brevibacterium sp r312. *Journal of General Microbiology* **132**, 2205-2208.

Thomas,J.G., Ayling,A. and Baneyx,F. (1997) Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from E. coli. To fold or to refold. *Appl. Biochem. Biotechnol.* **66**, 197-238.

Thuku,R.N., Brady,D., Benedik,M.J. and Sewell,B.T. (2009) Microbial nitrilases: versatile, spiral forming, industrial enzymes. *Journal of Applied Microbiology* **106**, 703-727.

Traut,T.W. and Jones,M.E. (1996) Uracil metabolism - UMP synthesis from orotic acid or uridine and conversion of uracil to beta-alanine: Enzymes and cDNAs. *Progress in Nucleic Acid Research and Molecular Biology, Vol 53* **53**, 1-78.

Tschumi,A., Nai,C., Auchli,Y., Hunziker,P., Gehrig,P., Keller,P., Grau,T. and Sander,P. (2009) Identification of apolipoprotein N-acyltransferase (Lnt) in mycobacteria. *J. Biol. Chem.* **284**, 27146-27156.

Tucek,M., Tenglerova,J., Kollarova,B., Kvasnickova,M., Maxa,K., Mohyluk,I., Svandova,E., Topolcan,O., Vlasak,Z. and Cikrt,M. (2002) Effect of acrylate chemistry on human health. *Int. Arch. Occup. Environ. Health* **75 Suppl**, S67-S72.

Uchida,K. (1999) Current status of acrolein as a lipid peroxidation product. *Trends Cardiovasc. Med.* **9**, 109-113.

Vagin,A. and Teplyakov,A. (1997) MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography* **30**, 1022-1025.

Varshavsky,A. (1997) The N-end rule pathway of protein degradation. *Genes Cells* **2**, 13-28.

Vidal-Ingigliardi,D., Lewenza,S. and Buddelmeijer,N. (2007) Identification of essential residues in apolipoprotein N-acyl transferase, a member of the CN hydrolase family. *J. Bacteriol.* **189**, 4456-4464.

Wang,H., Piatkov,K.I., Brower,C.S. and Varshavsky,A. (2009) Glutamine-specific N-terminal amidase, a component of the N-end rule pathway. *Mol. Cell* **34**, 686-695.

Wang,P., Matthews,D.E. and VanEtten,H.D. (1992) Purification and characterization of cyanide hydratase from the phytopathogenic fungus Gloeocercospora sorghi. *Archives of Biochemistry and Biophysics* **298**, 569-575.

Wang,W.C., Hsu,W.H., Chien,F.T. and Chen,C.Y. (2001) Crystal structure and site-directed mutagenesis studies of N-carbamoyl-D-amino-acid amidohydrolase from Agrobacterium radiobacter reveals a homotetramer and insight into a catalytic cleft. *J. Mol. Biol.* **306**, 251-261.

Wasternack,C. (1980) Degradation of pyrimidines and pyrimidine analogs--pathways and mutual influences. *Pharmacol. Ther.* **8**, 629-651.

Whiting,A.K. and Peticolas,W.L. (1994) Details of the acyl-enzyme intermediate and the oxyanion hole in serine protease catalysis. *Biochemistry* **33**, 552-561.

Wilmouth,R.C., Clifton,I.J., Robinson,C.V., Roach,P.L., Aplin,R.T., Westwood,N.J., Hajdu,J. and Schofield,C.J. (1997) Structure of a specific acyl-enzyme complex formed between beta-casomorphin-7 and porcine pancreatic elastase. *Nat Struct Biol* **4**, 456-462.

Wilmouth,R.C., Edman,K., Neutze,R., Wright,P.A., Clifton,I.J., Schneider,T.R., Schofield,C.J. and Hajdu,J. (2001) X-ray snapshots of serine protease catalysis reveal a tetrahedral intermediate. *Nat Struct. Biol* **8**, 689-694.

Wolf,B., Grier,R.E., Secor,M., Jr. and Heard,G.S. (1985) Biotinidase deficiency: a novel vitamin recycling defect. *J. Inherit. Metab Dis.* **8 Suppl 1**, 53-58.

Wolf,B. (2001) Disorders in biotin metabolism. Scriver,C.R., Beaudet,A.L., Sly,W.S. and Valle,D. (Eds.), The metabolic and molecular bases of inherited disease. *McGraw-Hill, New York*, 3935-3962.

Woods,M.J., Edgeworth,M.A. and Orsi,B.A. (1975) Studies on the active site of the amidase from Pseudomonas aeruginosa. *Biochem. Soc. Trans.* **3**, 1216-1219.

Wyatt, J. and Knowles, C. (1995) Microbial degradation of acrylonitrile waste effluents: the degradation of effluents and condensates from the manufacture of acrylonitrile. *Int. Biodeterior. Biodegrad.* **35**, 227-248.

Yamada,H. and Kobayashi,M. (1996) Nitrile hydratase and its application to industrial production of acrylamide. *Biosci. Biotechnol. Biochem.* **60**, 1391-1400.

Yamamoto,K., Oishi,K., Fujimatsu,I. and Komatsu,K. (1991) Production of R-(-)-mandelic acid from mandelonitrile by Alcaligenes faecalis ATCC 8750. *Appl. Environ. Microbiol.* **57**, 3028-3032.

Yamamoto,K., Ueno,Y., Otsubo,K., Kawakami,K. and Komatsu,K. (1990) Production of S-(+)-ibuprofen from a nitrile compound by Acinetobacter sp. strain AK226. *Appl. Environ. Microbiol.* **56**, 3125-3129.

Yan,J.X., Kett,W.C., Herbert,B.R., Gooley,A.A., Packer,N.H. and Williams,K.L. (1998) Identification and quantitation of cysteine in proteins separated by gel electrophoresis. *J. Chromatogr. A* **813**, 187-200.

Zalkin,H. and Smith,J.L. (1998) Enzymes utilizing glutamine as an amide donor. *Adv. Enzymol. Relat Areas Mol. Biol.* **72**, 87-144.

Zhang,Y., Kua,J. and McCammon,J.A. (2002) Role of the catalytic triad and oxyanion hole in acetylcholinesterase catalysis: an ab initio QM/MM study. *J Am. Chem Soc.* **124**, 10572-10577.

Appendices

Appendix I: Size exclusion chromatography protein standard curve that was used for calibrating the masses of the eluted mutant proteins. Panel (A) is the elution profile of the protein standards (Biorad), names and masses of which are tabulated in table (I) below. Panel (B) represents the fit of the retention time (Kav) versus the logarithm of the molecular masses (Log MM) of protein markers.

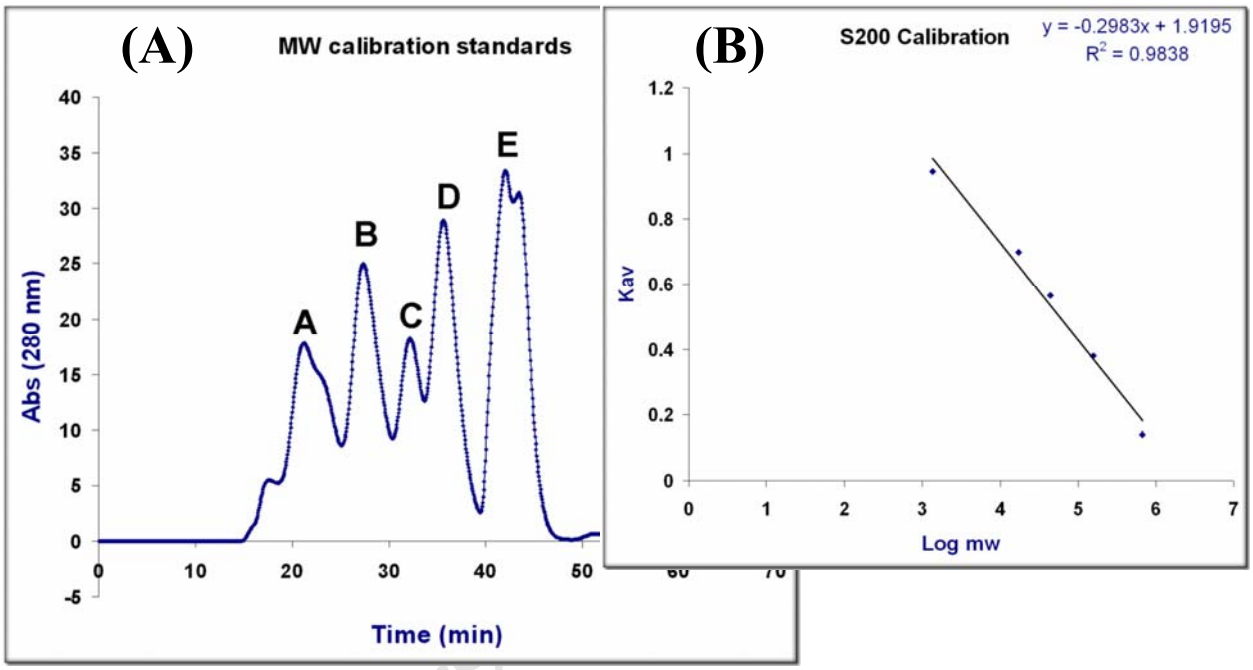
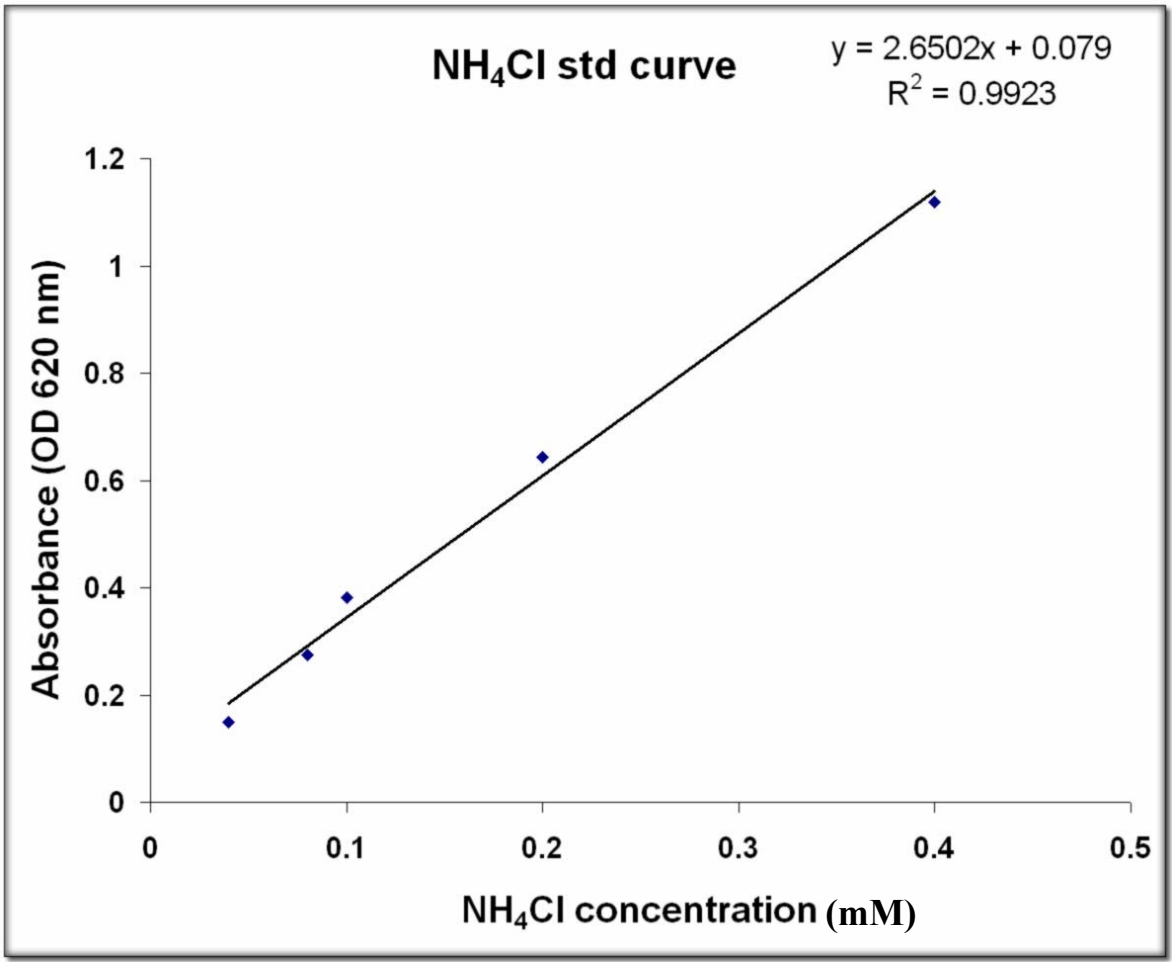


Table (I) Names and masses of the protein standards with the corresponding peak number on the elution profile in panel (A) above.

Peak number	Protein standard	MW (Da)	Log MW
A	Thyroglobulin	670 000	5.826075
B	γ-globulin	158 000	5.198657
C	Ovalbumin	44 000	4.643453
D	Myoglobin	17 000	4.230449
E	Vitamin B12	1 350	3.130334

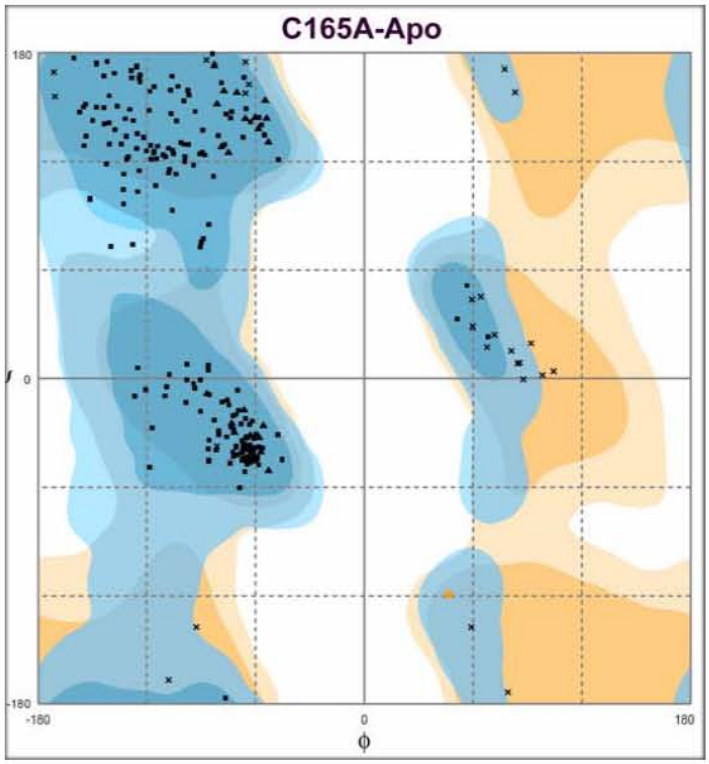
Appendix II: An example of an indophenol-blue ammonia assay standard curve that was constructed from the OD (620 nm) readings of triplicate samples containing 0.05 mM to 0.4 mM ammonium chloride (NH_4Cl) standards. The equation of the linear line was used to determine the amount of ammonia present in the reaction mixtures of the WT and the mutant proteins with various amide and nitrile substrates.



Appendix III: Ramachandran plots for all residues in the final models as calculated by *RAMPAGE* (Lovell et al. 2003). The plots are labeled with the name of the dataset and the statistics are given below each plot. Fig. (I) below represents the key to the plot coloring and residues' representation as generated by *RAMPAGE*.

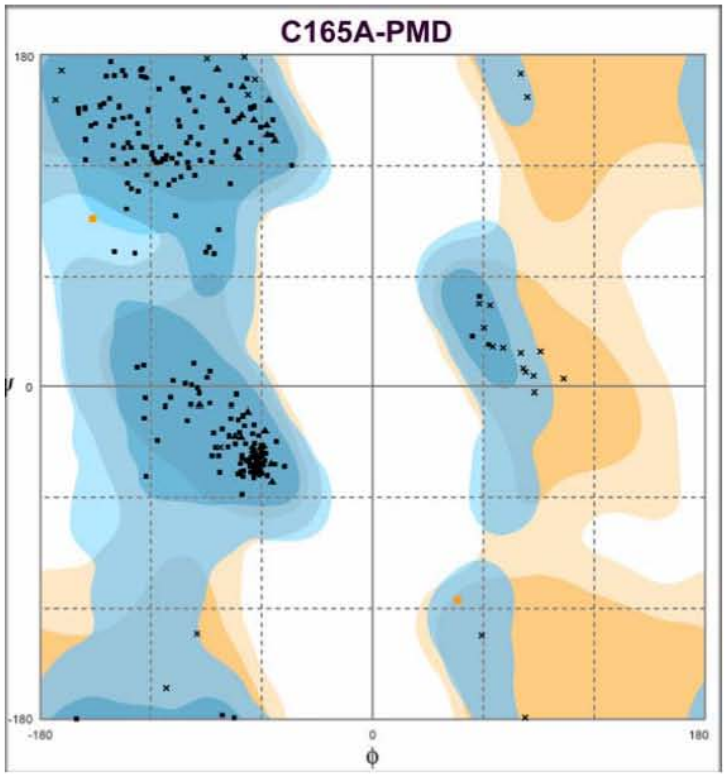
Fig. (I): Key to the Ramachandran plots coloring and residues' representation.

(A)



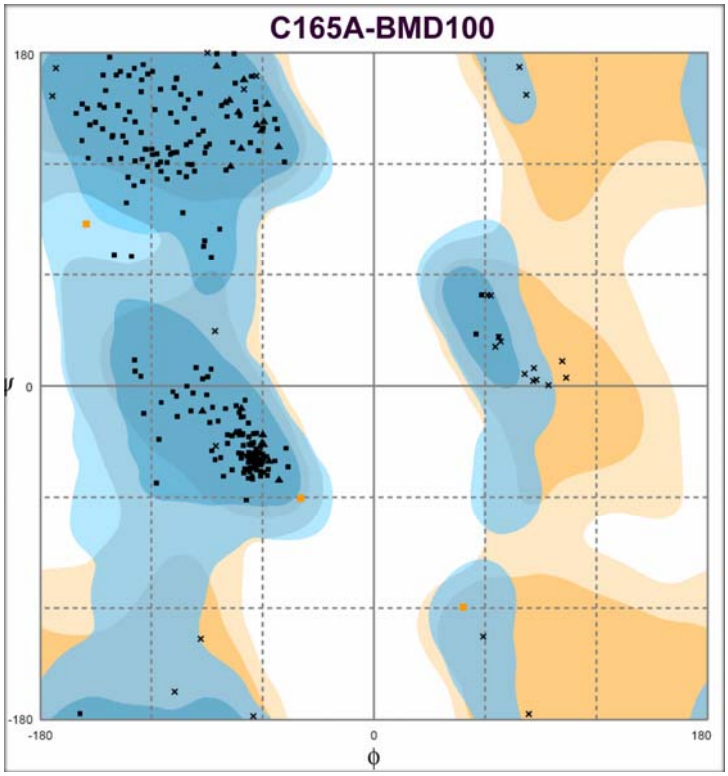
Number of residues in favoured region (~98.0% expected) : 251 (99.6%)
Number of residues in allowed region (~2.0% expected) : 1 (0.4%)
Number of residues in outlier region : 0 (0.0%)

(B)



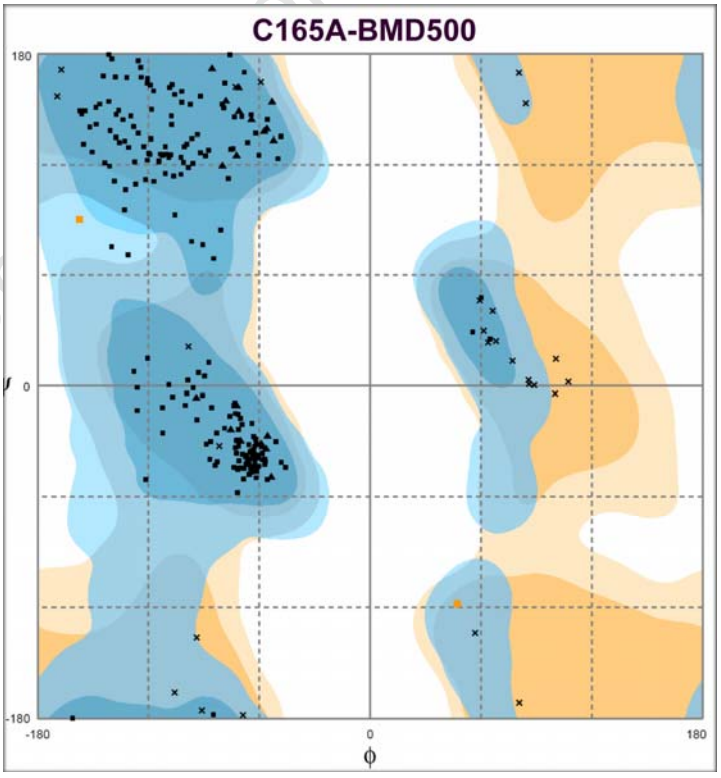
Number of residues in favoured region (~98.0% expected) : 252 (99.2%)
Number of residues in allowed region (~2.0% expected) : 2 (0.8%)
Number of residues in outlier region : 0 (0.0%)

(C)



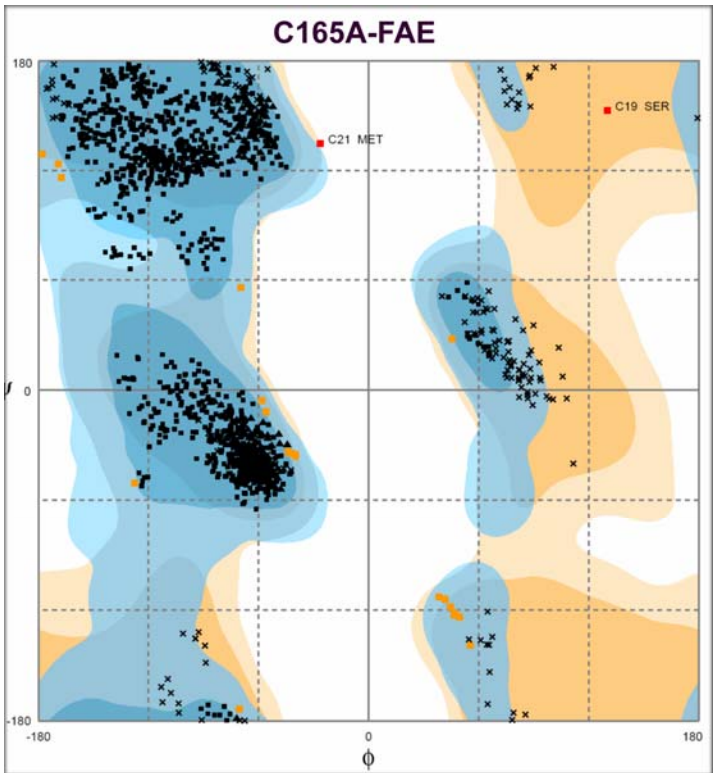
Number of residues in favoured region (~98.0% expected) : 254 (98.8%)
Number of residues in allowed region (~2.0% expected) : 3 (1.2%)
Number of residues in outlier region : 0 (0.0%)

(D)



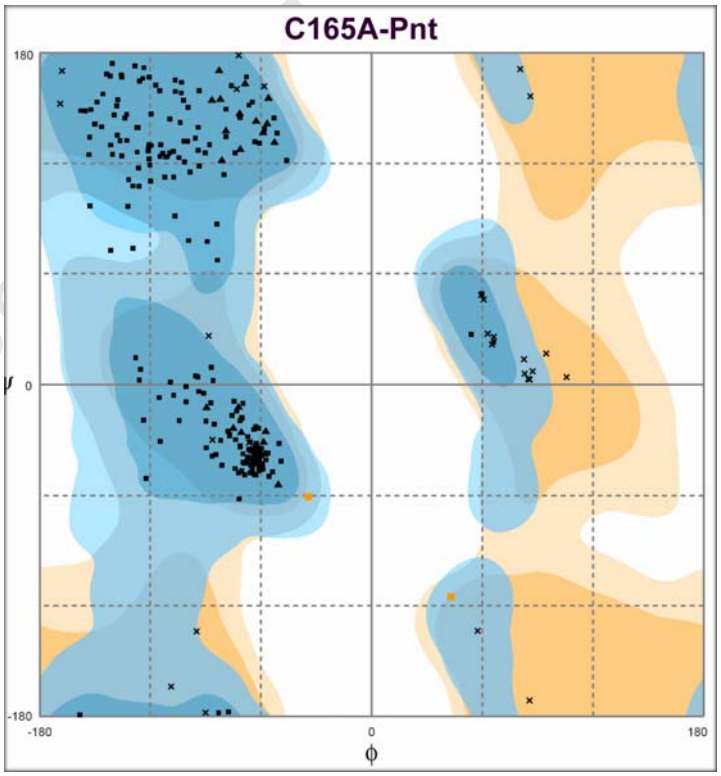
Number of residues in favoured region (~98.0% expected) : 254 (99.2%)
Number of residues in allowed region (~2.0% expected) : 2 (0.8%)
Number of residues in outlier region : 0 (0.0%)

(E)



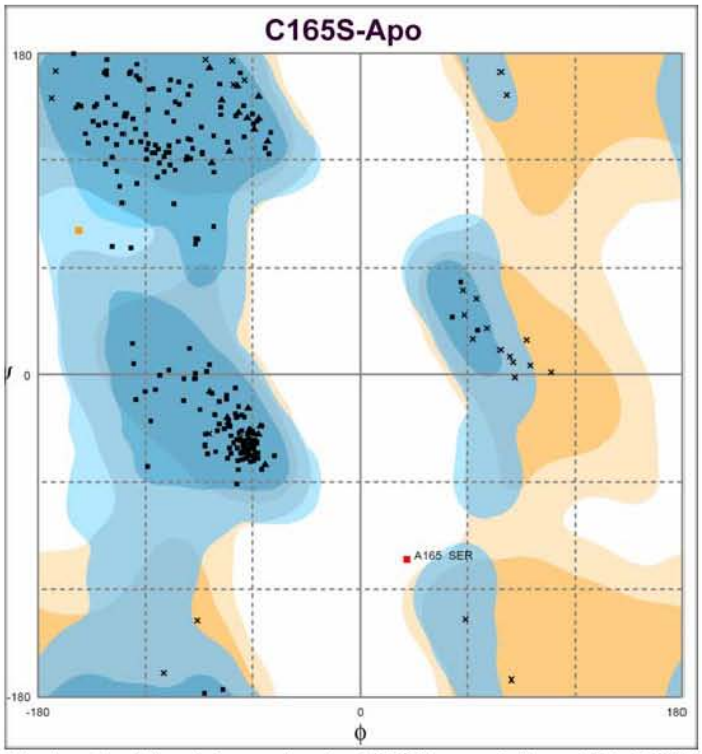
Number of residues in favoured region (~98.0% expected)	: 1520 (98.6%)
Number of residues in allowed region (~2.0% expected)	: 20 (1.3%)
Number of residues in outlier region	: 2 (0.1%)

(F)



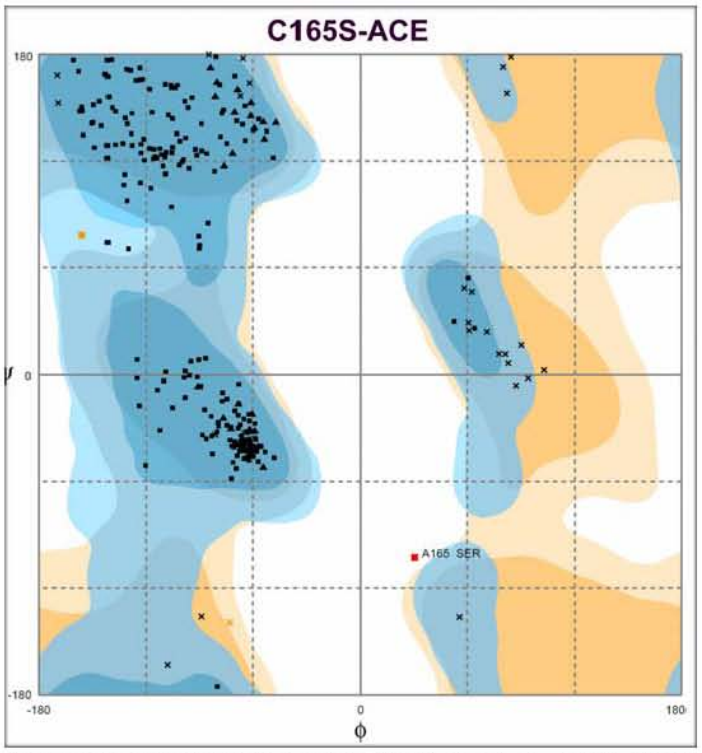
Number of residues in favoured region (~98.0% expected)	: 256 (99.2%)
Number of residues in allowed region (~2.0% expected)	: 2 (0.8%)
Number of residues in outlier region	: 0 (0.0%)

(G)



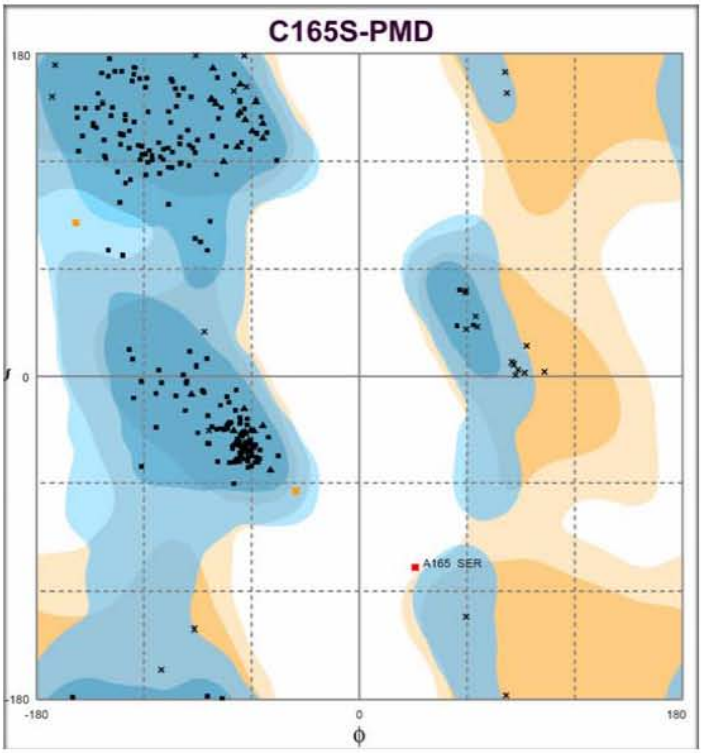
Number of residues in favoured region (~98.0% expected)	: 251 (99.2%)
Number of residues in allowed region (~2.0% expected)	: 1 (0.4%)
Number of residues in outlier region	: 1 (0.4%)

(H)



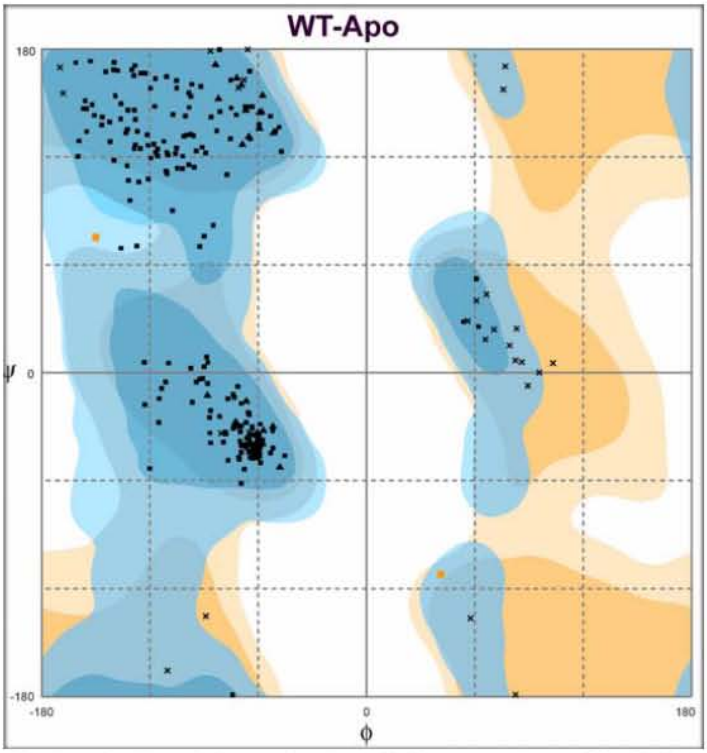
Number of residues in favoured region (~98.0% expected)	: 254 (98.8%)
Number of residues in allowed region (~2.0% expected)	: 2 (0.8%)
Number of residues in outlier region	: 1 (0.4%)

(I)



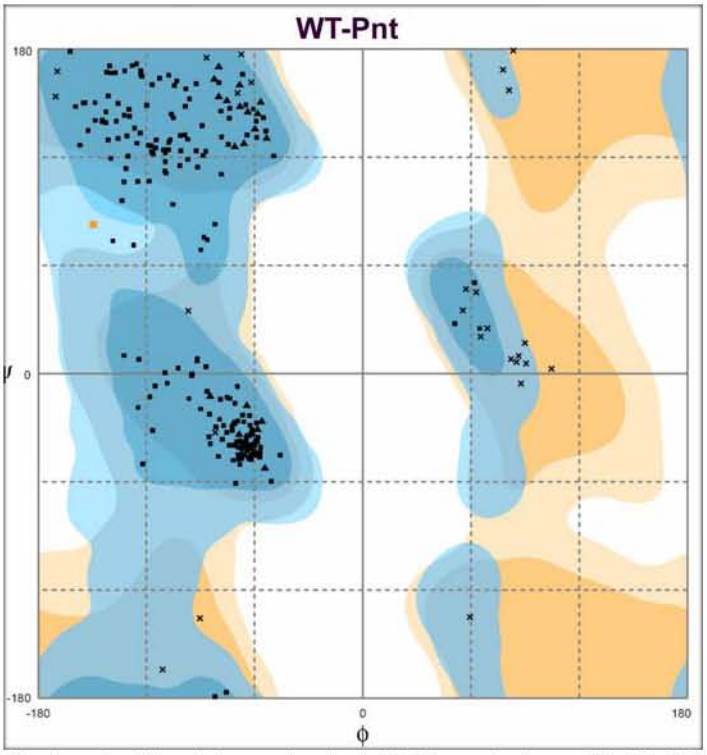
Number of residues in favoured region (~98.0% expected) : 254 (98.8%)
Number of residues in allowed region (~2.0% expected) : 2 (0.8%)
Number of residues in outlier region : 1 (0.4%)

(J)



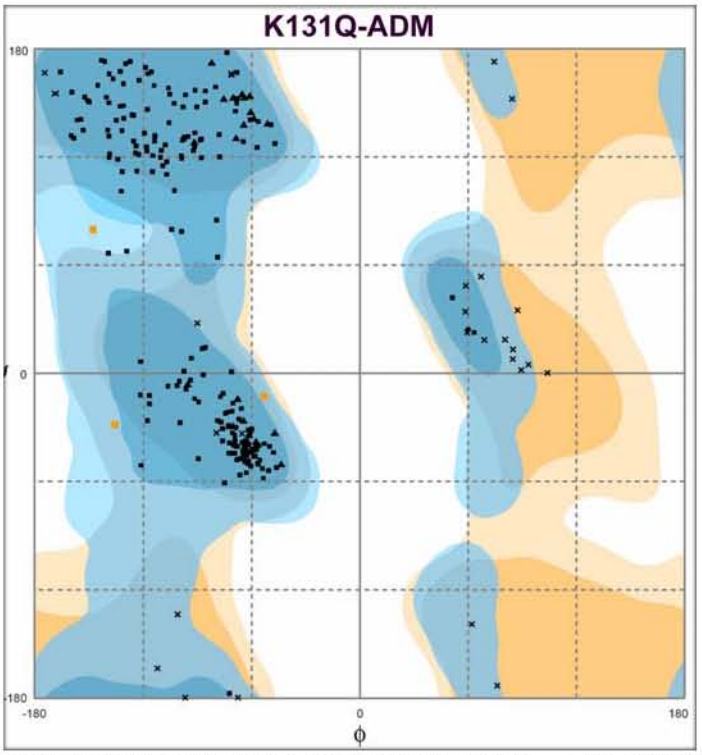
Number of residues in favoured region (~98.0% expected) : 253 (99.2%)
Number of residues in allowed region (~2.0% expected) : 2 (0.8%)
Number of residues in outlier region : 0 (0.0%)

(K)



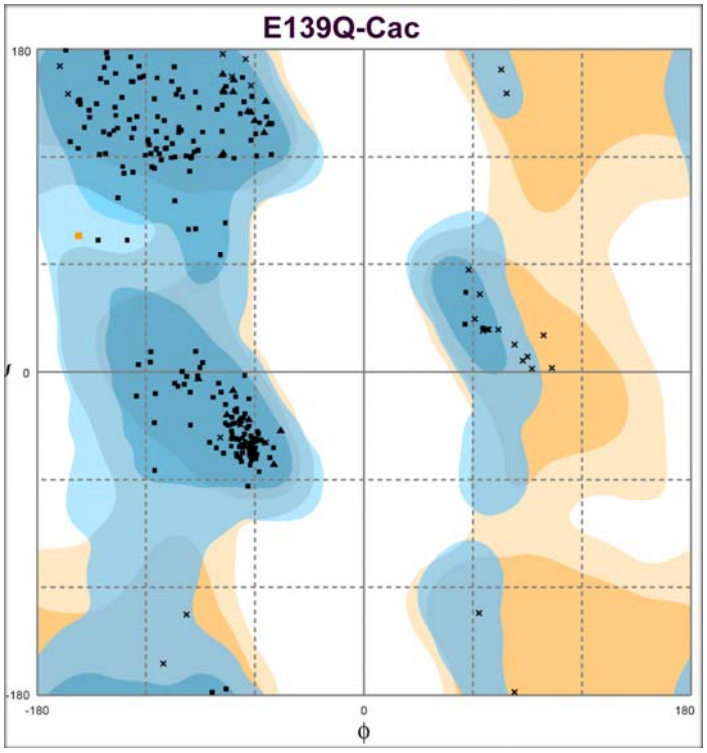
Number of residues in favoured region (~98.0% expected)	: 254 (99.6%)
Number of residues in allowed region (~2.0% expected)	: 1 (0.4%)
Number of residues in outlier region	: 0 (0.0%)

(L)



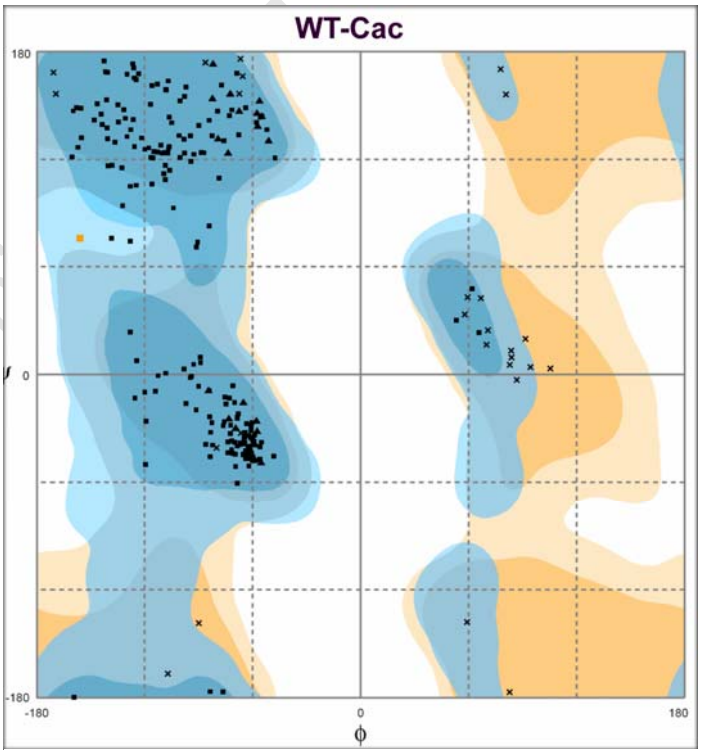
Number of residues in favoured region (~98.0% expected)	: 251 (98.8%)
Number of residues in allowed region (~2.0% expected)	: 3 (1.2%)
Number of residues in outlier region	: 0 (0.0%)

(M)



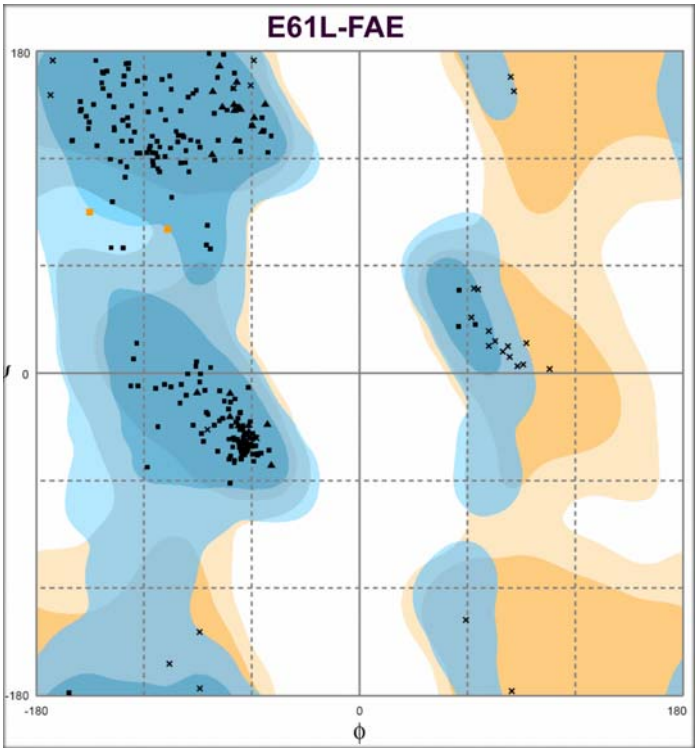
Number of residues in favoured region (~98.0% expected)	: 254 (99.6%)
Number of residues in allowed region (~2.0% expected)	: 1 (0.4%)
Number of residues in outlier region	: 0 (0.0%)

(N)



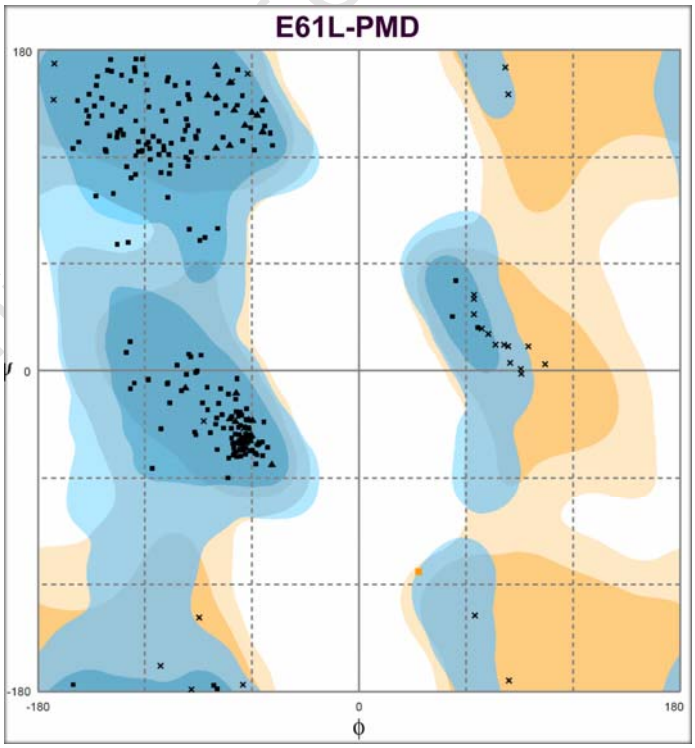
Number of residues in favoured region (~98.0% expected)	: 246 (99.6%)
Number of residues in allowed region (~2.0% expected)	: 1 (0.4%)
Number of residues in outlier region	: 0 (0.0%)

(O)



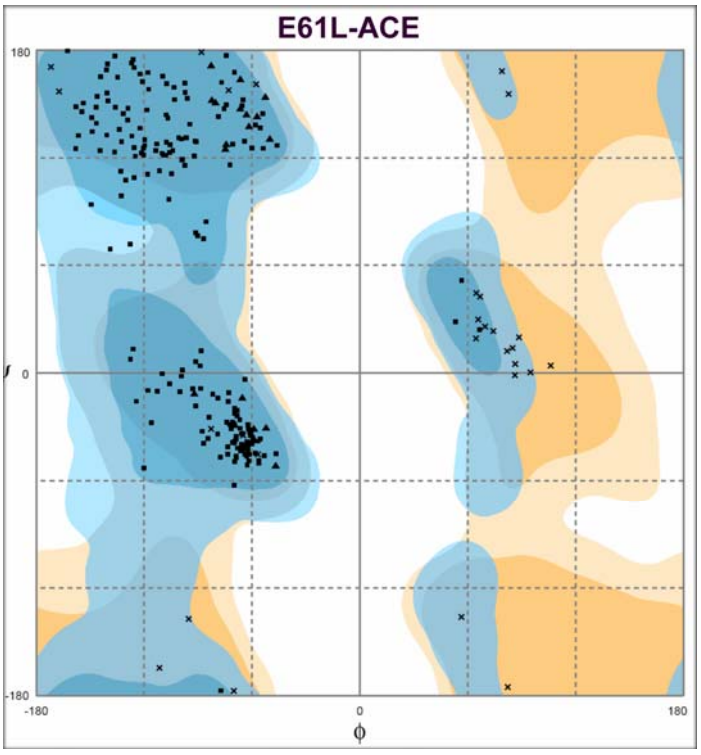
Number of residues in favoured region (~98.0% expected) : 253 (99.2%)
Number of residues in allowed region (~2.0% expected) : 2 (0.8%)
Number of residues in outlier region : 0 (0.0%)

(P)



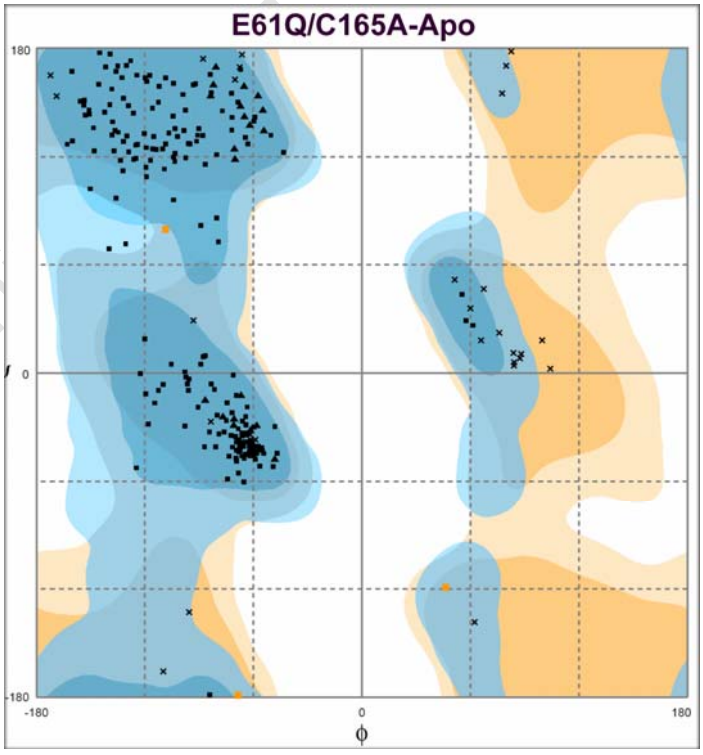
Number of residues in favoured region (~98.0% expected) : 256 (99.6%)
Number of residues in allowed region (~2.0% expected) : 1 (0.4%)
Number of residues in outlier region : 0 (0.0%)

(Q)



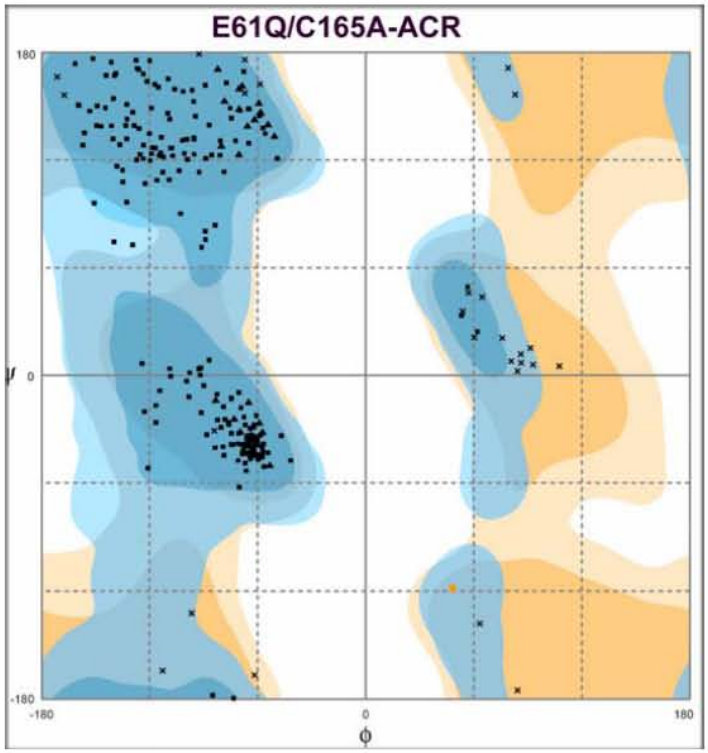
Number of residues in favoured region (~98.0% expected) : 251 (100.0%)
Number of residues in allowed region (~2.0% expected) : 0 (0.0%)
Number of residues in outlier region : 0 (0.0%)

(R)



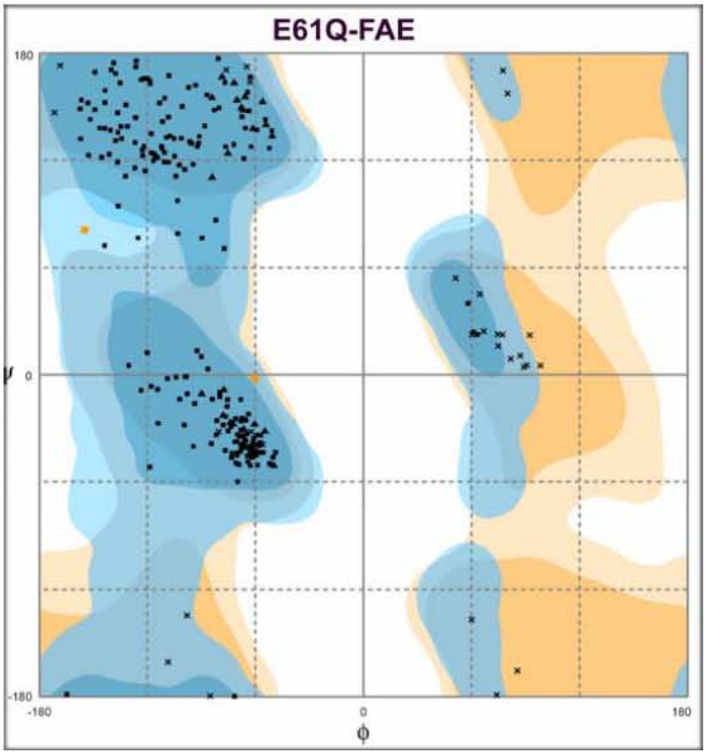
Number of residues in favoured region (~98.0% expected) : 254 (98.8%)
Number of residues in allowed region (~2.0% expected) : 3 (1.2%)
Number of residues in outlier region : 0 (0.0%)

(S)



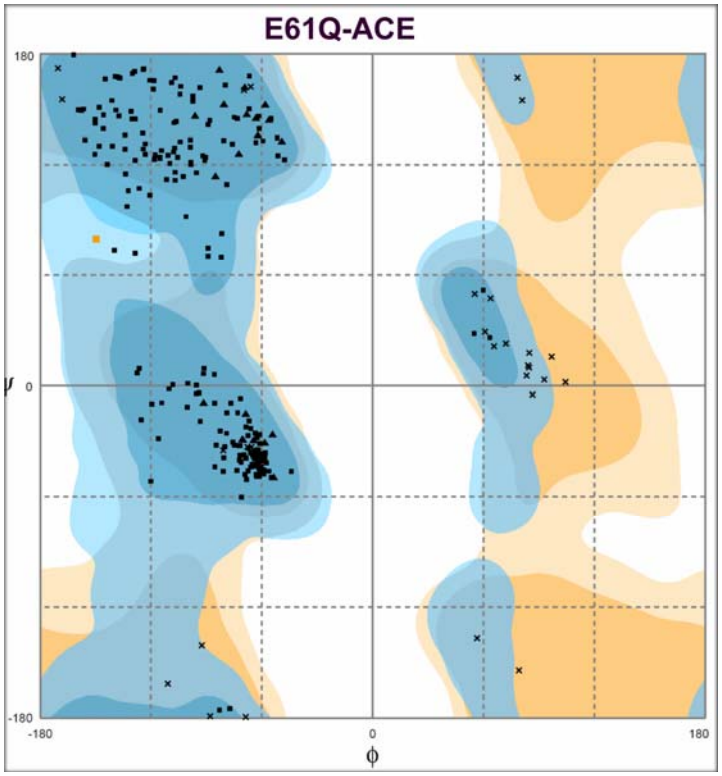
Number of residues in favoured region (~98.0% expected)	: 254 (99.6%)
Number of residues in allowed region (~2.0% expected)	: 1 (0.4%)
Number of residues in outlier region	: 0 (0.0%)

(T)



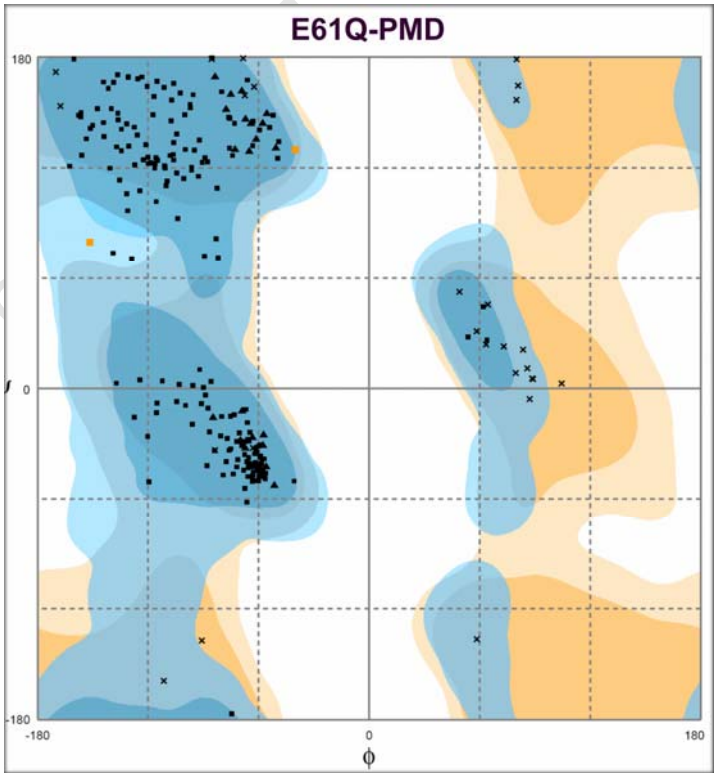
Number of residues in favoured region (~98.0% expected)	: 254 (99.2%)
Number of residues in allowed region (~2.0% expected)	: 2 (0.8%)
Number of residues in outlier region	: 0 (0.0%)

(U)



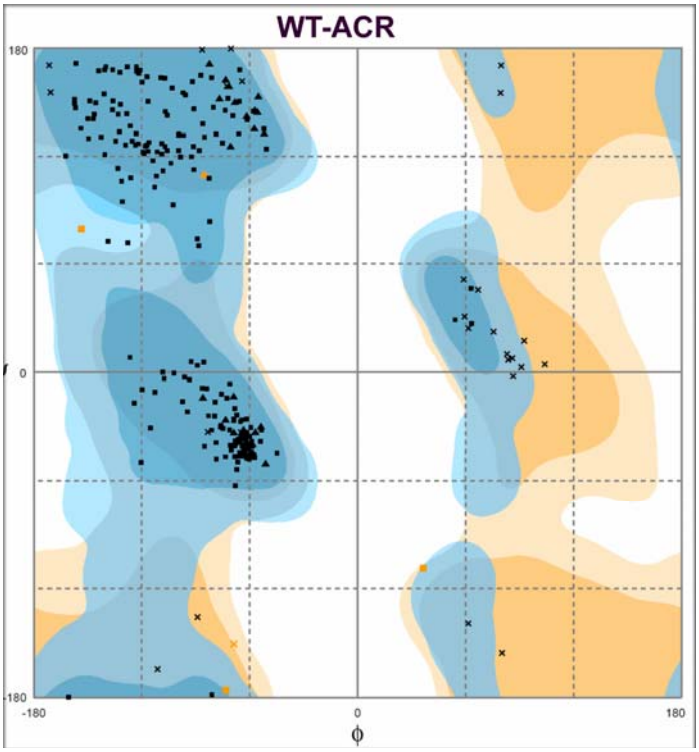
Number of residues in favoured region (~98.0% expected) : 253 (99.6%)
Number of residues in allowed region (~2.0% expected) : 1 (0.4%)
Number of residues in outlier region : 0 (0.0%)

(V)



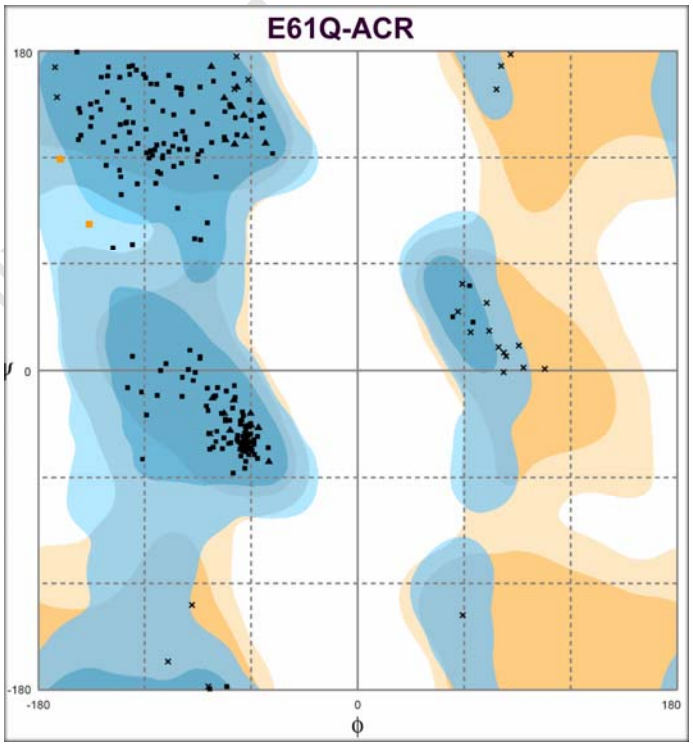
Number of residues in favoured region (~98.0% expected) : 249 (99.2%)
Number of residues in allowed region (~2.0% expected) : 2 (0.8%)
Number of residues in outlier region : 0 (0.0%)

(W)



Number of residues in favoured region (~98.0% expected) : 250 (98.0%)
Number of residues in allowed region (~2.0% expected) : 5 (2.0%)
Number of residues in outlier region : 0 (0.0%)

(X)



Number of residues in favoured region (~98.0% expected) : 242 (99.2%)
Number of residues in allowed region (~2.0% expected) : 2 (0.8%)
Number of residues in outlier region : 0 (0.0%)

Appendix IV: MALDI-TOF spectrum for the E61L NitN mutant reacted with propionamide (PMD), showing a weak PMD thioester intermediate fragment peak (*m/z* 2097) and a dominant free protein (unreacted Cys165) tryptic peptide peak (*m/z* 2041).