The Crystal structure of an aliphatic amidase from *Geobacillus pallidus* RAPc8

Serah Kimani

A minithesis submitted in partial fulfilment of the requirements for the degree of Master of Science (Structural Biology) at the University of Cape Town

Supervisors:

Associate Professor B. T. Sewell and Dr. Muhamed Sayed

February 2007

KEYWORDS

Active site

Amidase

Amides

Catalytic triad

Dimer

Geobacillus pallidus RAPc8

Hexamer

Homologue

Molecular replacement

Nitrilase

Nitriles

Superfamily

Trimer

ABBREVIATIONS

1D - 1-dimensional

2D - 2-dimensional

3D - 3-dimensional

6D - 6-dimensional

BSA - Bovine Serum Albumin

CC - correlation coefficient

CCP4 - Collaborative Computational Project, Number 4

DCase - N-carbamyl-D-amino acid amidohydrolase from *Agrobacterium* sp. KNK712

DM - Density modification

DTT - dithiothreitol

EM - Electron microscopy

FOM - Figure of Merit

HOMSTRAD - Homologous Structures Alignment Database

LLG - The Log Likelihood Gain

MSA - Multivariate Statistical Analysis

mM – millimolar

MR - molecular replacement

NCS - Non-Crystallographic Symmetry

NHase - Nitrile hydratase

Nit – Nitrilase domain of the *C. elegans* NitFhit fusion protein

ORF - Open Reading Frame

PamiE - A nitrilase-related aliphatic amidase from *Pseudomonas aeruginosa*

PCA - Principal Component Analysis

PC - Patterson correlation

PDB - Protein Data Base

RMSD - Root Mean Square Deviation

SDS-PAGE - sodium dodecyl sulfate-polyacrylamide gel electrophoresis

S/N - Signal-to-Noise

Vm - Matthews coefficient

ABSTRACT

The crystal structure of an aliphatic amidase from *Geobacillus pallidus* RAPc8

S. Kimani

M.Sc. minithesis, Department of Molecular & Cell Biology, University of Cape Town

Amidases are a group of carbon-nitrogen hydrolysing enzymes that catalyze the conversion of amides to corresponding carboxylic acids and ammonia. These enzymes are of great interest in synthetic industries where they are used for mass production of acidic products.

Aliphatic amidase from *Geobacillus pallidus* RAPc8 (RAPc8 amidase), which belongs to the nitrilase superfamily of enzymes, has recently been characterised biochemically. It shows both amide hydrolysis and acyl transfer activities, and also exhibits stereoselectivity for some enantiomeric substrates. This enzyme can therefore be exploited in large-scale production of enantio-pure compounds. Structural characterization of this amidase would yield insights into the basis of this substrate selectivity and activity. This would inform future experiments that aims at modifying this enzyme to alter its substrate specificity.

This work presents structural characterization of RAPc8 amidase. Gel filtration chromatography and electron microscopic analyses provided useful information on the quaternary structure of RAPc8 amidase. Crystals were grown, and an X-ray diffraction dataset to 1.9 Å collected using an in-house X-ray source. The space group of this data was determined to be primitive cubic $P4_232$, and the structure was solved by molecular replacement using the backbone of the hypothetical protein PH0642 from *Pyrococcus horikoshii* (PDB ID, 1j31) that had all non-identical side chains substituted with alanines, as a search probe. The molecular replacement rotational and translational searches were performed using *PHASER*. The model was rebuilt with *PHENIX* before refinement using *REFMAC5*. The final model was of high quality with minimal errors.

RAPc8 amidase is homohexameric in solution and has a four-layer α-β-β-α structural fold that highly resembles nitrilase superfamily enzymes. It has an extended C-terminal tail that is essential for strengthening the interacting dimer interfaces by participating in domain swapping. The active site pocket has Glu, Lys, Cys catalytic triad that is conserved in the nitrilase superfamily. The substrate binding pocket is small in size, explaining the specificity of this enzyme for short aliphatic amides.

These findings have made steps towards understanding the catalytic mechanism, and the basis for substrate specificity in this enzyme. It has also provided useful information on the overall structure, as well as the structure of the active site, not only for RAPc8 amidase but also for related enzymes, which will form the basis for designing future structural characterization work in the nitrilase-related amidases.

February 2007

DECLARATION

I declare that *The crystal structure of an aliphatic amidase from Geobacillus pallidus RAPc8* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Serah Kimani 27^{th} February 2007

Signed:

ACKNOWLEDGEMENTS

I would like to thank the following:

My supervisor, Associate Professor Trevor Sewell, for expert guidance and support

My co-supervisor Dr. Muhammed Sayed and fellow scientists who have contributed to this work; Vinod Agarkar for purifying and crystallizing RAPc8 amidase; Brandon Weber for performing gel filtration analysis and for preparing electron micrographs; Jason Von Rooyen for performing electron micrographs analysis

Jean Watermayer for continued support, encouragement and proofreading

All structural biology students and staff

Carnegie Foundation of the New York for supporting structural biology program, Polio Research Foundation and the University of Cape Town for funding

My family, for prayers, support, patience and encouragement

God almighty, the giver of all knowledge and wisdom

CONTENTS

CHAPTER 1.: LITERATURE REVIEW

1.1. General Introduction

The formation and cleavage of carbon-nitrogen (C-N) bonds are central processes in both eukaryotic and prokaryotic organisms. While the processes of peptide bond formation by ribosomes (Moore, Steitz, 2002) and non-ribosomal peptide synthetases (Keating et al., 2002), and the cleavage of the same through proteolytic activities of proteases (Rawlings et al., 2002) are well documented, the metabolism of non-peptide C-N bonds is still being investigated, and a body of knowledge is beginning to emerge. Non-peptide C-N bond hydrolysis reactions occur in plants, animals and fungi where they play an important role in the production of natural substances such as auxin and biotin, etc which are required for deamination of protein and amino acid substrates (Pace, Brenner, 2001). In plants particularly, these activities are implicated in nutrient metabolism, as well as in the degradation of toxic cyanogenic compounds (Piotrowski et al., 2001). On the other hand, C-N bond condensation reactions are important in biochemical processes, including post-translational modification of amino acids, proteins and other compounds. C-N bond reactions are also observed in bacteria and archea, particularly those that have an ecological relationship with plants and animals (Pace, Brenner, 2001).

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]. The hydrolysis of these compounds is mainly performed by nitrilase superfamily enzymes (Pace and Brenner, 2001) that attack either the cyano carbon of a linear nitrilase substrate or the planar carbon of an amide substrate, using a conserved cysteine residue (Stevenson et al., 1990; Bork, Koonin, 1994). Enzymes from other families, including signature amidases (Chebrou et al., 1996; Patricelli, Cravatt, 2000), N-terminal nucleophile hydrolases and amidotransferases (Zalkin, Smith, 1998) also acts on C-N bond-containing substrates, although they are structurally and mechanistically unrelated to nitrilase superfamily members (Brenner, 2002).

1.1.1. Nitrile metabolism and nitrile-degrading enzymes

Although nitrile-metabolism activities are relatively infrequent in plants and animals, they are commonly observed in bacteria (including *Acinetobacter, Corynebacterium, Arthrobacter, Pseudomonas, Klebsiella, Norcadia, Bacillus, Rhocococcus*, etc), that metabolize nitriles as a sole source of carbon and nitrogen (Banerjee et al., 2002). However the physiological role of nitrile-degrading enzymes in these microbes is not fully understood. Nitriles, which are products of aldoxime degradation (observed in plants; (Kato et al., 2000), as well as abiotic conversion of metal cyanides (Banerjee et al., 2002), enter a number of metabolic pathways (Figure 1-1), including hydrolysis, oxidation (oxygenase; Sawyer et al., 1984) and reduction (nitrogenase; Liu et al., 1997) by various enzymes.

Nitrile hydrolysis (Figure 1-1) in microbes follows two pathways: (1) A single enzymatic pathway that is catalyzed by nitrilases and that involves conversion of organic nitriles to corresponding acids and ammonia, and (2), a bi-enzymatic pathway that involves hydration of nitriles to corresponding amides by nitrile hydratases (NHases), followed by conversion of amides to corresponding organic acids and ammonia by amidases. Cyanide dihydratase and cyanide hydratase enzymes are closely related to nitrilases in terms of amino acid sequence similarities and protein structure, but unlike nitrilase enzymes which have a wide substrate specificity, these enzymes only use inorganic cyanide (H-C≡N) substrates efficiently to produce acid and amide products, respectively (O'Reilly, Turner, 2003).

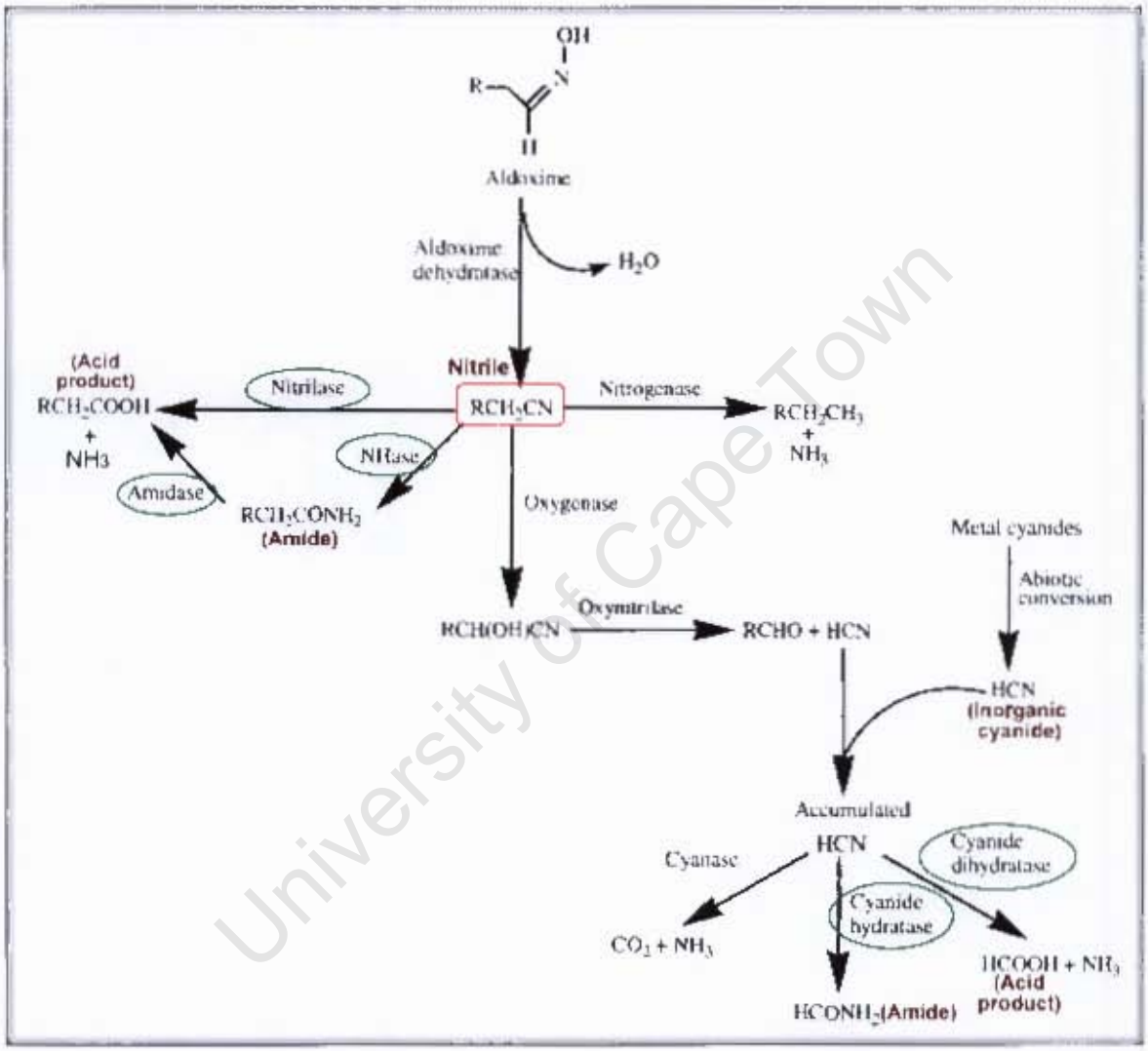


Figure 1-1: Pathways of nitrile metabolism. Enzymes from the nitrilase superfamily are circled in green and various nitrile metabolism products involving nitrilase superfamily enzymes are labelled. The Figure was taken from Banerjee et al. (2003) and modified slightly.

1.1.2. Applications of nitrile-degrading enzymes

Micro-organisms containing nitrile-metabolizing enzymes have a great potential as synthetic biocatalysts in chemical industries, as well as in environmental bioremediation (Banerjee et al., 2002). The *in vitro* use of nitrile-degrading enzymes is being explored.

Nitrile compounds are used in synthetic chemical industries as important intermediates for providing amides, amines, amidines, carboxylic acids, esters, drug intermediates and pharmaceuticals (Banerjee et al., 2002; Fournand, Arnaud, 2001), among other compounds. They are also useful for the manufacture of a variety of polymers including polyacrylonitrile (acrylonitriles) and nylon-6:6 (adiponitriles) (Banerjee et al., 2002). The possibility of utilizing the biotransformation capabilities of nitrile-degrading enzymes has been explored, as a replacement of the traditional chemical-based nitrile conversion methods (Figure 1-2) which have several drawbacks: Aside from being cost-ineffective, chemical hydrolysis of nitriles requires harsh conditions (Fournand, Arnaud, 2001; Banerjee et al., 2002) such as heating at strong acidic or alkaline pH, which makes selective transformation unachievable particularly in cases of labile substrates and products. In addition, formation of by-products such as toxic inorganic cyanides and salts (Banerjee et al., 2002) impedes the production of pure products. Biocatalytic conversions are therefore attractive, as the hydrolysis proceeds at mild pH and temperature conditions. The observed chemo-, regio-, and enantio-selective properties (Yamamoto et al., 1990; Yamamoto et al., 1991; Banerjee et al., 2002) of these enzymes can be utilized to produce enantio-pure products. An example of production mass biotransformation is the application of the nitrile hydratase from *Rhodococcus rhodochrous* J1 by the Nitto Chemistry Industry Company Ltd in Japan to produce 30 000 tons of acrylamide annually as reported by Yamada and Kobayashi (1996). Moreau and colleagues (1993) also reported the use of a nitrile hydratase and amidase couple in *Rhodococcus* sp. R312 strain to produce adipic acid from adiponitrile. Adipic acid is one of the raw materials in the manufacture of nylon-6:6.

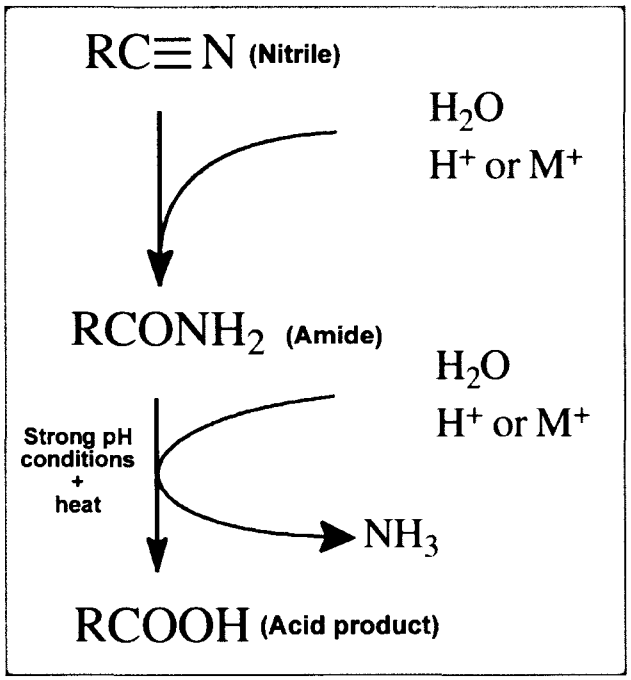


Figure 1-2: Chemical pathway of nitrile hydrolysis: The activation of carbon-nitrogen triple bond is achieved by either hydrogen protons or metal cationic species (M^+). The conversion of amide to acid is more resistant to hydrolytic cleavage, hence requires harsh conditions of temperature and pH for total hydrolysis. The figure was taken from Fournand and Arnaud (2001).

Most nitrile compounds are reported to be highly toxic, carcinogenic and mutagenic (Pollak *et al.*, 1991), and as industrial wastewater ends up in 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, NHases and amidases to biodegrade acrylonitrile-containing effluent from acrylonitrile-manufacturing industries has been reported (Wyatt & 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 as reported by Freyssinet and others (1996). Another example is the soil bacterium, *Agrobacterium radiobacter,* which has the potential of degrading bromoxynil herbicide (Muller, Gabriel, 1999).

1.2. The Nitrilase superfamily enzymes

Initially, members of the nitrile-hydrolyzing superfamily were reported to include nitrilases, cyanide hydratases, aliphatic amidases, β-ureidopropionases, β-alanine synthases and N-carbamyl-D-amino acid amidohydrolases (Bork, Koonin, 1994). On the basis of sequence similarity and domain fusion characteristics, Pace and Brenner (2001) re-classified nitrilase superfamily enzymes into 4 major groups (nitrilases, amidases, carbamylases and N-acyltransferases), distributed in 13 different branches. The four major reactions involving nitrilase superfamily enzymes are shown in Figure 1-3 below.

Branch 1 enzymes have nitrilase activity and consist of nitrilases, cyanide dihydratases and cyanide hydratases (O'Reilly, Turner, 2003; Pace, Brenner, 2001). Eight of the other branches (branches 2, 3, 4, 5, 6, 7, 8 and 9) consist of amidases with varying substrate specificities: branches 2, 3, and 4 consist of aliphatic amidases, amino-terminal amidases and biotinidases; branches 5 and 6 are amidases with carbamylase activity, comprising of β-ureidopropionases and carbamylases respectively; branches 7 and 8 are fusion proteins, with a nitrilase-related domain that has amidase activity specific for glutamine hydrolysis in prokaryotes (branch 7) and eukaryotes (branch 8); branch 9 consists of apolipoprotein N-acyltransferase enzymes that perform an amidase condensation reaction, transferring a fatty acid to polypeptide amino terminus. Branch 10 is likely to consist of fusion proteins. The only known member of this branch is the Nit domain of the worm (*C. elegans*) NitFhit "Rosetta stone" fusion protein, whose function is not known (Pace et al., 2000). The function of branch 11 enzymes is not very clear, but a new member of the group, N-carbamyl putrescine amidohydrolase that catalyzes the metabolism of arginine into spermidine and succinate (Nakada et al., 2001b) has been identified. This enzyme is related to the β–ureidopropionases of branch 5 (Nakada et al., 2001b). The function of branch 12 has not been confirmed, but these enzymes are thought to play a role in protein post-translational modification based on the fusion of their nitrilase-related domain with RimI N-terminal acetyltransferases (Brenner, 2002). Branch 13 consists of non-fused

outliers with no known function. A summary of the activity and the domain structure for the 13 branches of the nitrilase superfamily is found in Table 1-1.

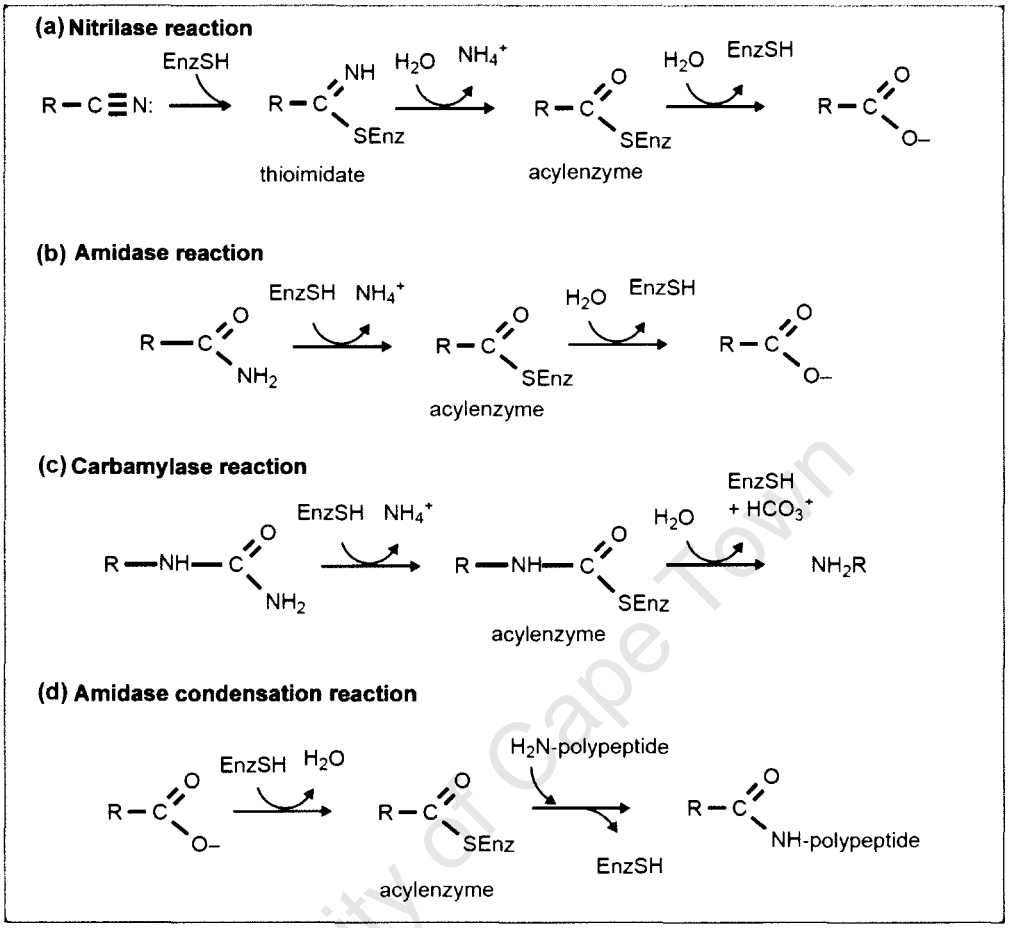
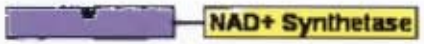


Figure 1-3: Reaction types in the four groups of nitrilase superfamily enzymes. Nitrilase activity is observed in branch 1 enzymes. Amidase activity is observed in branches 2-4 and in nitrilase-related domains of branch 7 and 8 enzymes. Carbamylase activity is observed in branches 5 and 6. Amidase condensation is observed in branch 9. All reactions have been proposed to proceed through acylenzyme intermediates as depicted. The figure was taken from Pace and Brenner (2001).

Four crystal structures of enzymes in different branches of the nitrilase superfamily are now available. These are: the Nit domain of the worm NitFhit fusion protein (Nit; PDB ID, 1ems); (Pace et al., 2000), *Agrobacterium* N-carbamyl-D-amino acid amidohydrolase (DCase; PDB ID, 1erz); (Nakai et al., 2000), putative C-N hydrolase from yeast (PDB ID, 1f89); (Kumaran et al., 2003) and hypothetical protein PH0642 from *Pyrococcus horishii* (PH0642; PDB ID, 1j31); (Sakai et al., 2004). Only the carbamylase (DCase, in branch 6) has been well characterized both structurally and biochemically (Chen et al., 2003), with a number of substrate-bound active site

mutant structures (1uf4, 1uf5, 1uf7 and 1uf8) (Hashimoto *et al.*, unpublished data) available in the Brookhaven Protein Data Bank (PDB). The remaining three structures (Nit, yeast C-N hydrolase and PH0642) have not been characterized in their enzymatic context, and their functions still remain to be determined; however, Mueller and colleagues (2006) have recently characterized an enzyme that is similar to the hypothetical PH0642 protein (1j31), as having nitrilase activity. The four nitrilase superfamily structures possess a novel four-layer α-β-β-α sandwich fold that is depicted in Figure 1-4.

Although the four proteins have relatively low average sequence identities (26%), several conserved motifs exist in the structures. This includes a novel Glu, Lys, Cys catalytic triad, which is similarly coordinated and positioned in all structures. In their biological context, the four nitrilase superfamily enzymes have a quaternary structure as summarised in Table 1-2 below. An interesting feature of the association of monomers in the formation of biological complexes is that a conserved dimer interface (Figure 1-5) exists in all four structural homologues, where the monomers involved in the dimer formation are related by 2-fold symmetry. An active site cleft exists in each monomer, as revealed by the analysis of the DCase structure (Nakai et al., 2000).

Nitrilase Branch number & name	Enzymatic reaction	Domain structure	Representative member of the branch Example specific reaction
1-Nitrilase	Nitrilase		***Rhodococcus* Nit A** (Kobayashi *et al.*, 1992) Acrylonitrile + $2H_2O$ → Acrylic acid + NH_3
2-Aliphatic amidase	Amidase		***Pseudomonas* amiE** (Novo *et al.*, 2002) Asparagine + H_2O → Aspartic acid + NH_3
3-Amino-terminal amidase	Amidase		***Saccharomyces* NTA1** (Baker *et al.*, 1995) N-terminal asparagine + H_2O → N-terminal aspartic acid + NH_3
4-Biotinidase	Amidase/ secondary amidase		**Human BTD** (Hymes & Wolf, 1996) Biocytin + H_2O → Biotin + lysine
5-β-Ureidopropionase	Amidase		**Rat BAS** (Kvalnes-Krick & 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-Prokaryote Gln-dependent NAD synthase	Amidase		***Saccharomyces* QNS1** (Pace & Brenner, 2001) Glutamine + H_2O → Glutamic acid + NH_3
8-Eukaryote Gln-dependent NAD synthase	Amidase		Similar reaction to number 7 above

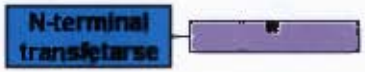
9-Apolipoprotein N-acetyltransferase	Reverse amidase		***Pseudomonas* cutE** (Piotrowski *et al.*, 2001) N-terminal DAG-modified Cys + palmitate → N-palmitylated protein + H_2O
10-Nit	unknown		**Worm NitFhit** (Pace *et al.*, 2000) Unknown reaction
11-N-carbamyl putrescine amidohydrolase	Predicted amidase		***Pseudomonas* AguB** (Nakada *et al.*, 2001) N-carbamyl putrescine + H_2O → Putrescine + CO_2 +NH_3
12-NB12	unconfirmed		***Sphingomonas* 1dhX** (Pace & Brenner, 2001) Unknown reaction
13-Non-fused outliers	unknown		Unknown reaction

Table 1-1: A summary of the 13 branches of the nitrilase superfamily enzymes. Enzymes in seven 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 the table is consolidated from nitrilase superfamily classification papers, Pace and Brenner (2001) and Brenner (2002).

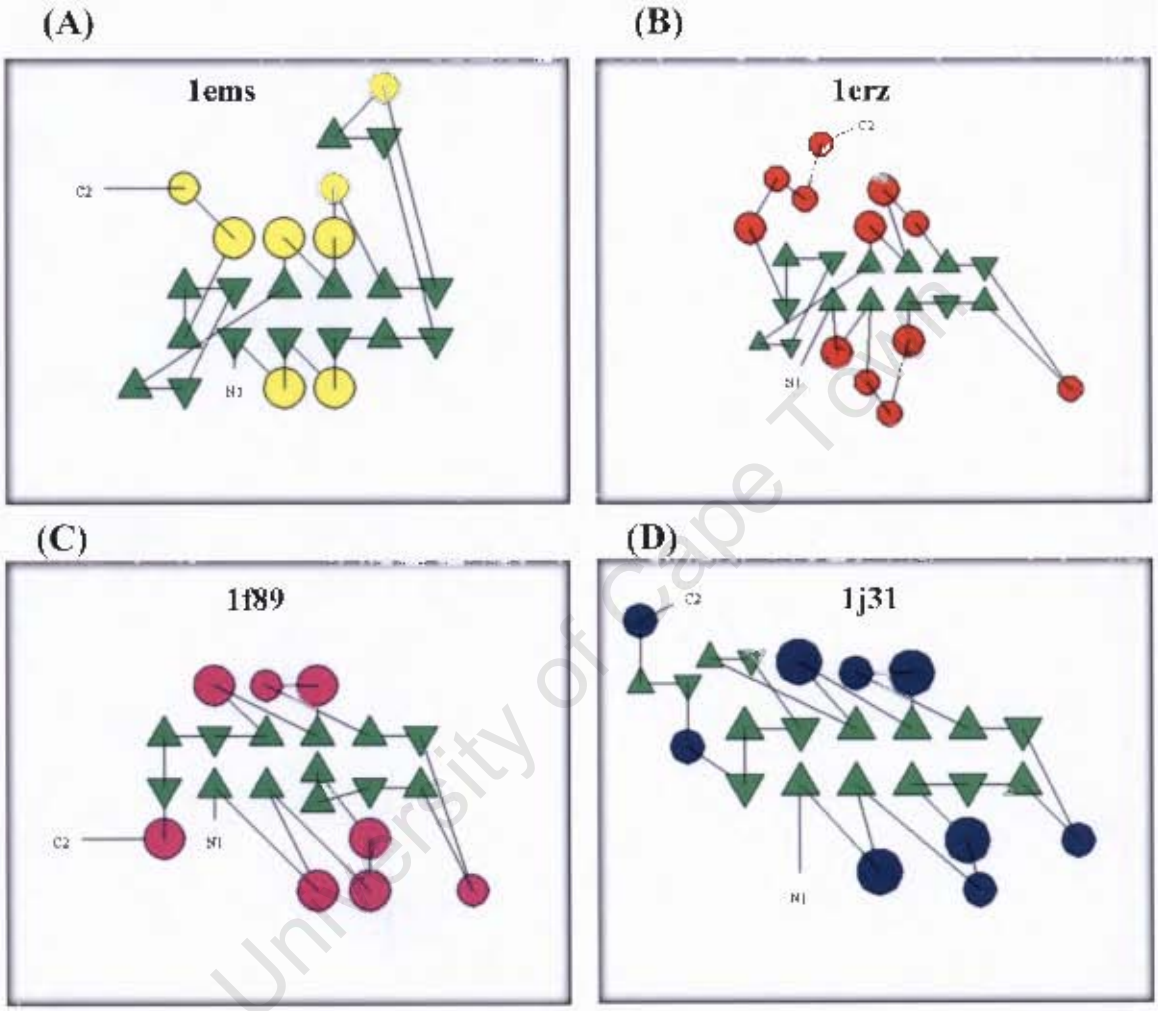


Figure 1-4: Structural folds of the four nitrilase superfamily structures as depicted by their topology diagrams. Helices are shown as circles, while β strands are shown as triangles. (A) is the topology diagram of 1ems, (B) of 1erz, (C) of 1f89 and (D) of 1j31. The four-layer α-β-β-α sandwich fold of the conserved hydrophobic core is clearly evident. The topology diagrams were generated with *TOPS* (Westhead et al., 1997).

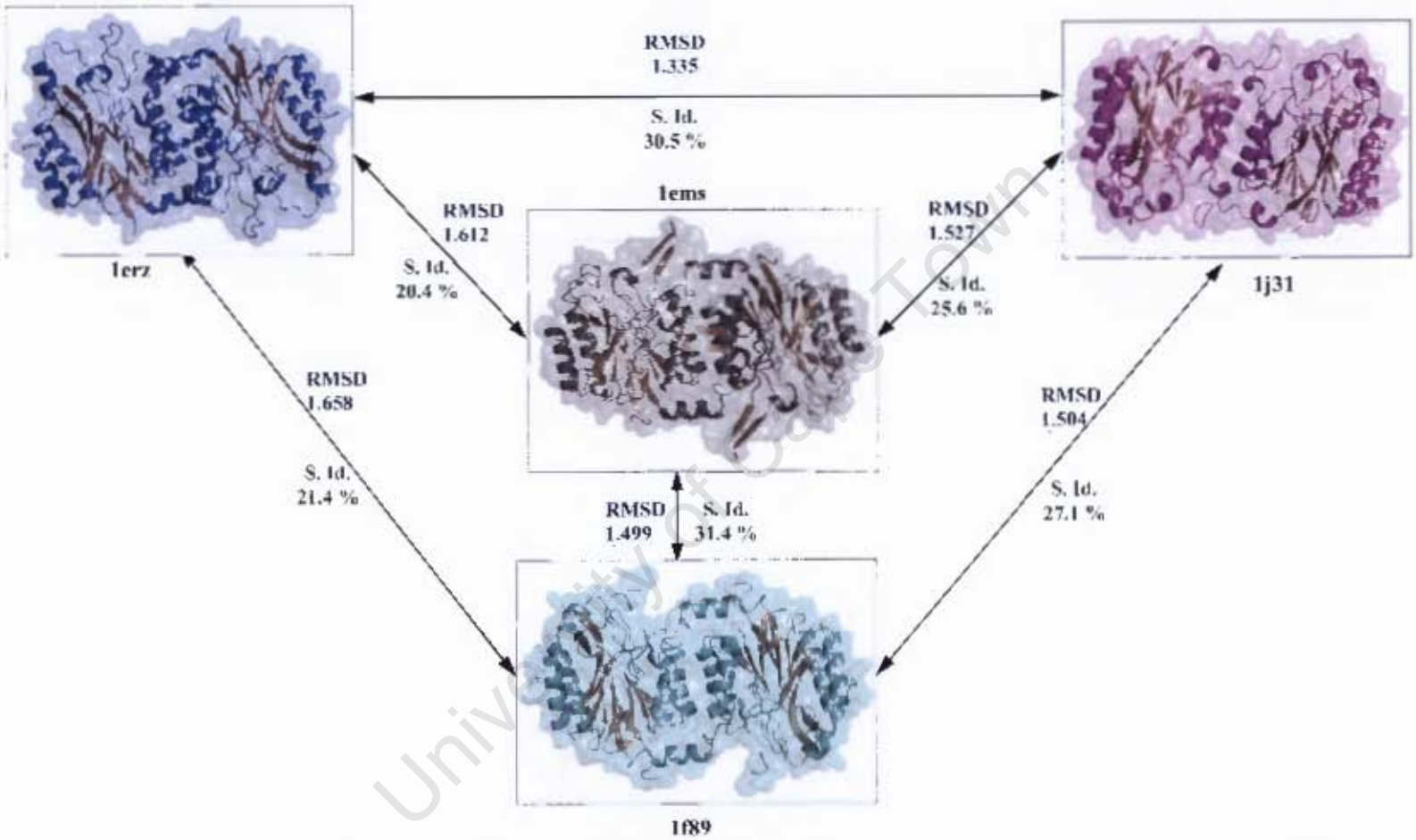


Figure 1-5: Transparent surface and cartoon representation of the structural fold in the existing four nitrilase superfamily structures. A dimeric form of each structure is rendered and they are all viewed down the conserved two-fold dimeric interface. The structures are labelled with their respective PDB ID and the secondary structure elements are shown in different colours to highlight the structural fold. The positional Root Mean Squared Deviation (RMSD) of superimposed C-alpha atoms between each pair of structures is labelled. The Sequence Identity (S. Id.) between each pair is also indicated. The four structures have a conserved four-layer α-β-β-α sandwich fold with the dimer from each structure forming an eight-layered α-β-β-α:α-β-β-α sandwich structure. The RMSD values in Å were determined using *ALIGN* (Cohen, 1997) and the images were rendered with *PyMOL* (Delano, 2004).

Table 1-2: Characteristics of the existing structures in the nitrilase superfamily. All the four proteins have a quaternary structure in solution.

PDB ID of the protein	**Branch in the nitrilase superfamily**	**Function/activity**	**Quaternary structure**
1erz	6	N-carbamyl -D-amino acid + H_2O ⟶ D-amino acid + CO_2 +NH4	Homotetramer
1ems	10	Unknown	Tetramer
1f89	Unknown	Unknown	Homodimer
1j31	Unknown	Unknown	Homodimer

NHases (Huang et al., 1997), N-terminal nucleophile hydrolases and triad amidotransferases (Zalkin, Smith, 1998), and signature amidases (Patricelli, Cravatt, 2000) are not members of the nitrilase superfamily, despite the fact that they all share common substrates.

1.3. The aliphatic amidases

Amidases belong to the hydrolase family, which is a subclass of acrylamide amidohydrolases, which are specific for linear amides (EC 3.5.1) and cyclic amides (EC 3.5.2). As mentioned earlier (Figure 1-1), amidases catalyze the hydrolysis of carboxylic acid amides to their corresponding free carboxylic acids and ammonia. They occur in both prokaryotes and eukaryotes, where in most cases they are coupled with nitrile hydratases (NHases), to drive the hydrolysis of nitriles in a bi-enzymatic pathway. Although this group of enzymes has not been sufficiently investigated, bacterial aliphatic amidases are the most extensively characterised, particularly due to their potential in the large scale production of acrylic as well as other acidic products in industry (Hughes et al., 1998).

Apart from the widely documented amide hydrolysis activity, some amidases are capable of transferring an acyl moiety to hydroxylamine to form hydroxamates (Figure 1-6) (Fournand, Arnaud, 2001). A good example is a study by Thiery and colleagues (1986) that demonstrated the ability of *Rhodococcus* species R312 wide spectrum aliphatic amidase to transfer acyl groups of amides, acids and esters to hydroxylamine, allowing the formation of the corresponding hydroxamic acids and ammonia. This enzyme also catalyzes the transfer of acyl moieties from short-chain amides to hydrazine (NH_2NH_2), leading to formation of acid hydrazines ($RCONHNH_2$). Other reactions that have been observed in a number of characterized amidases include acid transfer, ester hydrolysis and ester transfer (Fournand et al., 1998a; Fournand et al., 1998b; Fournand *et al.*, 1997). These reactions however are much slower, leaving amide hydrolysis and amide transfer reactions as the only industrially interesting amidase reactions (Fournand, Arnaud, 2001). As information on the biochemical characteristics of amidases continues to emerge, it is becoming clearer that amidases exhibit a wide range of substrate specificities; some are specific for aliphatic amides (Asano et al., 1982), others act on aromatic acid amides (Hirrlinger et al., 1996), while still others hydrolyze amides of α- or ω-amino acids (Stelkes-Ritter et al., 1995). Stereo-selectivity is an important aspect that is observed in some amidases (Mayaux et al., 1990; Mayaux et al., 1991; Hashimoto et al., 1991;

Hirrlinger et al., 1996); this can be exploited to allow production of enantiopure acids that would be very difficult to produce by other methods.

1.3.1. Classification of amidases

Although attempts to group amidases have been reported in a number of studies (Chebrou et al., 1996; Pace, Brenner, 2001; Fournand, Arnaud, 2001), their classification is not definitively formulated (Pertsovich et al., 2005). However, based on the amino acid sequence and structural organization, bacterial aliphatic amidases can now be broadly divided into two groups: Group 1 consists of amidases that are structurally related to the nitrilase superfamily enzymes, while group 2 mainly represents amidases belonging to the signature amidase family. These are structurally un-related to nitrilases (Pertsovich et al., 2005; Banerjee et al., 2002). Table 1-3 presents detailed information on some of the characterized bacterial amidases.

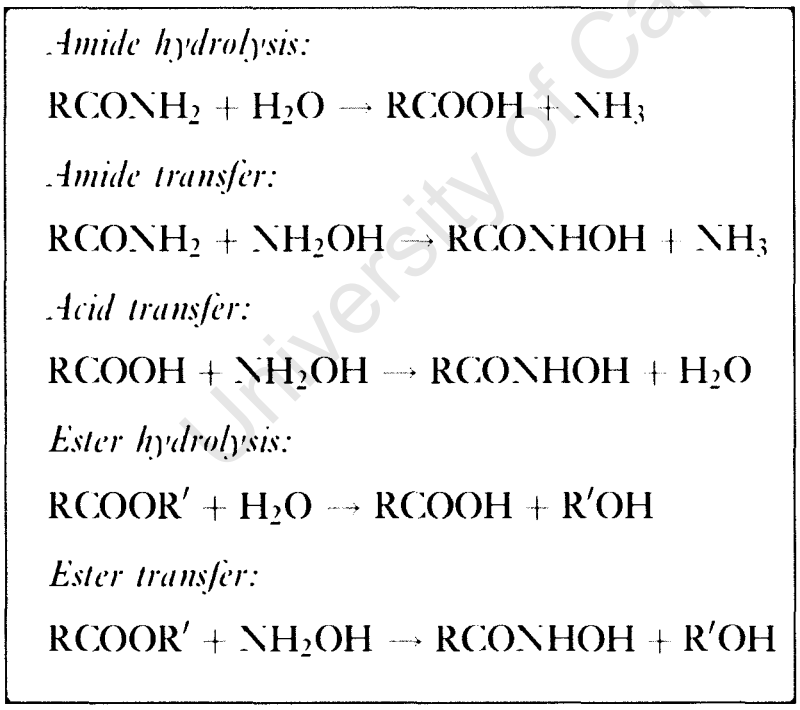


Figure 1-6: Reactions that are catalysed by amidases. Most of the characterized amidases are able exhibit both amide hydrolysis and amide transfer; these are the only reactions that are of interest in industry, as the rest are too slow to be of any economical use. The figure was taken from Fournand and Arnaud (2001).

1.3.1.1. Signature amidase enzymes

Signature amidases are characterized by the presence of an invariant GGSS (Gly-Gly-Ser-Ser) signature motif in their primary sequence (Chebrou et al., 1996; Kobayashi et

al., 1997). They also have aspartate and serine residues (Asp191 and Ser195 in the *Rhodococcus* J1 amidase numbering (Kobayashi et al., 1997) in the active site, in place of the nucleophilic cysteine (Novo et al., 1995) that is observed in the nitrilase superfamily. Amidases from this group exhibit a wide substrate specificity including aliphatic and aromatic amides, as well as amides of α-substituted carboxylic acids. Importantly, most of them exhibit stereoselectivity (Hirrlinger et al., 1996; Kobayashi et al., 1997). Most of the structurally characterized signature amidase enzymes have quaternary structure (Table 1-3), forming homodimeric and homooctameric complexes in solution (Kobayashi et al., 1997; Novo et al., 1995; Chebrou et al., 1996; Mayaux et al., 1991). Representative members of this group include amidases from *R. rhodochrous* J1 (Kobayashi et al., 1997), *Rhodococcus* sp. R312 (Fournand et al., 1998b; Chebrou et al., 1996), *Sulfolobus solfataricus* (Cilia et al., 2005) and *P. chlororaphis* B23 (Ciskainik et al., 1995).

Aside from C-N bond metabolism in prokaryotes, over 200 signature amidase enzymes are also found in archea and eukaryotic organisms forms (Labahn et al., 2002). Curnow and colleagues (1997) reported a signature amidase that transfers NH_3 from glutamine, resulting in the generation of properly charged $tRNA^{Gln}$ in eubacteria. Cravatt and others (1996) also reported a mammalian amidase that was involved in the degradation of neuromodulatory fatty acid amides. This enzyme was found to belong to an "amidase signature (AS) family" of enzymes that is characterized by a conserved stretch (approximately 130 amino acids) that is rich in glycine and serine residues, referred to as the "amidase signature sequence" (Mayaux et al., 1991; Chebrou et al., 1996). Further studies by Patricelli and Cravatt (2000) revealed that this amidase represented a large class of serine-lysine catalytic dyad hydrolases that closely resembled serine hydrolases (Boger et al., 2000).

Several crystal structures of signature amidase enzymes have been reported to-date. Figure 1-7 below presents details of the structural features of a peptide amidase from *Stenotrophomonas maltophilia* (Pam; PDB ID, 1m22); (Labahn et al., 2002). This is a regio-selective signature amidase that is involved in the hydrolysis of the C-terminal amide bond in peptide amides (Labahn et al., 2002). Pam has clear α/β sandwich fold, with a central β-sheet core surrounded by α-helices. This fold is conserved across the members of amidase signature family.

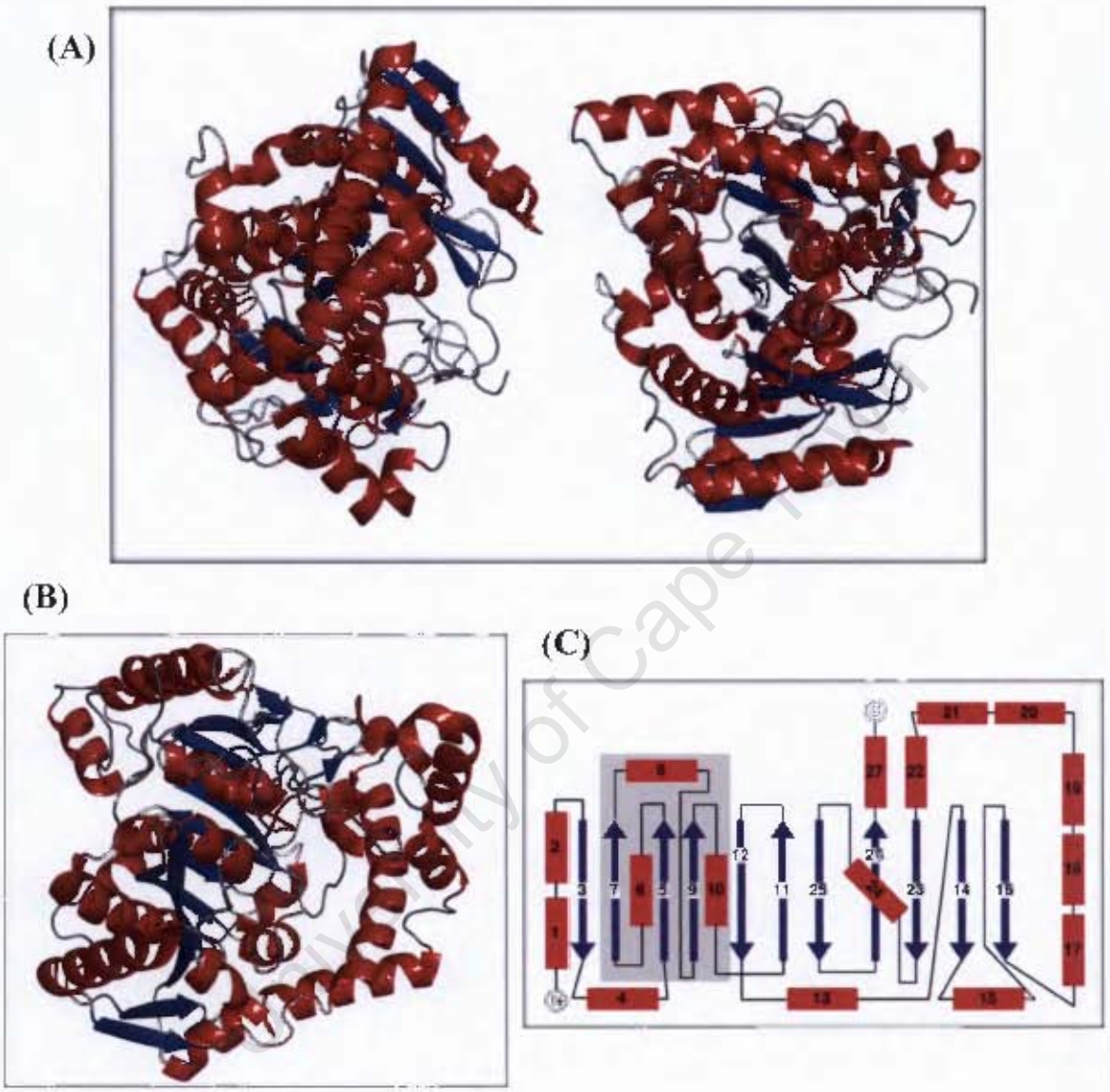


Figure 1.7: Structural features of a peptide signature amidase (Pam) from *Stenotrophomonas maltophilia* (PDB ID, 1m22); (Labahn *et al.*, 2002). Helices are coloured red, β strands blue and loops grey. (A) A cartoon representation of a homodimeric form of Pam. The associated dimers are rotated approximately 180° relative to one another. (B) A monomer of Pam, with the corresponding topology diagram (C). Pam has an α/β fold, with central mixed β-sheet core that is covered on either end by α-helices. The cartoons were rendered in *PyMOL* (Delano, 2004) and the topology diagram was taken from Labahn et al. (2002).

Microorganism	Amidase properties						Reference
	Amidase family	**Substrate specificity**	**Substrate selectivity**	**Number of amino acids**	**Subunit MW (kDa)**	**Quaternary structure**	
Rhodococcus rhodochrous M8	Nitrilase-related	Aliphatic amides - acetamide & propionamide	-	345	42±2	Homotetramer	Kotlova *et al.*, 1999
Brevibacterium sp. R312	Nitrilase-related	Wide spectrum	-	345	43	Homotetramer	Thiery *et al.*, 1986
Rhodococcus sp. Strain NCTR 4	Nitrilase-related	Aliphatic amides – acrylamide, acetamide, butyramide, propionamide, isobutyramide	-	-	44.5	-	Nawaz et al., 1994
Bacillus sp. BR449	Nitrilase-related	Aliphatic amides	-	348	38.6	Homohexamer	Kim & Oriel, 2000
Pseudomonas aeruginosa PAC142	Nitrilase-related	Aliphatic amides - acetamide	-	346	38.4	Homohexamer	Ambler *et at.*, 1987; Karmali *et al.*, 2000
Klebsiella pneumoniae NCTR 1	?	Aliphatic amides – acrylamide, acetamide, butyramide, propionamide	-	-	62	Monomer	Nawaz *et al.*, 1996
Brevibacterium strain R312	Signature amidases	2-aryl & 2-aryloxylpropionamides - 2-(4)-hydroxyphenoxylpropionamides	(R)-Enantiomer-selective	521	54.7	Homodimer	Mayaux *et al.*, 1990
Rhodococcus strain	Signature amidases	2-arylpropionamides – 2-phenylpropionamides	(S)-Enantiomer-selective	461	48.6	Homodimer	Mayaux *et al.*, 1991
Pseudomonas chlororaphis B23	Signature amidases	Aliphatic & aromatic amides – 2-phenylpropionamide	(S)-Enantio-selective for aromatic amides	-	54	Homodimer	Ciskanik *et al.*, 1995

Agrobacterium tumefaciens strain d3	Signature amidases	2-arylpropionamides – 2-phenylpropionamide	(S)-Enantiomer-selective	517	63	Homooctamer	Trott *et al.*, 2001
Rhodococcus erythropolis strain MP50	Signature amidases	Aromatic amides – 2-phenilpropionamides, α-chlorophenylacetamide	(S)-Enantiomer-selective	525	55	Homooctamer	Trott *et al.*, 2002
Rhodococcus erythropolis MP50	Signature amidases	Broad spectrum, including 2-arylpropionamides - 2-phenylpropionamide	(S)-enantiomer-selective	525	61	Homooctamer	Hirrlinger *et al.*, 1996
Ochrobactrum anthropi SV3	?	D-amino acids with aromatic or hydrophobic side chains	D-stereo-selective	363	40	Monomer	Komeda & Asano, 2000
Pseudomonas sp. MCI3434	?	Piperazine-2-tert-butylcarboxamine, -alaninamide, piperazine-2-carboxamide	R-stereo-selective	274	30.1	Monomer	Komeda *et al.*, 2004

Table 1.2: Properties of some of the characterized amidases in the literature. Question marks indicate cases where the classification is not certainly confirmed, while dashes indicate cases where no information on that particular property is available. Most of the nitrilase-related amidases are homotetrameric and homohexameric in solution while the signature amidase family enzymes are mostly homodimeric or homooctameric in solution. The amidase from *Klebsiella pneumonae* NCTR 1 (Nawaz *et al.*, 1996) has all the characteristics of nitrilase-related amidases including a nucleophilic cysteine, except that it has metal ions in the active site and is monomeric in solution. No other characterized amidase has been found to have coordinated metal ions in the active site, which makes this enzyme unique. Amidases from *Ochrobactrum anthropi* SV3 (Komeda & Asano, 2000) and *Pseudomonas* sp. MCI3434 (Komeda *et al.*, 2004) which are monomeric in solution, show stereo-selectivity, but no information is available as to whether an "amidase signature sequence" has been observed. Their substrates are slightly different from the signature amidases, which suggests that these enzymes might belong to a family of their own.

1.3.1.2. Nitrilase-related amidases

A combination of biochemical studies and primary sequence analyses has suggested that amidases from this group are sulphydryl enzymes (Novo et al., 1995; Novo et al., 2002; Farnaud et al., 1999) containing a conserved nucleophilic cysteine (Novo et al., 1995) that is involved in catalysis. Only bacterial amidases have been extensively characterized in this family. These include amidases from *Rhodococcus erythropolis* (Soubrier et al., 1992), *Rhodococcus* sp. R312 (Fournand et al., 1998b; Soubrier et al., 1992; Thiery et al., 1986), *Bacillus stearothermophilis* (Cheong, Oriel, 2000), *Pseudomonas aeruginosa* (Novo et al., 1995), *Bacillus* sp. BR449 (Kim, Oriel, 2000), *Hericobacter pylori* (Skouloubris et al., 1997) and our amidase of interest *Geobacillus pallidus* RAPc8 amidase (Cameron et al., 2005), among others. Unlike signature amidases which are either homodimeric or homooctameric in their active form, most of the characterized nitrilase-related aliphatic amidases form homotetrameric and homohexameric complexes in solution (Pertsovich et al., 2005). These enzymes show substrate preference for short-chain aliphatic amides (Fournand, Arnaud, 2001; Pertsovich et al., 2005). Details of some of the characterized nitrilase-related amidases are given in Table 1-3.

Although no crystal structures of nitrilase-related amidases have been determined to-date, a homology model of *Pseudomonas aeruginosa* amidase (PamiE; PDB ID, 1k17); (Novo et al., 2002), based on the crystal structure of the Nit domain of worm NitFhit fusion protein (Nit; PDB ID, 1ems); (Pace et al., 2000) now exists. The model (Figure 1-8) shows a conserved α-β-β-α sandwich fold that resembles the conserved structural fold of the nitrilase superfamily structures. Analysis of the three-dimensional (3D) PamiE model identified a Glu59, Lys134, Cys166 catalytic triad, that was also supported by mutagenesis studies (Farnaud et al., 1999; Novo et al., 2002). These catalytic residues were found to align with the Glu54, Lys129, Cys169 triad of the Nit crystal structure, after the two molecules were superimposed. Similar catalytic triad residues were also observed in the models of *R. erythropolis*, *H. pylori* and *B. stearothermophilus* amidases that were also based on Nit as the template (Novo et al., 2002). This confirmed the

conservation of the Glu, Lys, Cys catalytic triad across the nitrilase superfamily members and also supports the classification of these amidases in the nitrilase superfamily.

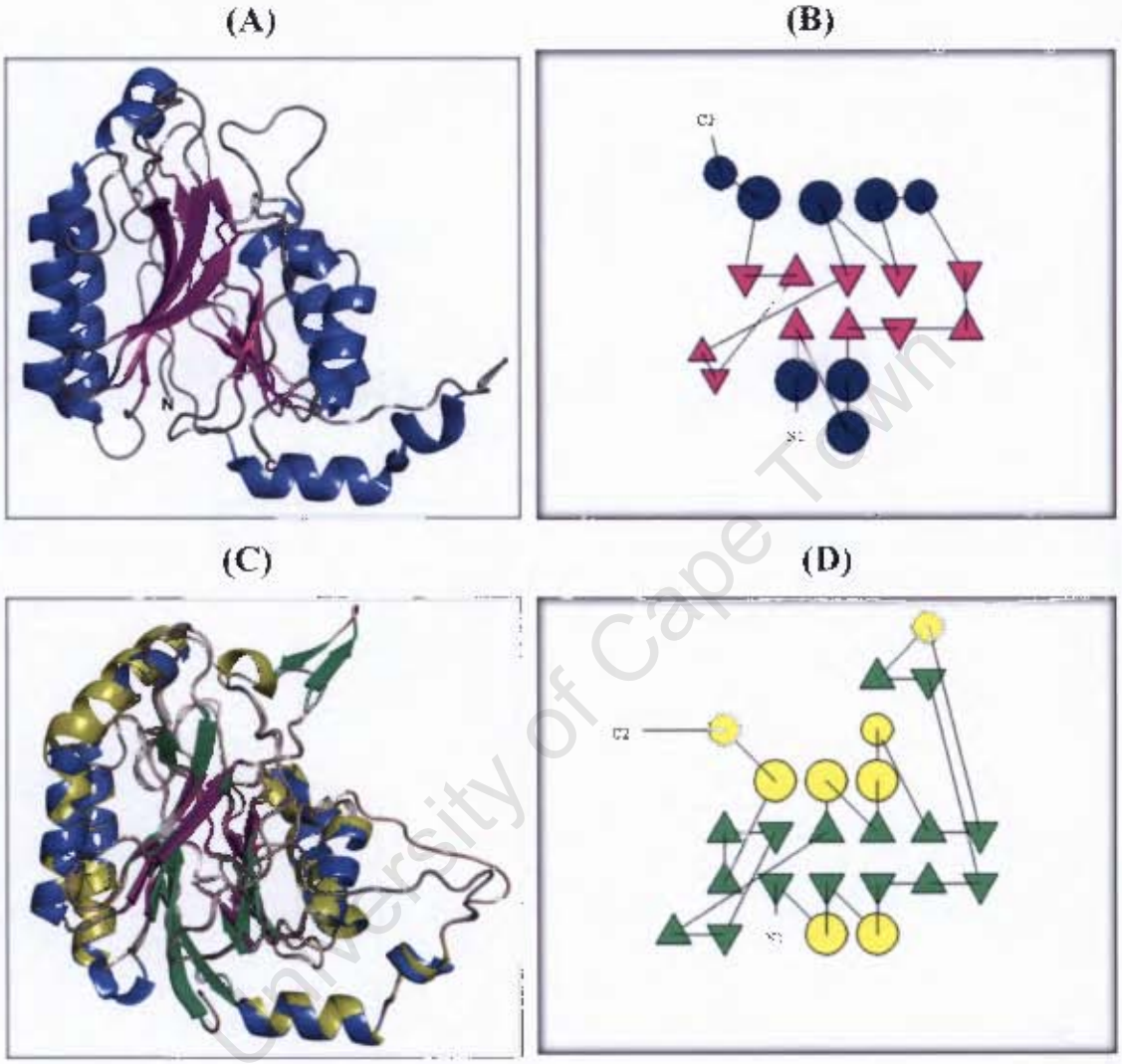


Figure 1-8: Modelled structural fold of an amidase from *Pseudomonas aeruginosa* (PamiE; PDB ID, 1k17): (A) is a cartoon diagram of PamiE (Novo et al., 2002), and (B) is a secondary structure topology diagram of the same protein (PamiE). Helices are shown in slate/blue, β-sheets are shown in magenta, and loops in grey. (C) is a cartoon representation of superposed structures of PamiE and Nit domain of NitFhit fusion protein (Nit; PDB ID, 1ems); (Pace et al.,2000) and (D) is a topology diagram of Nit protein. Helices are coloured yellow in Nit, β-sheets green and the loops wheat. The two proteins have a conserved fold (α-β-β-α sandwich architecture) of the core region, confirming their homology, and their belonging to the same nitrilase superfamily, despite their functional differences. The structures were superimposed with *ALIGN*

(Cohen, 1997), the cartoons were rendered with *PyMOL* (Delano, 2004) and the topology diagrams were generated using *TOPS* (Westhead et al , 1997).

Comparison of this modelled structure with that of the signature amidase Pam reveals that although amidases catalyze similar reactions, the two groups are very different structurally (Figure 1-9).

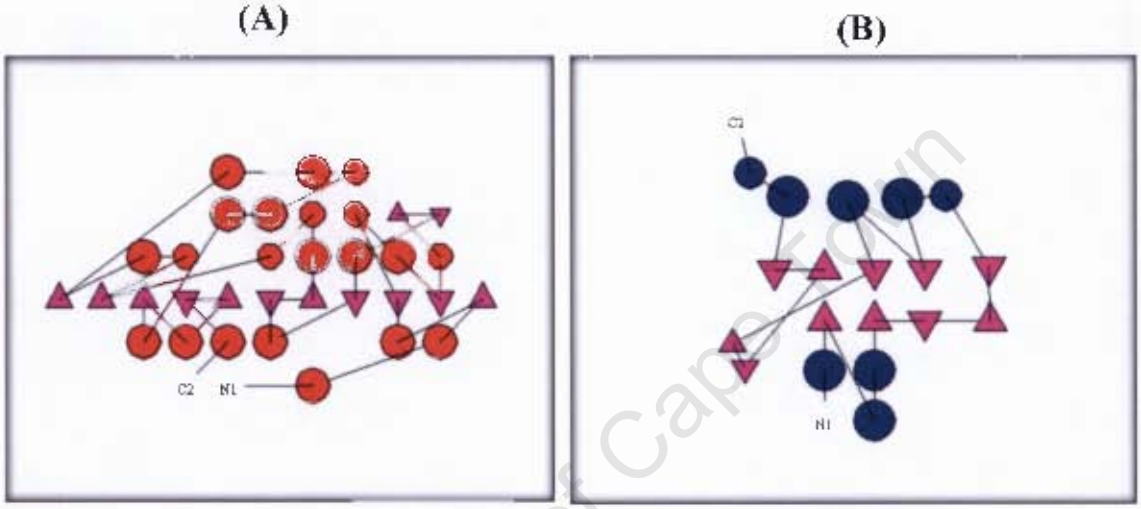


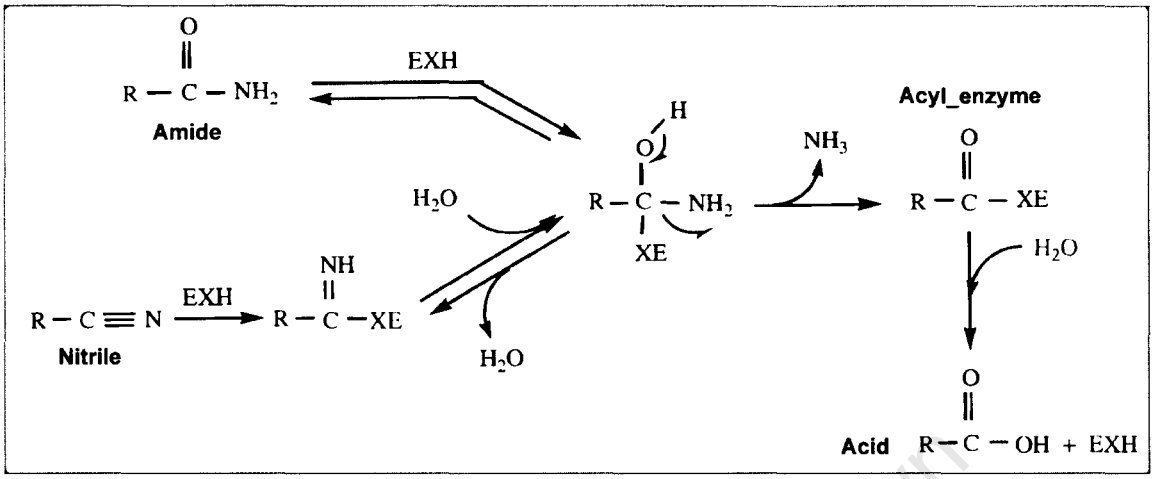
Figure 1-9: A side-by-side comparison of the structural fold differences between signature amidases and nitrilase superfamily aliphatic amidases: (A), a topology diagram of a **signature amidase** from *Stenotrophomonas maltophilia* (Pam; PDB ID, 1m22); (Labahn et al., 2002) and (B), a topology diagram of a theoretical model of a **nitrilase-related** aliphatic amidase from *Pseudomonas aeruginosa* (PamE; PDB ID, 1k17); (Novel et al., 2002). Although the two have similar functions, they exhibit completely different structural folds. The topology diagrams were generated with *TOPS* (Westhead et al., 1997).

Some organisms have been found to express a wide spectrum of amidases. An example is *Rhodococcus* sp. R312 bacteria that has been reported to express at least four types of amidases; an aliphatic amidase (Novo et al., 1995), an enantioselective amidase that hydrolyzes aryloxy propionamides (Mayaux et al., 1990), a novel amidase hydrolyzing dinitriles (Moreau et al., 1993) and an α-amino acid amidase that is specific for L-α-amino amides (Banerjee et al., 2002; Fournand, Arnaud, 2001). It is also worth noting that, although bacterial amidases function in the same nitrogen metabolism pathway as NHases, the existence of metal ions (like iron and cobalt) in the active site of amidases is

very rare. This has only been reported in an amidase from *Klebsiella pneumoniae* by Nawaz and colleagues (1996).

1.3.2. Mechanism of catalysis by amidases

In 1998, Kobayashi and colleagues (Kobayashi et al., 1998) reported the interesting finding that an amidase from *Rhodococcus rhodochrous* J1 also catalyzes the hydrolysis of nitriles to corresponding acids and ammonia. They then proposed an amidase hydrolytic mechanism that included nitrile hydrolysis, as shown in Figure 1-10 (A) below. In the proposed mechanism, the carbonyl group of the amide undergoes nucleophilic attack, resulting in the formation a tetrahedral intermediate. With the removal of ammonia, the tetrahedral intermediate is converted to an acyl-enzyme complex, which is then hydrolysed to an acid upon the addition of a water molecule. This mechanism had previously been suggested by Maestracci and others (1986), who proposed a similar mechanism (Figure 1-10 (B)) for acyl transfer activity that they had observed in the aliphatic amidase from *Rhodococcus* sp. R312.



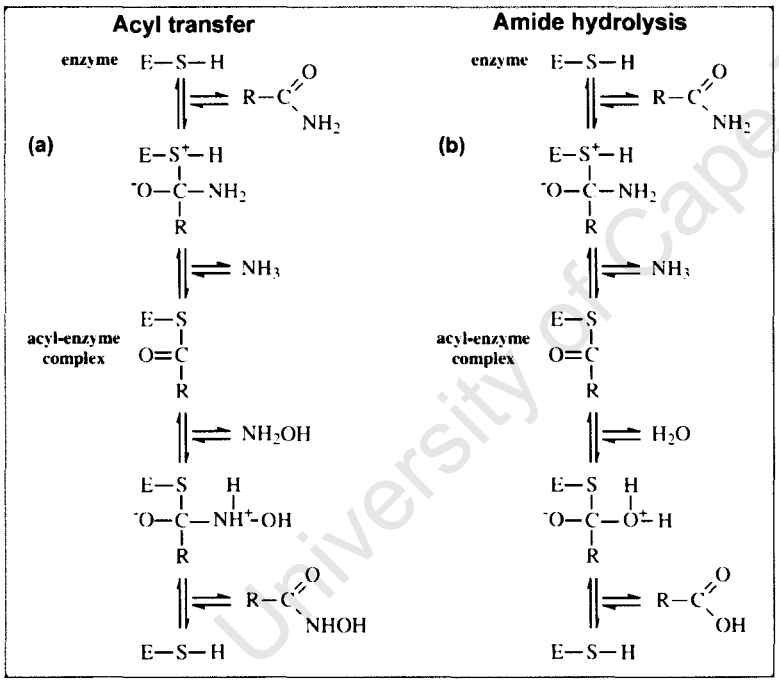


Figure 1-10: Catalytic mechanism of amidases: (A), amide and nitrile hydrolysis mechanism proposed by Kobayashi and others (1998). (B), amide hydrolysis and acyl transfer mechanisms proposed by Maestracci and colleagues (1986); the two mechanisms are proposed to proceed via the formation of an acyl-enzyme intermediate, followed by the transfer of the acyl group to either water (hydrolysis) or hydroxylamine (acyl transfer) co-substrates. Figure (A) is taken from Banerjee et al. (2002), while Figure (B) is taken from Fournand and Arnaud (2001).

1.3.3. *Geobacillus pallidus* RAPc8 amidase

Pereira and colleagues (1998) identified a thermostable amidase in a *Bacillus* isolate (sp. RAPc8) that is currently named *Geobacillus pallidus* (RAPc8), based on its 16S RNA sequence. An Open Reading Frame (ORF) encoding a putative 348-amino acid amidase (MW 38.6 kDa) was located 127 bp upstream (Figure 1-11) of ORFs encoding the β and α (in that order) subunits of a cobalt-containing nitrile hydratase (NHase). This gene organization was also observed in *Bacillus* sp. BR449 (Kim, Oriel, 2000), which also encodes a cobalt-containing NHase and a nitrilase-related aliphatic amidase. This differs from the gene organization observed in organisms expressing a NHase-signature amidase couple (Figure 1-11), the main difference being the order of the genes for α and β subunits of NHases. The amidase from *Geobacillus pallidus* RAPc8 (RAPc8 amidase) gene was cloned and over-expressed by Cameron and others (2005), who identified 100% sequence identity between this amidase and an aliphatic amidase from *Bacillus* sp. BR449 (Kim, Oriel, 2000). The DNA sequences however shared 99% identity due to six silent nucleotide substitutions (Cameron et al., 2005).

Although the physiological role of RAPc8 amidase is still not clear, there is a high likelihood that it might be involved in the metabolism of aldoximes (Figure 1-1), forming the third stage in the three-step pathway that involves aldoxime dehydratase, NHase and amidase. This speculation is based on the report by Kato and colleagues (1999), who showed a link between the three enzymes (aldoxime dehydratase, NHase and amidase) and aldoxime metabolism in *Rhodococcus* sp. YH3-3.

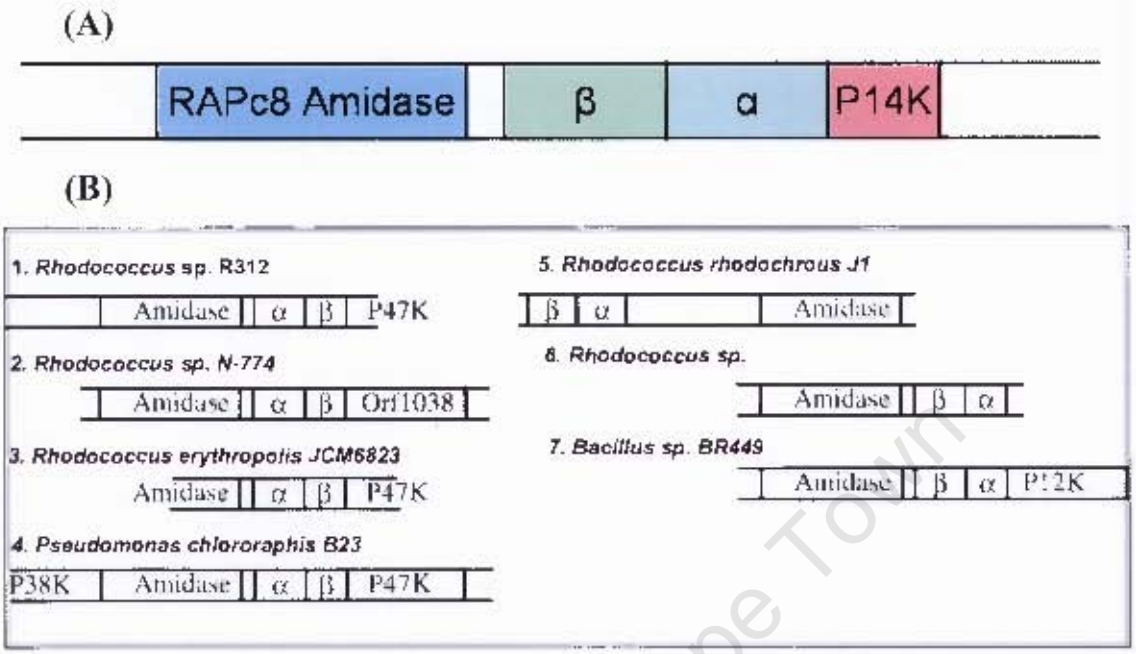


Figure 1-11: Structural organization of genes encoding different bacterial NHases and amidases: (A), shows the gene organization of *Geobacillus pallidus* RAPc8 while (B) shows NHase and amidase genes from other organisms. β and α are genes coding for NHase subunits. The sequence after the NHase gene (P14K in RAPc8 amidase), encodes a small protein that was found to be necessary for proper folding and activity of NHase (Cameroon et al., 2005). *G. pallidus* gene structure resembles that of *Bacillus* sp. BR449 (Kim, Oriel, 2000). Figure 1-11(B) was taken from Fournand and Arnaud (2001).

In a preliminary analysis done in our group (results not shown), a Blastp (Altschul et al., 1997) search using the primary structure of RAPc8 amidase identified 10 sequences having an identity level of 100-79%, with the first 100 sequences having identities between 100 and 32%. The sequences that were highly identical to RAPc8 amidase belonged to nitrilase-related aliphatic amidases from various organisms including *Bacillus* sp. BR449 (100% identical), *Geobacillus stearothermophilus* (89%) and *Pseudomonas aeruginosa* (PamiE) (81%). Sequence alignments showed a number of highly conserved regions, including the nitrilase superfamily catalytic triad residues, Glu, Lys and Cys, and flanking sequences. Glu59, Lys134 and Cys166 of *Pseudomonas aeruginosa* amidase (PamiE) aligned with residues at similar positions (Glu59, Lys134

and Cys166) in RAPc8 amidase. The similarity between the primary structure of the nitrilase superfamily amidases and RAPc8 amidase, as well as the existence of the characteristic Glu, Lys, Cys catalytic triad in RAPc8 amidase, was a clear indication that RAPc8 amidase was a member of the nitrilase superfamily.

1.3.3.1. Substrate specificity in RAPc8 amidase

Makhongela and colleagues (in press) have recently characterized RAPc8 amidase in terms of activity, substrate specificity and the effect of immobilization on different scaffolds. This is a first study of the kind on this enzyme. They have found that RAPc8 amidase is a highly thermostable enzyme, with activity between 50 and 60 °C and maximum specific activity at pH 7.0. It exhibits both amide hydrolysis and acyl transfer activities (Makhongela et al., in press), as observed with other characterised amidases.

Similar to the substrate specificities observed in other nitrilase-related amidases, RAPc8 actively hydrolyzes (Table 1-4) short-chain aliphatic amides such as acrylamide, propionamide and acetamide. It is however only moderately active on substituted short-chain and mid-length aliphatic amides (such as diacetamide, lactamide, fluoroacetamide and isobutyramide) and has no activity at all on long-chain aliphatic amides (hexanoamide), aromatic substrates or urea amides (Makhongela et al., in press). This enzyme has substantial level of chiral selectivity on D-enantiomers, particularly on D-lactamide and to some extent D-alaninamide. No activity was observed on L-lactamide.

A similar trend of substrate preference was observed with the acyl transfer activity when various amide substrates were used as acyl donors in the presence of hydroxylamine reagents (Makhongela et al., in press). Similar to hydrolysis reactions, RAPc8 amidase readily transfers acyl moieties from low molecular weight aliphatic amides (Table 1-5) to hydroxylamine to form hydroxamic acids. The amidase is not active on long-chain aliphatic amides (such as hexanamide), aromatic amides (benzamide) or L-isomers of either short- or mid-length aliphatic amides (including L-alaninamide and L-leucinamide). Among the amide substrates tested, isobutyramide showed highest relative

activity for acyl transfer (Table 1-5), which probably points to a shift in chain length specificity from hydrolysis to acyl transfer (Makhongela et al., in press).

RAPc8 amidase is completely inhibited by oxidizing agents such as hydrogen peroxide (H_2O_2), while the activity is reduced several fold by the presence of modifying metal ions like Fe^{3+}, Co^{2+}, Cu^{2+}, Zn^{2+} (Makhongela et al., in press). On the other hand, the presence of reducing agents like dithiothreitol (DTT) at millimolar (mM) concentrations not only restores the activity of the inactivated enzyme, but also stabilizes the acyl transfer activity against oxidation; this explains why the enzyme activity was lost in the absence of DTT during purification. All these observations are clearly consistent with the presence of a catalytically active cysteine thiol group. The activity of RAPc8 amidase is however not affected by the presence of either serine protease inhibitor PMSF or metal-chelating agent EDTA. This is a strong indication that neither an active site serine residue nor coordinated ions are present in the active site (Makhongela et al., in press).

Table 1-4: Hydrolytic activity of RAPc8 amidase on a number of tested amide substrates. High activity is noted on the short-chain aliphatic amides, while there is no activity at all with long-chain and aromatic amides. The table was borrowed from Makhongela et al. (in press).

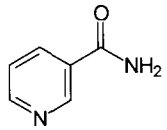
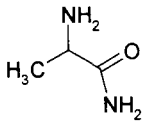
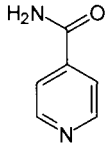
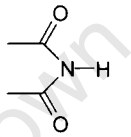
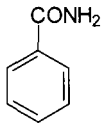
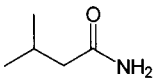
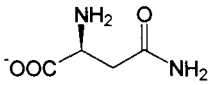
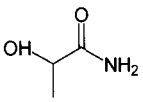
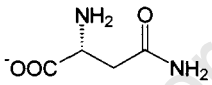
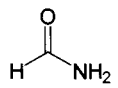
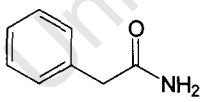
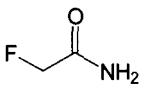
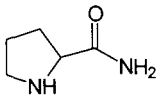
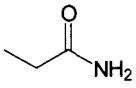
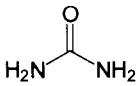
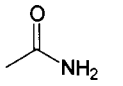
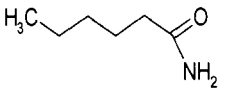
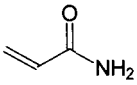
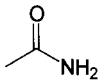
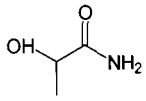
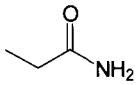
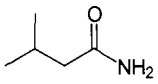
SUBSTRATE /STRUCTURE		RELATIVE ACTIVITY (%)	SUBSTRATE /STRUCTURE		RELATIVE ACTIVITY (%)
Nicotinamide		**0**	L-Alaninamide		**3**
Isonicotinamide		**0**	Diacetamide		**14**
Benzamide		**0**	Isobutyramide		**16**
L-Asparagine		**0**	Lactamide		**17**
D-Asparagine		**0**	Formamide		**30**
DL-Phenylalanine		**0**	Flouroacetamide		**165**
L-Prolinamide		**0**	Propionamide		**67**
Urea		**1.5**	Acetamide		**100**
Hexanamide		**0**	Acrylamide		**102**

Table 1.4: Acyl transfer activity of RAPc8 amidase. Isobutyramide has the highest acyl transfer activity, despite its low level of amide hydrolysis. The table was taken from Makhongela et al. (in press).

SUBSTRATE/STRUCTURE		RELATIVE ACTIVITY (%)
Acetamide		**100**
Lactamide		**85**
Propionamide		**95**
Isobutyramide		**350**

1.4. Motivation and study objectives

Motivation

Although the physiological role of *Geobacillus pallidus* sp. RAPc8 amidase (RAPc8 amidase) is not clear, its recent biochemical characterization (Makhongela et al., in press) has revealed the specificity of this enzyme for short-chain aliphatic amides, both in hydrolysis and acyl transfer. RAPc8 amidase also shows stereoselectivity for a number of tested enantiomeric substrates. However the molecular and structural basis of the observed substrate specificity and selectivity is still not understood.

It is now known that members of the nitrilase superfamily form oligomeric complexes in solution. The four enzymes whose crystal structures have been determined in the family (PDB IDs; 1erz, 1ems, 1f89 & 1j31: Nakai et al., 2000, Pace et al., 2000, Kumaran et al., 2003 & Sakai et al., 2004 respectively) have quaternary structure; 1erz and 1ems are tetrameric while 1f89 and 1j31 are dimeric in their active forms. The characterized amidases have an even number of subunits, ranging from 2 to 8 (Kotlova et al., 1999), while the nitrilase branch of the superfamily are reported to form complexes of 1 to 26 subunits (O'Reilly, Turner, 2003; Banerjee et al., 2002). The quaternary structure of RAPc8 amidase and its implications (if any) on the activity is still not known.

Members of the nitrilase superfamily of enzymes share a conserved Glu, Lys, Cys catalytic triad (Brenner, 2002), yet they show significant differences in the reaction (amidase, nitrilase, carbamylase, acyl transferase reactions among others) that they catalyze, in substrate specificity and in catalytic activity. Understanding the features responsible for these differences will require detailed structural comparison between enzymes of different branches. Although four crystal structures of nitrilase superfamily enzymes have been determined (a carbamylase (DCase) and 3 with unknown function), there is none from the amidase branch.

Finally, sequence comparisons between members from different branches in the nitrilase superfamily have revealed that RAPc8 amidase has an elongated C-terminal tail that is also present in some members of the nitrilase branch, but absent in most of the members in the superfamily. The role (either functional and/or structural) of this extended C-terminal tail in these enzymes is not known.

Project aims

The overall objective of the study was to determine the crystal structure of *Geobacillus pallidus* sp. RAPc8 amidase (RAPc8 amidase). The specific objectives were:

1. To determine the biological unit of the active RAPc8 amidase enzyme in solution using gel filtration and electron microscopy,
2. To collect X-ray diffraction data from RAPc8 amidase crystals, and to use crystallographic methods (specifically molecular replacement) to determine crystal packing as well as to solve crystal structure of RAPc8 amidase,
3. To determine the role of the elongated C-terminal tail in RAPc8 amidase and
4. To examine the active site of RAPc8 amidase in order to determine the structural basis of substrate specificity and selectivity and possibly the catalytic mechanism.

Significance and impact

In the past two decades, biotechnological industries have been exploring the enzymatic hydrolysis of nitriles and amides as a method of obtaining a broad spectrum of useful carboxylic acids. As mentioned elsewhere, this has been driven by attempts to find alternatives to the traditional acid- or base-catalyzed amide and nitrile hydrolytic methods, that involve harsh conditions of temperature and pH, and hence are incompatible with the labile structures of many industrially relevant compounds, introduce unwanted by-products, decrease product yield and increase the overall production costs (Fournand, Arnaud, 2001; Banerjee et al., 2002; Mylerova, Martinkova, 2003). The potential of nitrile-degrading enzymes in catalyzing useful enantio-, chemo-

and regio-selective biotransformations has been discovered. The use of these enzymes in chemical-free bioremediation of the environment has also been explored (Muller, Gabriel, 1999; Banerjee et al., 2002; Wyatt, Knowles, 2003). Ongoing biochemical characterization studies (Makhongela et al., in press) have revealed the capability of RAPc8 amidase to drive both hydrolytic and acyl transfer activities with a significant level of enantio-selectivity. The stereoselectivity of RAPc8 amidase will allow production of enantiopure acids that are difficult to produce by traditional chemical methods. Immobilization of RAPc8 amidase on Eupergit C beads (Makhongela et al., in press) has been found to enhance its activity, increasing the production yield several fold. This enzyme thus has great potential in synthetic industries.

Structural characterization of RAPc8 amidase will shed light on the basis of substrate specificity and selectivity as well as the catalytic mechanism not only in this enzyme but also in the nitrilase-related amidases. The long term focus will be to extend the utility of this enzyme both in industries and probably in waste degradation, to fine-tune its activity so that it is more selective, or even to alter its substrate specificity. Structural information is needed to inform such experiments.

The research will in the long run have an impact in the biotechnological industries in South Africa, and across the continent. The outcome of this will allow the generation of highly efficient enzymes, whose use will ensure safe and effective production protocols, high quality pure products and increased yields at lower production costs. In cases where nitrile- or amide-containing industrial wastes present a challenge for safe disposal, there is the potential to construct waste treatment systems using microorganisms that contain amide-degrading enzymes, ensuring safe disposal of these wastes. This will go a long way in preventing environmental degradation resulting from the accumulation of these wastes in the environment.

CHAPTER 2.: MATERIALS AND METHODS

2.1. Biological unit molecular weight determination

The molecular weight of the amidase from *Geobacillus pallidus* RAPc8 amidase (RAPc8 amidase) biological unit was determined by gel filtration and electron microscopy (EM).

2.1.1. Gel filtration

Gel filtration was performed according to the published protocol of Makhongela et al. (in press). In brief, an S-300 HR gel exclusion chromatography column (GE-Healthcare) was pre-equilibrated with a buffer containing 150 mM NaCl and 50 mM potassium phosphate buffer (pH 7.0). The protein sample was applied to the column and eluted with 20 mM Tris, 150 mM NaCl, 1 mM DTT (pH 7.4) over a total volume of 120 ml. Fractions were collected at 3 min intervals with a Gilson FC 203B fraction collector at a flow rate of 0.5 ml/min. RAPc8 amidase homogeneity and sub-unit molecular mass were determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and non-denaturing gel electrophoresis (native-PAGE) using 15 % gels stained with Coomassie blue R-250. Samples were assayed for protein concentration using a Bradford protein determination kit (BioRad) with bovine serum albumin (BSA) as the standard.

2.1.2. Electron microscopy and image analysis

Negatively-stained electron micrograph preparation and EM image processing and analysis were performed using the protocols that are published in Makhongela et al. (in press). In summary, purified amidase (3 µl) at a concentration of 0.05 mg / ml was applied to carbon films which had been glow-discharged for 30s, rinsed on two droplets of water, blotted and negatively stained with 2% uranyl acetate. Images were recorded at minimum dose (approx 100 el/A^2) using a Zeiss 912 electron microscope set for zero energy loss imaging at 120kV and 50,000x magnification on a Proscan 2k x 2k CCD camera. The magnification on the CCD had previously been determined to be 2.28Å per pixel under these conditions.

Projection images (12,698) of the putative RAPc8 amidase were manually selected using *Ximdisp* graphical program (Crowther et al., 1996). The selected particles were then boxed, filtered and normalized using *SPIDER* (Frank et al., 1996). The normalized projection images were rotationally and translationally aligned using the reference-free alignment method after which they were subjected to K-means Multivariate Statistical Analysis (MSA) to put them in classes of similar views, using *SPIDER*. The images in each class were then averaged to generate class averages with improved Signal-to-Noise (S/N) ratio. The aligned images were also subjected to Principal Component Analysis (PCA) using methods described by Frank (2006). This was done to determine the symmetry of the elements that comprised the images.

2.2. Secondary structure prediction

Protein Structure Prediction Server (*PSIPRED*) (McGuffin et al., 2000) was used to predict the secondary structure profile for RAPc8 amidase primary sequence.

2.3. Protein purification and crystallization

RAPc8 amidase protein was purified and crystallized using materials and methods that we have published (Agarkar et al., 2006). After expression in *E. coli* BL21 (DE3) cells and cell lysis, the protein was heated to 45 °C for 20 min in a buffer containing 10 mM DTT, 1 mM EDTA, 200 mM NaCl and 10% glycerol. The precipitants were removed by centrifugation. The supernatant was applied to a HighLoad 16/10 Phenyl-Sepharose column (Amersham Biosciences) pre-equilibrated in a buffer containing 1.7 M ammonium sulphate, 50 mM potassium phosphate, 5 mM DTT, pH 7.4. The bound protein was eluted with a linear gradient of ammonium sulphate (1.7-0 M). This was followed by another chromatographic step using a Hiprep 16/10 Q-sepharose FF column (Amersham Biosciences) pre-equilibrated with a potassium phosphate buffer (20 mM) that also contained 2 mM DTT and 1 M NaCl at pH 7.4. The column was eluted with a linear gradient of NaCl (0.1-1 M). The fractions containing pure protein were pooled, and concentrated, then taken through a gel filtration chromatography step as described earlier.

The enzyme was crystallized using the hanging-drop vapor diffusion method, using 4 µl of protein and 4 µl reservoir solution. After a number of crystallization conditions screens and optimization steps, diffraction quality crystals were grown at 22 °C in the presence of 1.2 M sodium citrate, 400 mM NaCl and 100 mM sodium acetate at pH 5.6.

2.4. X-ray diffraction data collection

Crystals were visually inspected for quality, and their dimensions were measured using an in-lens graticule on a stereomicroscope (Leica Microsystems, Wetzlar, Germany). Images were acquired on a Leica Z16 APO (KLI500 LCD) stereomicroscope connected to a computer, using the *IM500* program (Leica Microsystems, Wetzlar, Germany).

The X-ray data were collected on the in-house X-ray diffraction machine at the University of Western Cape, South Africa. The X-ray diffraction machine consists of a Rigaku RUH3R copper rotating-anode X-ray source operated at 40 kV, 22 mA; a Rigaku R-axis IV+ image plate camera (Rigaku MSC, Tokyo, Japan); an X-stream 2000 low-temperature system (Rigaku MSC, Houston TX, USA); and an AXCO PX50 glass capillary optic with a 0.1 mm focus (Australian X-Ray Capillary Optics, Parkville VIC, Australia). Data from crystals mounted on a cryoloop (Hampton Research) were collected at a temperature of 100 K with a crystal-to-detector distance of 157.1 mm.

Images data frames covering an oscillation angle of 0.5° per frame and spanning a range of 48° (between 25° and 73°) were collected for 55 minutes per frame, to make a total of 96 images.

2.5. Data processing and characterization

The indexing, integration, scaling and merging of the raw X-ray diffraction data were performed using *Denzo* and *SCALEPACK* (Otwinowski, Minor, 1997). Determination

of the Laue group and subsequently the space group was were attempted using various strategies as implemented with the crystallographic programs, *Denzo* and *SCALEPACK*, *POINTLESS* (Collaborative Computational Project Number 4 (CCP4)., 1994; Evans, 2006) and *PHASER* (CCP4, 1994; (Read, 2001). The measured data were characterized in terms of asymmetric unit contents by calculating a Matthews coefficient (Vm) (Matthews, 1968) using *PHASER*. The program *POLARRFN* (CCP4, 1994) was used to calculate a fast self-rotation function in order to discern the presence of a local rotation axis and its orientation in the asymmetric unit.

2.6. Checking for the quality of diffraction data

In addition to data quality statistics obtained from the merging and scaling step in *SCALEPACK* (Otwinowski, Minor, 1997), the *Xtriage* program of the *PHENIX* (Python-Based Hierarchical Environment for Integrated Xtallography) suite (Adams et al., 2002) was used to check for crystal twinning, translational pseudo-symmetry, possible outlier reflections (Read, 1999) and ice rings-related problems in the merged RAPc8 amidase data.

2.7. Molecular replacement

2.7.1. Search for homologues

A search for distant structural homologues of RAPc8 amidase was performed using the structure-sequence threading tool, *GenTHREADER* (Jones, 1999). Structural alignments between the identified homologues and RAPc8 amidase sequence were performed manually with the guidance of the *GenTHREADER* output, the predicted RAPc8 amidase secondary structure profile and the pair-wise superposition of the structures using *ALIGN* (Cohen, 1997).

2.7.2. Preparation of the search models for molecular replacement

The coordinate files of the identified homologues were edited manually before molecular replacement (MR) trials. Parts of the models that had a high likelihood of being different from RAPc8 amidase structure were removed, including non-protein atoms such as water molecules and metal ions, regions containing deletions and/or insertions, and extra domains that were not homologous to RAPc8 amidase. Extra subunits were also removed from the models leaving only regions of the model that were consistent with the molecular content of RAPc8 amidase asymmetric unit. A series of search models were prepared for use in different MR trials; these included models lacking the variable loop regions, mutated models with RAPc8 amidase side chains, and mixed poly-alanine models with alanines replacing non-identical residues. Mutation of the models' side chains was performed using both *SCWRL* (Canutescu et al., 2003) and *CHAINSAW* (Schwarzenbacher et al., 2004), and was guided by structural alignments between RAPc8 amidase and the homologues.

2.7.3. Molecular replacement rotational and translational searches

Since RAPc8 amidase MR searches were not trivial, several strategies as discussed in the Results and Discussion section were attempted. Various MR programs were used, including *MOLREP* (CCP4, 1994; Vagin, Teplyakov, 1997), *AMoRE* (CCP4, 1994); Navaza, 1994), *EPMR* (Kissinger et al., 1999), *REPLACE* suite (*GLRF* and *TF*) (Tong, 1993) and *PHASER* (CCP4, 1994; Read, 2001). Confirmation of molecular replacement (MR) solution and rigid body refinement were performed using *PHASER*.

2.8. Model rebuilding and refinement

Rigid body refinement was performed with *PHASER* (CCP4, 1994; Read, 2001), and initial sigma-A weighted electron density maps were generated using *REFMAC5* (CCP4, 1994; Murshudov et al., 1997). Since model fitting, rebuilding and refinement process was not straight-forward, several strategies were employed as discussed in the Results and Discussion chapter. These included (among other strategies) solvent

flattening of the initial electron density map using *RESOLVE* (Terwilliger, 2004a). The initial model (substituted with RAPc8 amidase side chains) was rebuilt in *PHENIX* (Adams et al., 2002). This was followed by alternating iterative cycles of manual model building, fitting and rebuilding using the crystallographic graphical program *O* (Jones et al., 1990) and restrained refinement using *REFMAC5*.

2.9. Final model validation

The accuracy, precision and correctness of the refined RAPc8 amidase model was assessed using various validation tools, including *PROCHECK* (CCP4, 1994; Laskowski et al., 1993), *WHATCHECK* (Hooft et al., 1996) and *MOLPROBITY* (Lovell et al., 2003).

2.10. RAPc8 amidase structure analysis

The online *CASTp* (Computed Atlas of Surface Topography of proteins) server (Binkowski et al., 2003) was used to analyze the pockets and cavities on the surface or interior of the complete RAPc8 amidase model. *BAVERAGE* program (CCP4, 1994) was used to calculate average B-factors for the final model, both for the main chain and the side chains. The topology diagrams were generated with *TOPS* (Westhead et. al., 1999) and modified using *TOPDRAW* (Bond, 2003).

Unless otherwise stated all molecular diagrams were rendered with *PyMOL* (Delano, 2004).

CHAPTER 3: RESULTS AND DISCUSSION

3.1. Molecular weight of RAPc8 amidase

3.1.1. Gel filtration

Gel filtration analysis of purified amidase from *Geobacillus pallidus* RAPc8 (RAPc8 amidase) on the preparative column resulted in two peaks (Figure 3-1): a 426 kDa peak near the void volume of the column and a 220 kDa peak eluting as a shoulder of the first peak. The 220 kDa peak is approximately equivalent to 6 subunits of 38 kDa each.

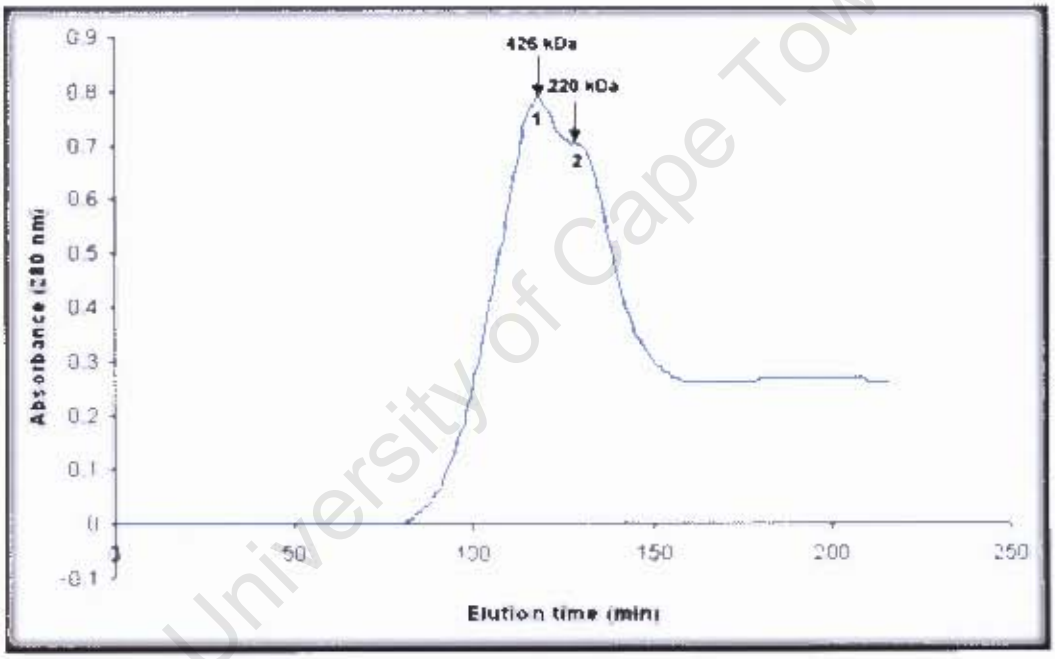


Figure 3-1: Gel filtration profile of purified RAPc8 amidase. The two peaks are labelled. The relative molecular masses were calculated based on a standard profile of Bovine Serum Albumin (BSA).

Analysis of the purified enzyme by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) (Figure 3-2) indicated over 99% homogeneity. A single band at approximately 38 kDa corresponding to the theoretical 38.6 kDa subunit molecular weight of a putative RAPc8 amidase, as derived from the translated gene sequence (Cameron et al., 2005), was observed. A high level of purity and

homogeneity is necessary for crystallization as impurities and heterogeneity can hinder successful crystallization.

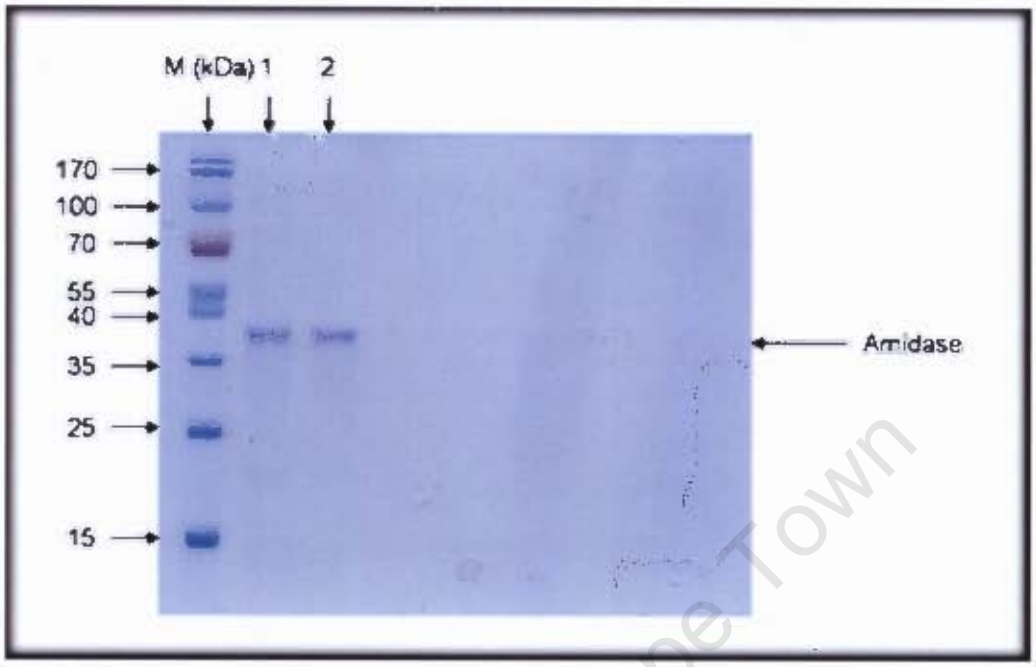


Figure 3-2: SDS-PAGE of purified RAPe8 amidase. **Lane M:** molecular weight markers; **lane 1:** A fraction of peak 1 from gel filtration chromatography; **lane 2:** A fraction of peak 2 from gel filtration chromatography. The single bands at 38 kDa indicate over 98% purity, with each peak representing a different multimeric form of the amidase.

The electrophoretic mobility of active enzyme on a native acrylamide gel electrophoresis suggested a complex of approximately 230 kDa, consistent with a homohexameric native structure (Results not shown).

3.1.2. Electron microscopy

Electron microscopy (EM) (Figure 3-3) of various protein-containing fractions from the two gel filtration peaks showed that the 426 kDa peak consisted mainly of aggregates of 10 nm particles arranged in short chains of various sizes while the 220 kDa peak contained 10 nm particles which were dispersed as single particles. The 220 kDa particles correspond to a complex of approximately six 38 kDa subunits, suggesting a homohexameric form of RAPe8 amidase. A large number of particles in the 220 kDa peak were roughly square in shape while a few had a distinct triangular

appearance, making them easy to identify. Following reference-free alignment and k-means based multivariate statistical analysis classification and averaging of single particles using *SPIDER* (Frank et al., 1996), projections with the triangular shape clustered together giving a clear class average with a 3-fold symmetry (Figure 3-3(C)).

Due to the difficulty in satisfactory classification of the images with the approximate square appearances, the aligned particles were subjected to Principal Component Analysis (PCA) (Bretaudiere, Frank, 1986). In brief, PCA is a linear transformation method in which 2-dimensional (2D) projection image matrices are first transformed into 1-dimensional (1D) image vectors. The corresponding 1D eigenvectors are then computed, after which 2D eigenimages are reconstituted from the chosen eigenvectors. The presence of symmetry in the 2D eigen images indicates underlying symmetry in the components of the images. The aligned projection images gave rise to eigenimages (Figure 3-4) with clear 2-, 3-, and 6- fold symmetry, among many other shapes. The presence of 6-fold eigenimages indicated that the particles had D3 point group symmetry. It was therefore concluded that the roughly square images, of which the 2-fold eigenimage is a component, were projections perpendicular to the triangular projections. This is consistent with the RAPc8 amidase being homohexameric in solution, and indicates the possible arrangement of the subunits in the complex; i.e. trimers of dimers possibly arranged around a 3-fold axis, to give rise to D3 point group symmetry.

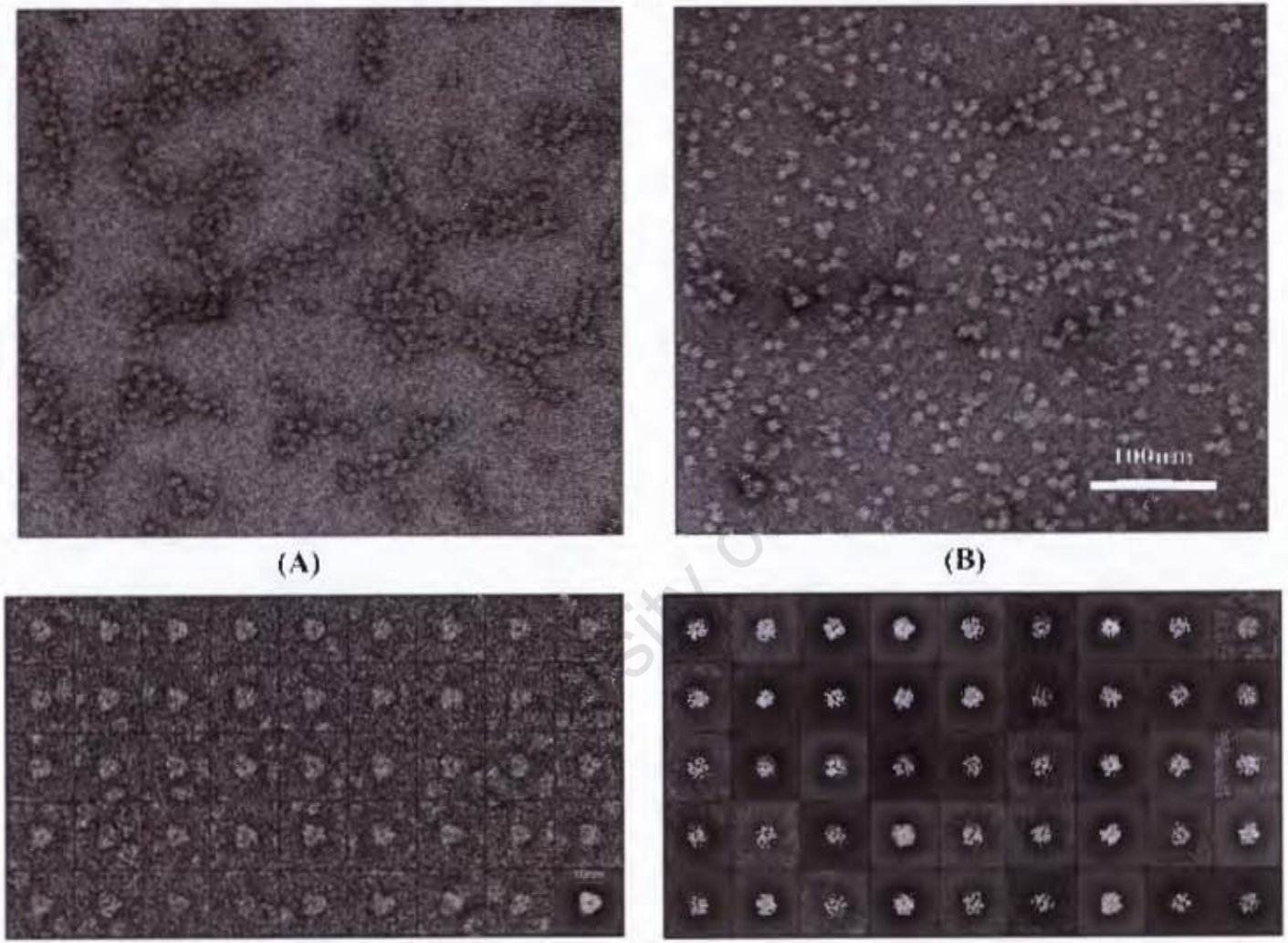


Figure 3-3: Electron microscopy analysis of fractions of purified RAPe8 amidase from Gel filtration: **(A)** aggregated amidase particles in short chains from fractions in the 426 kDa gel filtration peak, **(B)** singly dispersed amidase particles from the 220 kDa gel filtration peak. **(C)** some of the 10 nm particles with a distinct triangular appearance and their class average (lower right corner) with 3-fold symmetry and **(D)** some of the class averages from the indistinguishable projections with a roughly square appearance. Images were generated in graphical program *WEB* (Frank *et al.*, 1996).

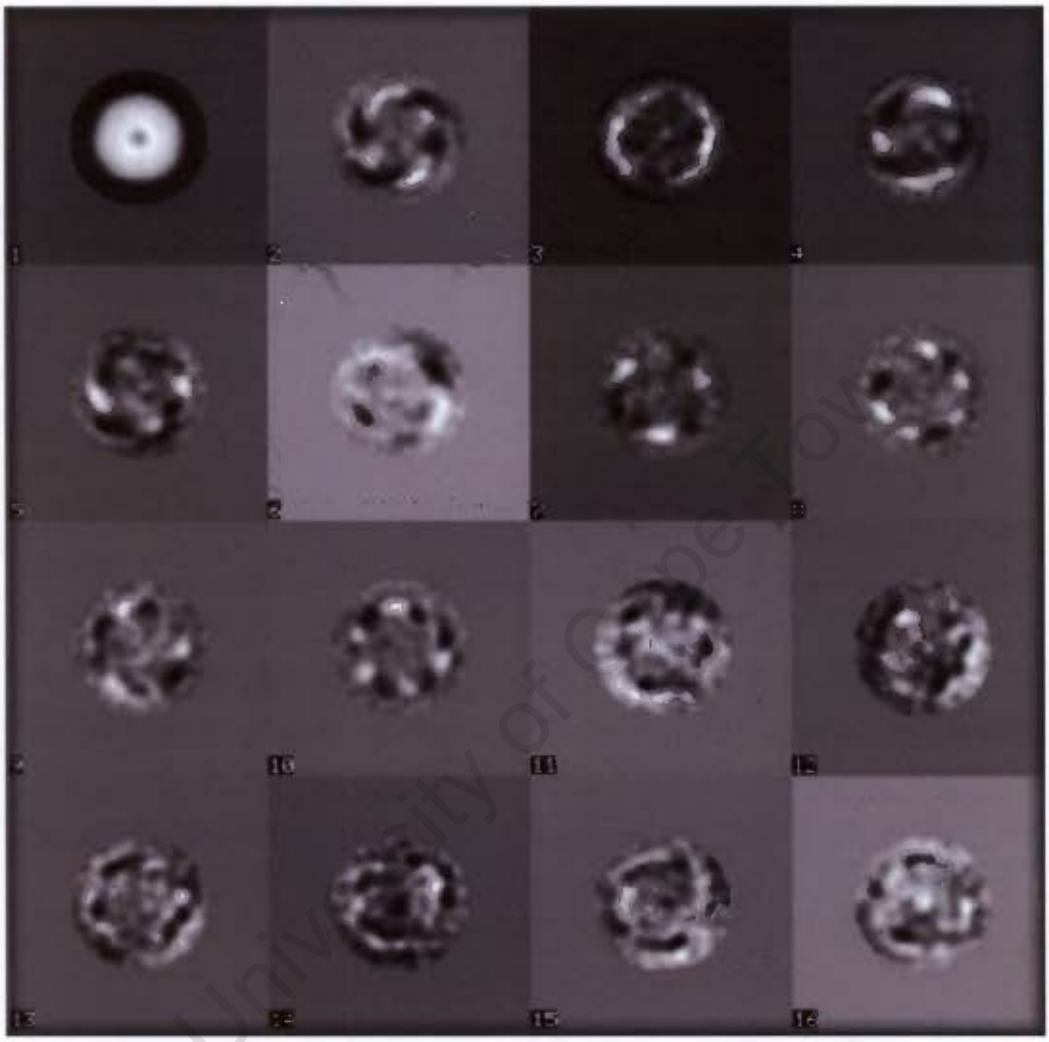


Figure 3-4: Eigenimages from Principal Component Analysis of the aligned image projections. **1**: Average of all the images; **2**: eigenimage having 2-fold symmetry; **8 and 11**: 3-fold symmetric eigenimages **10**: eigenimage with 6-fold symmetry. The other numbers represent other shapes. The image was prepared using *WEB* (Frank et al., 1996).

3.2. Predicted secondary structure for RAPc8 amidase

Advances in technology through bioinformatics approaches have made it possible to predict the secondary structure of a protein from its sequence only, with varying degrees of success resulting from the use of different methods. The secondary structure of RAPc8 amidase was predicted using a secondary structure prediction server, *PSIPRED* (McGuffin et al., 2000), a program that utilizes neural networks that have been trained to recognize certain patterns like solvation parameters (Jones, 1999b) in the sequence in order to predict the secondary structure.

The secondary structure elements of RAPc8 amidase were found to follow a pattern (Figure 3-5) that resembled that of the nitrilase superfamily homologues, suggesting a similar conserved 3-dimensional (3D) fold of the core region. The C-terminal region (which is more extended relative to other nitrilase superfamily enzymes) was predicted to have secondary structure elements including helices, which indicated that the C-terminus in the crystal was possibly highly ordered, most likely with a structural role.

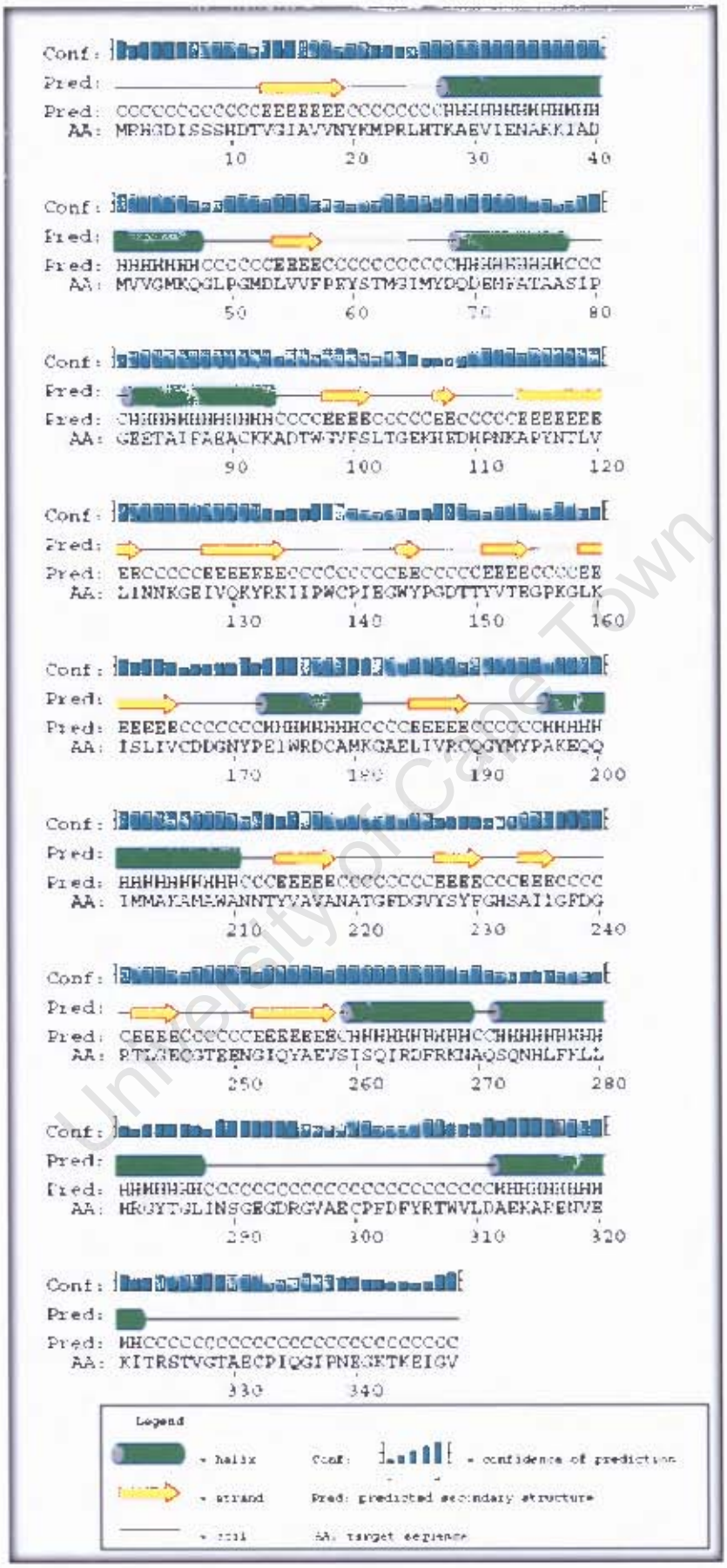


Figure 3-5: Predicted secondary structure prediction of RAPc8 amidase.

3.3. X-ray diffraction data collection

Diffraction-quality crystals were obtained using the protocol we have described (Agarkar *et al.*, 2006). The crystals (Figure 3-6), which had a tetrahedral shape with dimensions of approximately 0.2 mm x 0.1 mm x 0.1 mm, were selected for X-ray diffraction experiments. They diffracted well, producing visible spots beyond 1.8 Å resolution and having a mosaicity of 0.375°. A typical X-ray diffraction image is shown in Figure 3-7 below.

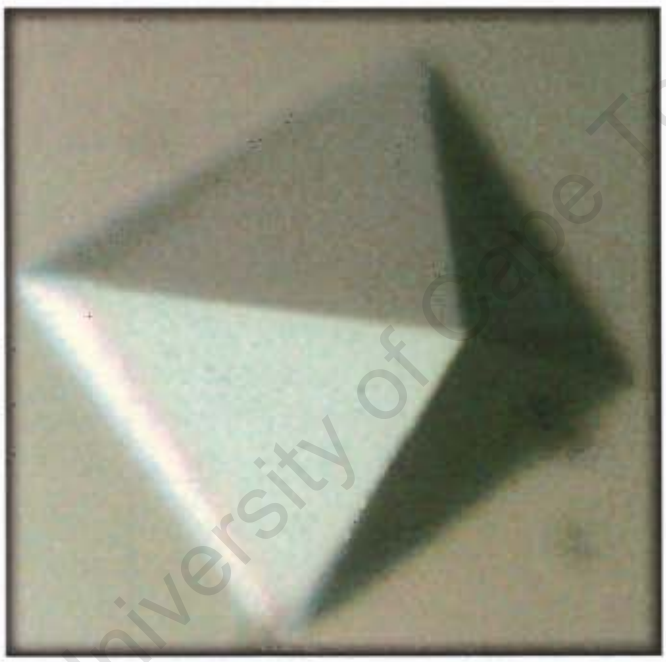


Figure 3-6: The image of a crystal from RAPc8 amidase.

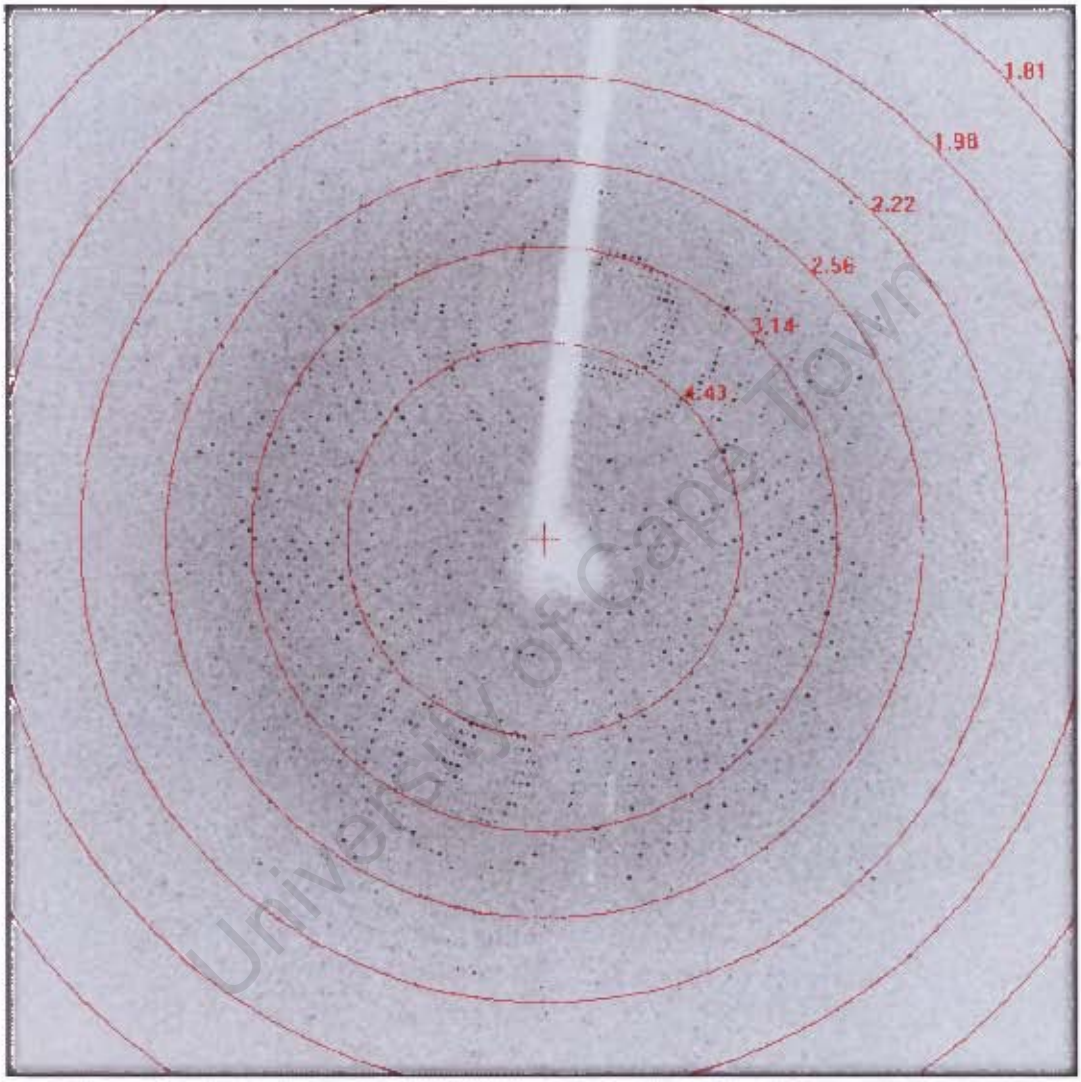


Figure 3-7: RAPc8 amidase crystal X-ray diffraction image. The spots are visible beyond 1.8 Å resolution.

3.4. Processing of X-ray diffraction data and space group determination

The 96 X-ray diffraction images were indexed and integrated using *Denzo* (Otwinowski, Minor, 1997). *Denzo,* which works in Fourier space, analyzes the layout of the diffraction spots to check for the most common spacings between them and assigning them to the most consistent Bravais lattice. The indexing procedure therefore assigns miller indices (h, k, l) to each reflection and also gives initial estimates of unit cell parameters, crystal orientation and mosaicity. The refinement procedure improves these estimates for one diffraction image. Integration of the entire dataset using *Denzo* involves the use of the estimated parameters to calculate spot profiles for each image, after which a profile fitting method is used to predict spots in all the images. The Bravais lattice (unit cell) of the RAPc8 amidase crystal was unequivocally determined as Primitive Cubic. There are two possible crystallographic point groups in primitive cubic unit cell; $23 < 432$ in order of ascending symmetry.

Scaling and merging of the data were performed using the HKL suite program *SCALEPACK* (Otwinowski, Minor, 1997). Each image in a dataset may have its recorded intensities on a different scale due to various physical reasons, including: variations in intensity of the incident beam, absorption of diffracted rays, radiation damage, incorrect calibration of the detector and shadows from the beam stop among others. In the process of merging symmetry-equivalent reflections from different images into a set of unique reflections, *SCALEPACK* applies scale factors that put all observations on a common scale. It then does post-refinement that optimizes unit cell and crystal orientation parameters based on the knowledge of the observed locations of all reflections in the entire dataset. The data scaling and merging procedure also provides valuable statistics on the quality of data as well as information on the space group.

In an attempt to discern the space group (or the point group) of the RAPc8 amidase crystal with *SCALEPACK* (Otwinowski, Minor, 1997), the data were scaled and merged in the space groups with the highest symmetry from each point group (space

group $P2_13$ for 23 point group and space groups $P4_132$, $P4_232$ and $P4_332$ for 432 point group). The merging statistics were good, with reasonable numbers of outliers in all cases. *SCALEPACK* output a list of likely axial reflections in all cases. However, most of the listed reflections in all the four space groups were too strong (with high signal-to-noise ratio) for truly systematically absent reflections. At this stage, it was hard to tell if the strong axial reflections were due to background noise. Since neither the presence (or absence) of systematic absences nor the merging statistics could offer reliable guide in distinguishing between the space groups, the data were merged in the two space groups with the lowest symmetry (P23 and P432) and subjected to further analysis.

3.4.1. Characterization of RAPc8 amidase datasets

(A) - Matthews coefficient (volume) calculation

The Matthews coefficient (Vm) (Matthews, 1968) allows estimation of the number of macromolecular subunits in the asymmetric unit. It is calculated as:

Vm = V / MW * Z * X where,

V is the volume of the unit,

MW is the molecular mass of the molecule,

Z is the number of the asymmetric units in the unit cell, and

X is the number of molecules in the asymmetric unit.

The program *PHASER* (Collaborative Computational Project Number 4 (CCP4)., 1994; Read, 2001) was used to calculate Vm for both P23- and P432-merged datasets to 2 Å resolution.

P23-merged data had a Matthews coefficient of 2.41 with 2 amidase molecules of 38,596 Daltons each and 48.90% solvent in the asymmetric unit. P432-merged data on the other hand, had a Matthews coefficient of 2.41 with 1 amidase molecule of 38,596 Daltons and 48.90% solvent in the asymmetric unit. The Vm values in both datasets were within the empirically observed range as reported by Matthews (1968).

(B) - Self rotation function calculation

If the Matthews coefficient suggests 2 or more molecules in the asymmetric unit, a self rotation function of the native data is usually calculated. This is important in determining if the molecules in the asymmetric unit are related by a rotational symmetry axis, and if so, the orientation of the axis in the unit cell (direction of the rotation axis). It is a useful criterion with which to check for the presence of Non-Crystallographic Symmetry (NCS). In principal, the self rotation function compares a native Patterson with itself, with the peaks corresponding to the rotation that superimposes multiple copies of the molecule in the asymmetric unit. Peaks corresponding to crystallographic symmetry are similar in height to the origin peak. The program *POLARRFN* (CCP4, 1994) that calculates fast self-rotation function in polar angles was used to calculate a self rotation function for the RAPc8 amidase P23 and P432 datasets. Even though there was no possibility of NCS in the P432 asymmetric unit, the dataset was included in the calculations for comparison purposes.

Peaks (Figure 3-8) corresponding to crystallographic four-fold, three-fold and two-fold axes (kappa = 90°, 120°, and 180° respectively) were observed in P432 dataset as expected. The same peaks were observed in P23 dataset but in this case, the four-fold peaks were non-crystallographic. Four-fold NCS in the P23 dataset was not consistent with two molecules in the asymmetric unit, as suggested by the Matthews coefficient. These results clearly indicated that the amidase crystal belonged to the 432 point group; consistent with the crystallographic four-, three- and two-fold axes, and one molecule in the asymmetric unit.

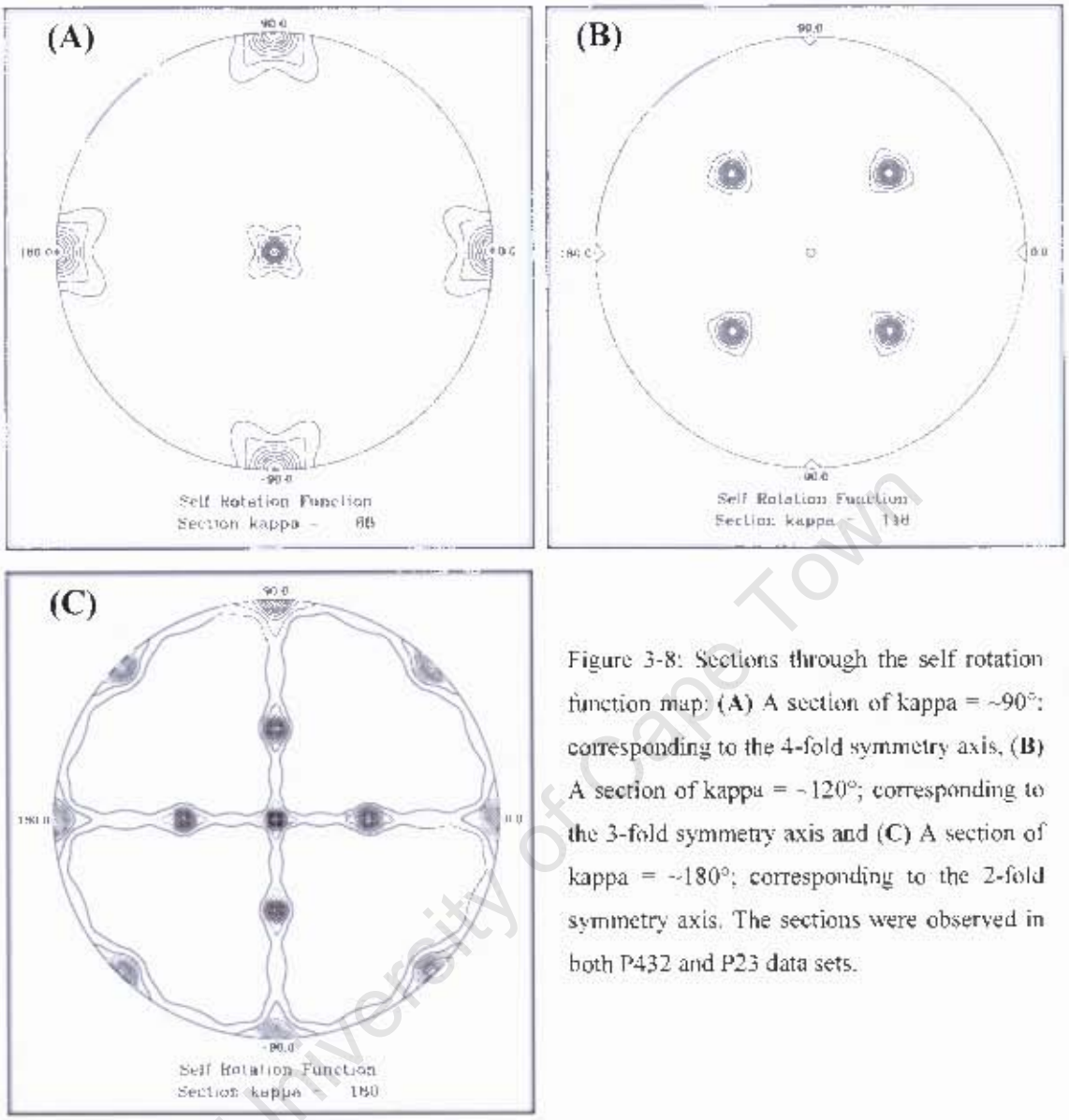


Figure 3-8: Sections through the self rotation function map: **(A)** A section of kappa = ~90°; corresponding to the 4-fold symmetry axis, **(B)** A section of kappa = ~120°; corresponding to the 3-fold symmetry axis and **(C)** A section of kappa = ~180°; corresponding to the 2-fold symmetry axis. The sections were observed in both P432 and P23 data sets.

To confirm the point group, and to obtain full information on the space group, further analysis of the data was performed using a pre-released version of *POINTLESS* (CCP4, 1994; Evans, 2006).

(C) - Determination of Laue group, point group and space group using *POINTLESS*

POINTLESS (CCP4, 1994: Evans, 2006) uses maximum likelihood targets to score the agreement (reflected by a correlation coefficient) between reflections that are likely to be related by rotational symmetry (Evans, 2006), hence providing

information on the Laue group of the diffraction data. As clearly described by Evans (2006), *POINTLESS* reads unmerged intensities, determines the Lattice with the highest symmetry from the unit cell parameters, scores each rotational symmetry element separately, groups the symmetry elements into subgroups and compares the symmetry subgroups to obtain a net score for each subgroup. Upon successful determination of Laue and point group symmetries, *POINTLESS* proceeds to test for systematic absences through Fourier analysis of signal intensity-to-noise ratio ($I/\sigma(I)$ for each zone in which some reflections are likely to be systematically absent.

Unmerged RAPc8 amidase crystal data that had been indexed, integrated and scaled in the P1 space group using *SCALEPACK* (Otwinowski, Minor, 1997) having unit cell parameters $a = b = c = 130.5$Å, $\alpha = \beta = \gamma = 90°$, were used. The m-3m Laue group had the highest score (Z-score = 13.81) and all the expected rotational symmetry elements (2-, 3- & 4-fold axes) were clearly present (Table 3-1). Since 432 is the only point group corresponding to m-3m Laue group (Suh et al., 1993), 432 was therefore confirmed to be the point group for RAPc8 amidase crystal.

Table 3-1: Scores for each symmetry element in Laue group m-3m. Correlation coefficient Z-score (Z-CC) was used to score the agreement of the symmetry-related observations in each rotational symmetry element. The 2-fold, 3-fold and 4-fold axes are clearly present in RAPc8 amidase crystal, suggesting m-3m Laue group.

	Symmetry element & operator	**Correlation coefficient Z-scores (Z-CC)**	**Correlation coefficient (CC)**
1	Identity	13.82	0.96
2	2-fold [101]	13.89	0.96
3	2-fold [10-1]	13.98	0.97
4	2-fold [01-1]	13.92	0.97
5	2-fold [011]	13.70	0.95
6	2-fold [1-10]	13.79	0.96
7	2-fold k [010]	13.94	0.97
8	2-fold [110]	13.93	0.97
9	2-fold h [100]	13.93	0.97
10	2-fold l [001]	14.07	0.98
11	3-fold [111]	13.79	0.96
12	3-fold [1-1-1]	13.72	0.95
13	3-fold [1-11]	13.69	0.95
14	3-fold [11-1]	13.65	0.95
15	4-fold h [100]	13.62	0.94
16	4-fold k [010]	13.90	0.96
17	4-fold l [001]	13.74	0.95

A test of the m-3m Laue group for systematic absences, suggested a possibility of axial reflections on the 4(2) [b] screw axis (Table 3-2). However, most of the flagged reflections (list not shown) had very high ($I/\sigma(I)$) values; similar to the axial reflections that had been flagged by *SCALEPACK* during data scaling and merging.

Table 3-2: Possibility of systematic absences in zones 4(2) and 4(1). Peak height is the value in Fourier space at the relevant screw (translation) point relative to the origin. This should be close to 1.0 if systematic absences are present. Probability is the likelihood for the symmetry element to be present.

Zone	Peak Height	Standard Deviation	Probability	Reflection	Condition
Screw axis 4(2) [b]	0.659	0.138	0.129	0k0	k=2n
Screw axis 4(1) [b]	-0.164	0.086	0.000	0k0	k=2n

A negative peak height and a probability of 0 for the systematic absences along 4(1) screw axis (Table 3-2), ruled out the possibility of having $P4_132$ or $P4_332$ as the space groups for the RAPc8 amidase. *POINTLESS* (CCP4, 1994) could not distinguish between P432 and $P4_232$ space groups.

(D) - Molecular replacement success as a strategy for distinguishing between P432 and $P4_232$ as the possible space groups for RAPc8 amidase crystal.

Based on the assumption that the right space group would yield a correct molecular replacement (MR) solution. MR searches were performed using the programs *MOLREP* (CCP4, 1994; Vagin, Teplyakov, 1997), *AMoRe* (CCP4, 1994; Navaza, 1994) and *EPMR* (Kissinger et al., 1999) with both P432- and $P4_232$-merged datasets, using four nitrilase superfamily homologues as search probes. No clear MR solution was obtained with either of the two datasets (data not shown). The RAPc8 amidase crystal space group could therefore not be determined using this strategy.

(E) - Manual packing of the P432 cubic unit cell

Following the insights obtained from the electron microscopy (EM) analysis of the RAPc8 amidase biological complex (i.e. the likelihood of RAPc8 amidase being homohexameric in solution, with each hexamer being comprised of trimers of dimers with a clear 3-fold axis), it was found necessary to explore the packing of P432 and $P4_232$ unit cells manually. Trimers of dimers can exist in either of these two space groups.

A P432 cubic unit cell (Figure 3-9) has 3-fold axes at the corners of the cube directed towards the centre, 2-fold axes at the middle of the sides of the cube, also directed towards the centre and 4-fold axes on the faces of the cube also pointing towards the centre. The structurally-related nitrilase superfamily enzymes have a conserved dimer axis that is likely to be present in RAPc8 amidase complex. Assuming one subunit per asymmetric unit (from the Matthews coefficient), the dimer interface would lie on the crystallographic 2-fold axis. The arrangement of the subunits to form a hexamer comprising of trimers of dimers, would place three dimers around the crystallographic 3-fold axis. With this arrangement in place, the orienting and positioning of the model(s) in the unit cell is reduced from a 6-dimensional (6D) to a 2D search problem; i.e. distance (translation) of the dimer from the centre of the cube and rotation of the dimer around the 2-fold axis. A sample of a packed P432 unit cell, with a 7 Å translation towards the centre of the cube and a rotational angle of -35° around the 2-fold axis is shown in figure 3-10 below.

Figure 3-9: A sample of P432 cubic unit cell with the 4-fold axes on the faces of the cube (green) and the 3-fold axes on the corners of the cubes (red). The 2-fold axes (not marked) are located in the middle of the sides of the cube. The figure was taken from Dale Minerals International website (Online).

A simulation of all possible combinations of translation distance(s) and rotation angle(s) using one of the nitrilase superfamily homologues as the search probe did not give rise to a packing arrangement (Figure 3-10) similar to the observed 3-fold

symmetric projection images (Figure 3-3 (c)) from the electron micrographs. In all cases, unlikely biological packing in which the 2-fold axes intersected with the 3-fold axes at a position remote from perpendicular was observed. A check of clashes between molecules in the packed unit cell, calculation of phases and electron density maps confirmed that the RAPc8 amidase homohexamer could not pack in a P432 unit cell.

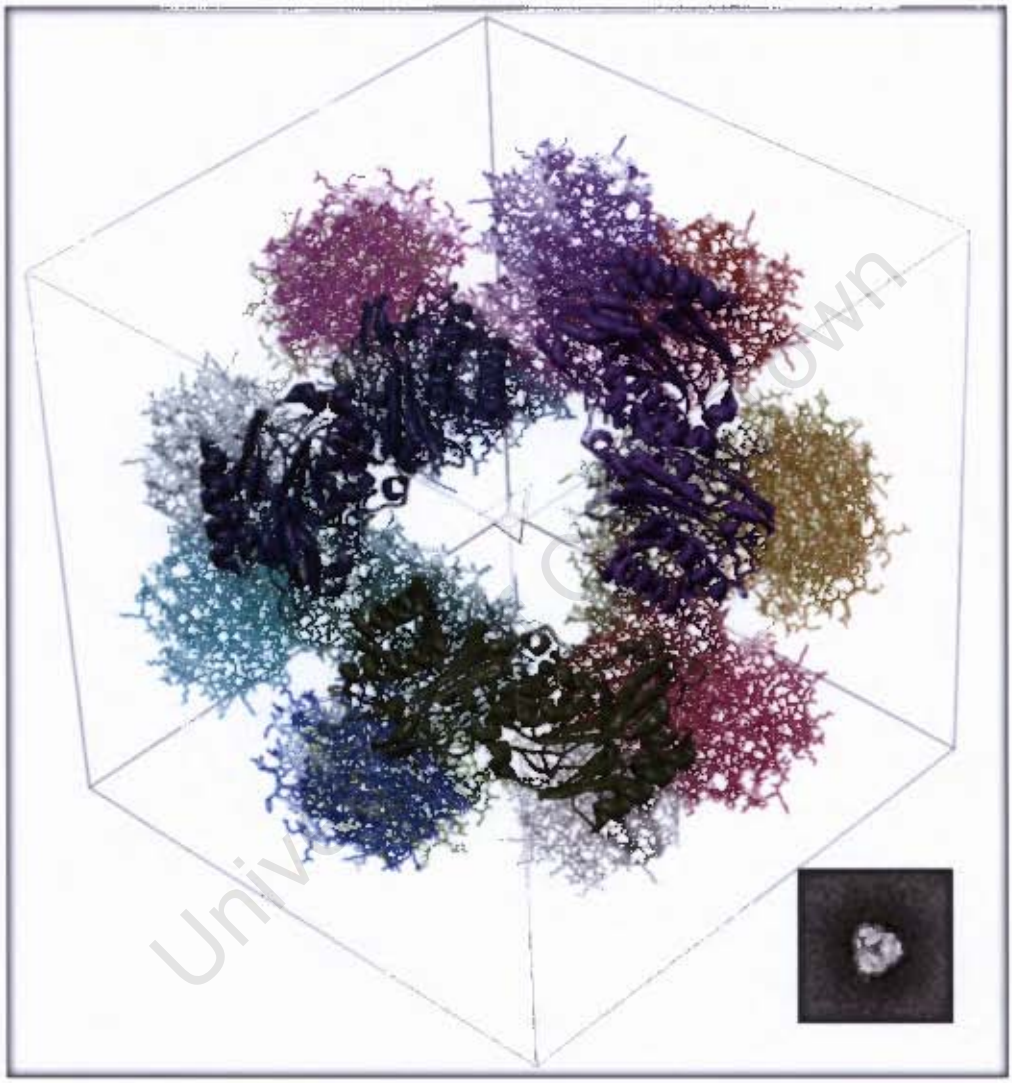


Figure 3-10: A P432 cubic unit cell packed with 24 molecules of a RAPc8 amidase homologue (1f89), with a translation distance of 7 Å towards the centre of the cube and a -35° rotation around the 2-fold axis, as viewed along the 3-fold axis. The six subunits forming one hexamer (trimer of dimers) are rendered as cartoons. The hexameric packing does not resemble the 3-fold symmetric projections observed in the electron micrographs (inset). The packed unit cell was rendered with *UCSF-Chimera* (Pettersen et al., 2004).

Packing in the $P4_232$ unit cell was not exhaustively attempted. However the space group of the RAPc8 amidase crystal was confirmed to be $P4_232$ when the MR solution was obtained. This solution was determined after the phasing program *PHASER* (CCP4, 1994; Read, 2001) was set to perform exhaustive exploration of translations of many rotation function peaks (including smaller ones) in all alternative space groups, as described later in the MR section. There are only ten deposited structures with the space group $P4_232$ in the Brookhaven Protein Data Base (PDB), six of which are viruses.

(F) - Visualization of systematic absences in the RAPc8 amidase $P4_232$ space group

Systematically-absent reflections on the 4(2) screw axis were visualized as pseudo-precession images (Figure 3-11) of the hk0 zone in reciprocal space using *HKLVIEW* (CCP4, 1994). Merged RAPc8 amidase structure factor amplitudes were used to generate the images. All odd reflections in both h and k principal axes were found to be extremely weak (absent), relative to the even ones. This led to the conclusion that, the high signal-to-noise ratio that was observed in the detected axial reflections was partly a contribution from background noise.

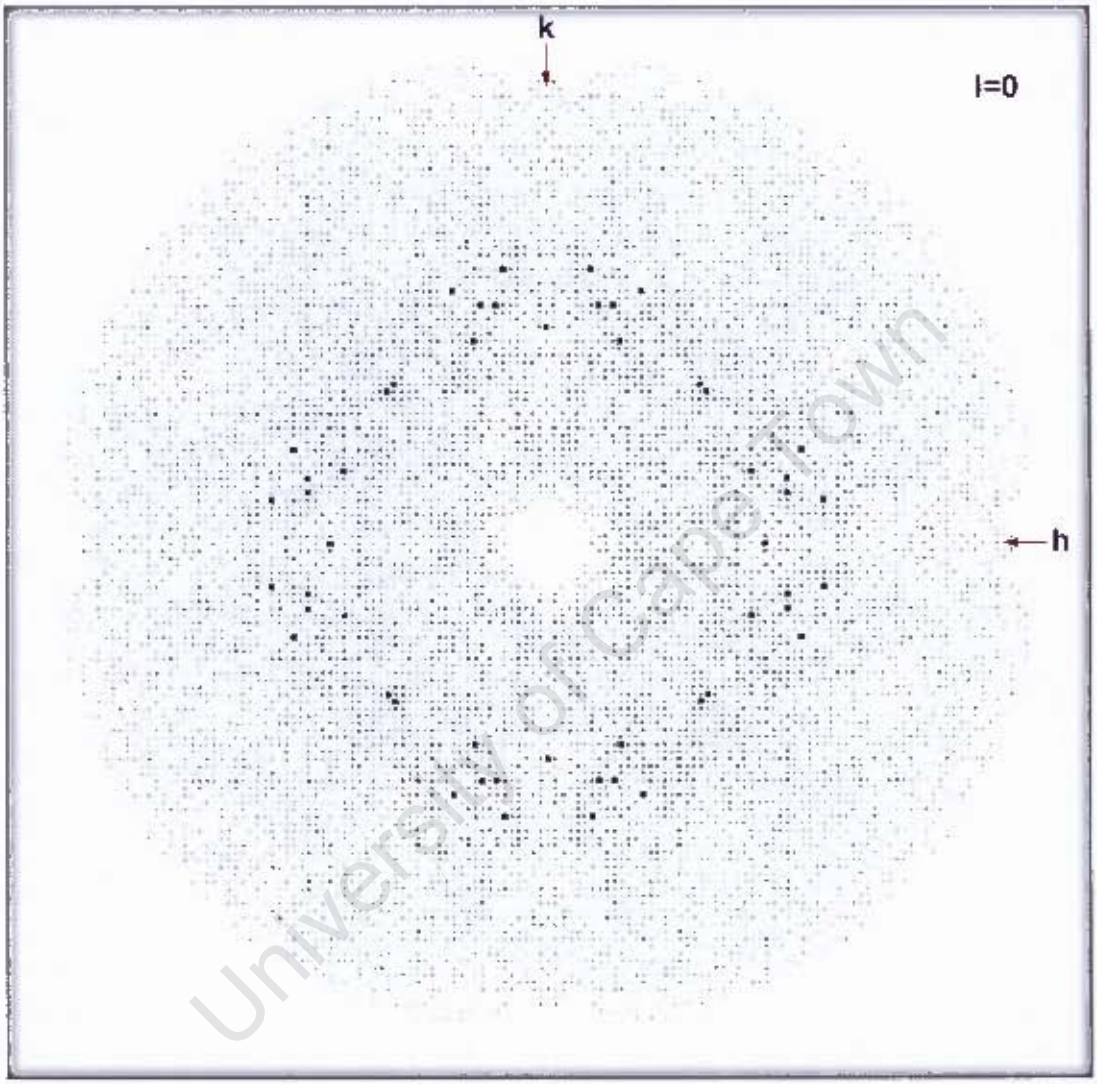


Figure 3-11: hk0 zone for RAPc8 amidase diffraction data. The darkness of the spots (square) is a consequence of the amplitudes.

3.5. Assessment of diffraction data quality

The quality of a 3D crystal structure is dependent on the information content of the diffraction data (Weiss, 2001). The overall quality of diffraction data can be deduced from the analysis of the agreement between equivalent reflections after scaling, which also helps in identifying parts of the data that agree poorly with the rest of the data (Evans, 2006). This allows for omission of the 'bad' data during the structure determination process.

Aside from the data quality statistics generated by *SCALEPACK* (Otwinowski, Minor, 1997) upon scaling and merging of symmetry-related reflections, merged RAPc8 amidase data was also analysed using the *Xtriage* program of the *PHENIX* suite (Adams et al., 2002). *Xtriage* is a command line utility that combines twin analysis with other data quality indicators to automatically assess the diffraction data. The following parameters were analysed:

Checking for bad parts of the data

A plot of merging R-factor (R_{sym}) against batch/frame number (Evans, 2006) (Figure 3-12) showed no parts of the data that were problematic: the R_{sym} values were low (11.1% on average), with no big shifts between the images. All 96 images were good, and all data were usable.

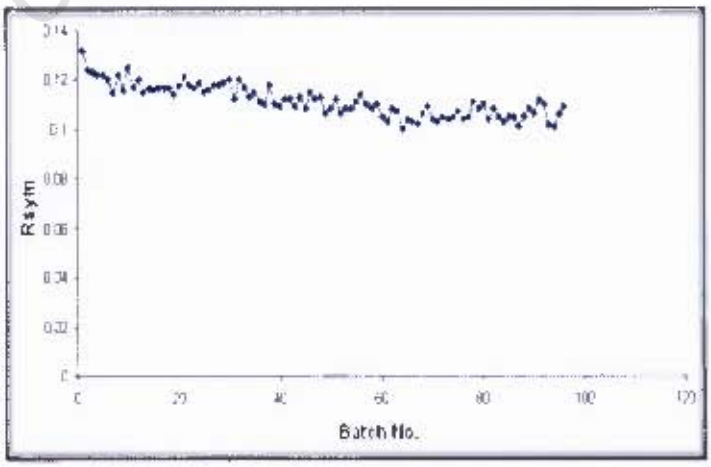


Figure 3-12: A plot of R_{sym} against batch number showed no problematic frames. The images had an overall R_{sym} of 11.1%.

A plot of scale- and B-factors against Batch/frame number (Evans, 2006) (Figure 3-13) showed constant scale factors but a slight increase in B-factors with increasing exposure time (batch number). The slight increase in B-factors was possibly as a result of mild radiation damage during the data collection process, which is expected. The radiation damage-associated deterioration was minimal and therefore unlikely to affect the quality of the collected RAPc8 amidase data.

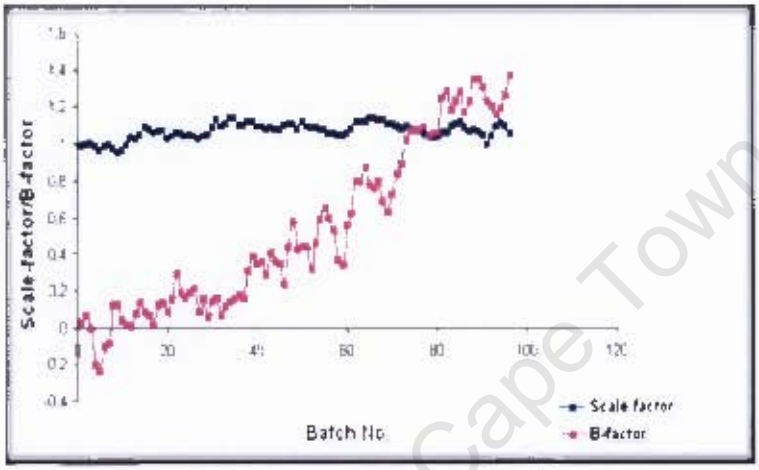


Figure 3-13: A plot of scale factors (Blue squares) and B-factors (pink squares) against batch number. The slight increase in B-factors is an indication of crystal deterioration as more frames were collected.

A Wilson plot (mean intensity as a function of resolution) (Figure 3-14) of all the observed RAPc8 amidase data was normal, with a reasonably straight line for data above 4 Å (corresponding to scattering from ordered atoms) and an expected dip at about 5 Å (where the disordered solvent contribution to structure factor amplitudes is high).

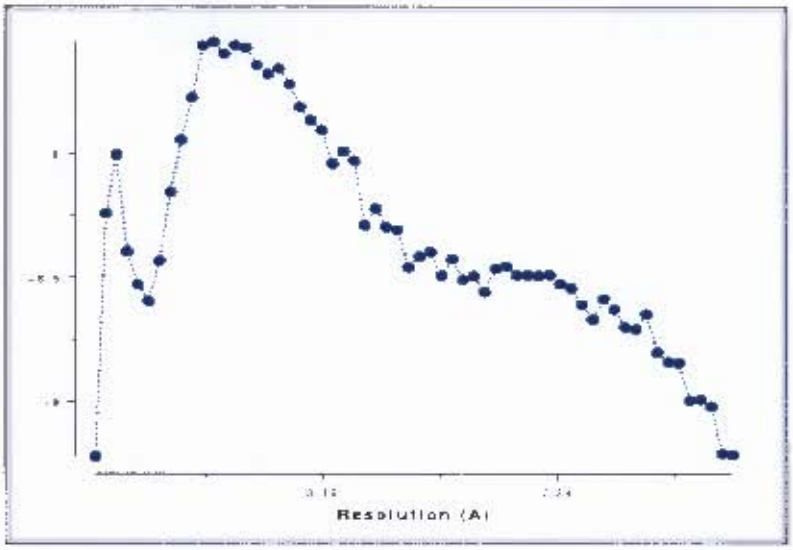


Figure 3-14: A Wilson plot of all RAPc8 amidase data. The plot was 'normal' with no bad region (s) identified.

Other quality checks

Analysis of the merged data with the *Xtriage* program found no possible outliers among acentric and centric reflections (Read, 1999). A check by the same program did not detect any ice rings-related problems in the ice ring-sensitive resolution ranges. Although no twinning was expected in the $P4_232$ space group, crystal twinning analysis found no merohedral or pseudo-merohedral twin operators as expected. Native Patterson peaks analysis showed no probability of translational pseudo symmetry in RAPc8 amidase data.

Data collection, crystallographic parameters and overall data quality statistics as assessed by various indicators are presented in Table 3-3 below.

Overall, the collected RAPc8 amidase data was of high quality (Table 3-3), likely to yield a high quality model upon structure determination. An average B-factor of 12.70 $Å^2$ was indicative of a highly ordered crystal, unlikely to have many regions of high flexibility, mobility and/or disorder. Although diffraction spots had been observed beyond 1.8 Å on the images, a nominal resolution (extent to which features that are close to each other in space can be differentiated) cut-off of 1.90 Å was decided on, based on the signal-to-noise ratio as deduced from ($I/\sigma(I)$), the merging R-factor

(R_{sym}) and the χ^2 value of the outer resolution shell. This ensured that very weak reflections were omitted from the data set.

An overall merging R-factor (R_{sym}) of 11% indicated high accuracy in the measurement of identical reflections, while a χ^2 value of close to unity (1.04) was an indication of how well the declared scaling error model matched the actual errors in the reflections data. The data was of high redundancy (7.9) suggesting high accuracy in the averaged measurements, an indication of intrinsically high quality data (Weiss, 2001). A completeness level of 99.3% indicates that almost all of the theoretically expected unique reflections had been measured.

Table 3-3: Crystallographic parameters, data collection statistics and quality indicators of RAPc8 amidase diffraction data. Under the data quality parameters, both low resolution (inner) and high resolution (outer) data shells are presented.

Data Collection Statistics: Crystal parameters	
Wave length (Å)	1.5418
Rotation range (°)	48
Space group	$P4_232$
Unit cell Parameters	a = b = c = 130.39Å; α = β = γ = 90.00°
Unit cell Volume ($Å^3$)	2216322
Matthews Coefficient ($Å^3$ Da^{-1})	2.41
Solvent content (%)	48.90
Number of molecules per asymmetric unit	1
Data Collection statistics: Data quality parameters	
Nominal Resolution Range (Å) (inner shell; outer shell)	46.1 – 1.90 (46.1 - 4.09; 1.97 - 1.90)
Total observations (unique)	240879 (30300)
Completeness (%) (inner shell; outer shell)	99.30 (99.70; 94.80)
Redundancy (inner shell; outer shell)	7.90 (8.30; 3.90)
Signal-to-noise ratio (I/σ(I)) (inner shell; outer shell)	18.27 (43.12; 3.23)
R_{sym} [†] (%) (inner shell; outer shell)	11.0 (6.10; 31.5)
χ^2 (inner shell; outer shell)	1.04 (1.33; 0.59)
Wilson plot average B-factor ($Å^2$)	12.70

[†]$R_{sym} = \sum \left| I - \langle I \rangle \right| / \sum \langle I \rangle$ in which I is a measured intensity and <I> is the average intensity from multiple measurements of symmetry-related reflections.

3.6. Molecular replacement

The phase information of X-ray diffraction data is lost during data collection. Molecular replacement (MR) (Rossmann, 1990) is an appropriate method for obtaining the phase when known homologous structures exist. MR encompasses techniques which aim at orienting and positioning the search model (of a previously solved structure) so that it coincides with the position of the unknown (target) protein in the crystal. If the search molecule is correctly orientated and positioned in the target cell, a preliminary model of the target structure (MR solution) is obtained, which provides estimated phase information for the target protein, allowing electron density maps to be calculated. The MR solution is optimized by rigid body refinement after which the structure is taken through alternating cycles of map calculation, model fitting, rebuilding and refinement to help obtain the correct positions of atoms in the target protein and also to remove the bias introduced by the starting model.

Orienting the search model requires three rotational parameters (α, β, γ) while positioning it relative to the origin requires 0 (for P1 space group), 2 (for polar space groups) or 3 (for all other space groups) translational parameters. The MR search is therefore a 6D problem for most space groups, with billions of possible combinations (Hazes, 2003). To simplify the MR procedure, a "divide and conquer" approach is used where the problem is broken into a 3D rotational search to obtain the orientation of the molecule, followed by a 3D translational search to position the model in the unit cell. In the traditional MR methods, the rotational search compares intramolecular vectors in the Patterson map (an interatomic vector Fourier map calculated using the square of the structure factor amplitudes and a phase of zero) of the observed data with those calculated from the search model, as the model is rotated to match the observed data. The translational searches on the other hand are performed by two main methods traditionally: (1) by comparing intermolecular vectors in the Patterson map of the target protein with those of the model as the model is moved around in the unit cell or (2) by an R-factor search method which basically involves the calculation of an R-factor as the search model and its symmetry-related molecules are moved through the unit cell of the target protein. The correct position gives the lowest R-

factor. A correlation coefficient (CC) is also calculated in some cases, to measure the agreement between calculated and observed structure factors as the search model is moved in the unit cell.

The success of MR is largely dependent on the similarity (percentage sequence identity) between the search model and the target protein (Hazes, 2003). In addition to similarity, the completeness of the search model (extent to which the search model represents the target protein) also plays a substantial role. Of great importance too, is the completeness and overall quality of the observed data.

The MR procedure for RAPc8 amidase comprised of four main steps as described by Rossmann (1990). These were: (1) finding solved structures that are homologous to RAPc8 amidase for use as search models, (2) determining the relative orientation of the search probe in RAPc8 amidase unit cell, (3) defining the position of the model in the unit cell relative to the origin and (4) calculating the phases using the correctly-oriented and -positioned model.

3.6.1. Search for homologues

Since no known structure of *Geobacillus pallidus* RAPc8 amidase in any other form (in another functional state or in different space group) was available, *GenTHREADER* (Jones, 1999a) was used to search for existing homologous structures in the Brookhaven Protein Data Base (PDB) (Berman et al., 2000). *GenTHREADER* is a protein fold recognition tool that uses a protein sequence-to-structure threading technique to identify pairs of proteins (either homologous or analogous) that have similar folds. It first utilizes the *PSIPRED server* (McGuffin et al., 2000) to predict the sequence of the secondary structure elements in the target protein and then uses *HOMSTRAD* (Homologous Structures Alignment Database) (Mizuguchi et al., 1998) to pull out related homologues.

A search for RAPc8 amidase structural homologues produced four independent crystal structures with certainty (average score of 0.97): the hypothetical protein PH0642 from *Pyrococcus horikoshii* (PDB ID, 1j31; Sakai et al., 2004), a putative C-

N hydrolase from yeast (PDB ID, 1f89; Kumaran et al., 2003), the N-carbamyl-D-amino acid amidohydrolase (DCase) from *Agrobacterium* sp. KNK712 (PDB ID, 1erz; Nakai et al., 2000) and the Nit domain of the worm NitFhit fusion protein (Nit) (PDB ID, 1ems; Pace et al., 2000). The structures had varying percentage sequence identity to the amidase, as presented in Table 3-4 below. All the four homologues belong to the nitrilase superfamily, one of which (DCase; 1erz) has known function. A homology model of an aliphatic amidase from *Pseudomonas aeruginosa* (PamiE) (PDB ID, 1k17; Novo et al., 2002) based on the crystal structure of Nit (1ems) also exists, which is 81.4% identical to RAPc8 amidase.

Table 3-4: A comparison between RAPc8 amidase and solved structures of distant homologues in the nitrilase superfamily. The number of residues in each protein is highlighted in bold (black) whereas the number of identical residues at conserved positions and percentage sequence identity are highlighted in red and blue respectively. The four homologues are on average 19.2% identical to RAPc8 amidase. All comparisons were done using *GenTHREADER*.

Protein/PDB ID	RAPc8 Amidase	1j31	1f89	1erz	1ems
RAPc8 amidase	**348**	61	59	56	58
1j31	21.4%	**262**	71	80	67
1f89	20.7%	27.1%	**271**	59	85
1erz	18.8%	30.5%	21.4%	**303**	62
1ems	15.8%	25.6%	31.4%	20.4%	**296**

To identify the structurally-conserved features, coordinates from the four homologues were first superimposed (Figure 3-15) by pair-wise α-carbon-atom alignment using *ALIGN* (Cohen, 1997). A structural alignment (Figure 3-16) between RAPc8 amidase and the four homologues was then generated manually, guided by the predicted RAPc8 amidase secondary structure elements (Figure 3-5), as well as by *GenTHREADER* (Jones, 1999) and *ALIGN* pair-wise alignments.

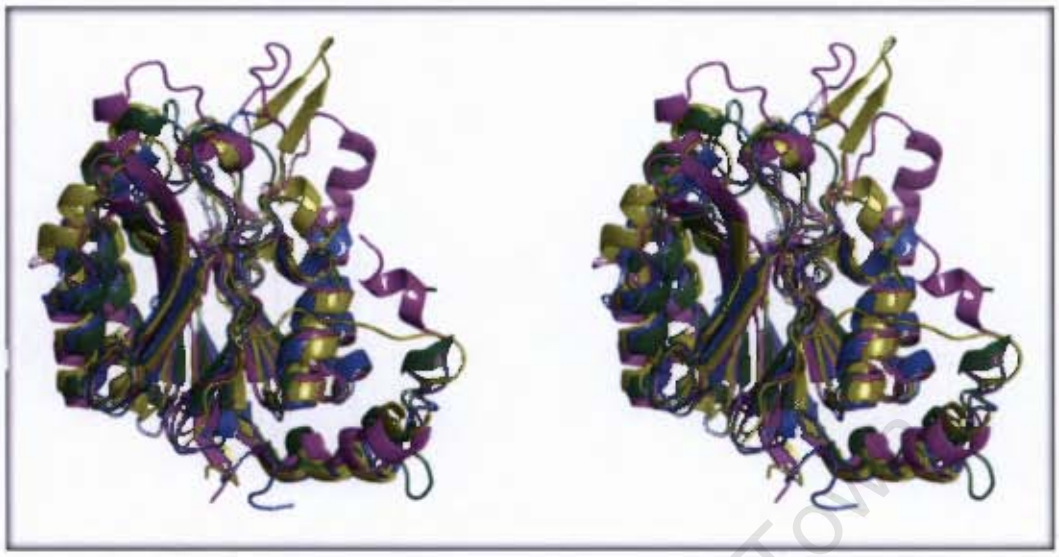


Figure 3-15: A stereo view of cartoon representation of the superimposed monomers from the four nitrilase superfamily crystal structures; 1j31 is shown in green, 1f89 in blue, 1erz (DCase) in purple and 1ems (Nit) in yellow. The four proteins have a conserved α-β-β-α hydrophobic core, with major variability being observed in the external loop regions. The cartoon was rendered using *PyMOL* (Delano, 2004).

The predicted secondary structure elements in RAPc8 amidase (Figure 3-5) coincide with the conserved regions in the homologues (Figure 3-16). Although the sequence identity between RAPc8 amidase and the distant homologues is only 19.2% (on average), the α-β-β-α structural fold is conserved in these proteins; consistent with the idea that structure is better conserved than sequence during evolution, especially the hydrophobic core and the enzymatic active site.

Before rotational and translational MR searches were performed, the four homologues (1j31, 1f89, 1erz & 1ems) were superimposed using *ALIGN* (Cohen, 1997) to put them on the same orientation prior to MR trials. This was necessary to allow the rotation function with different models to be directly comparable.

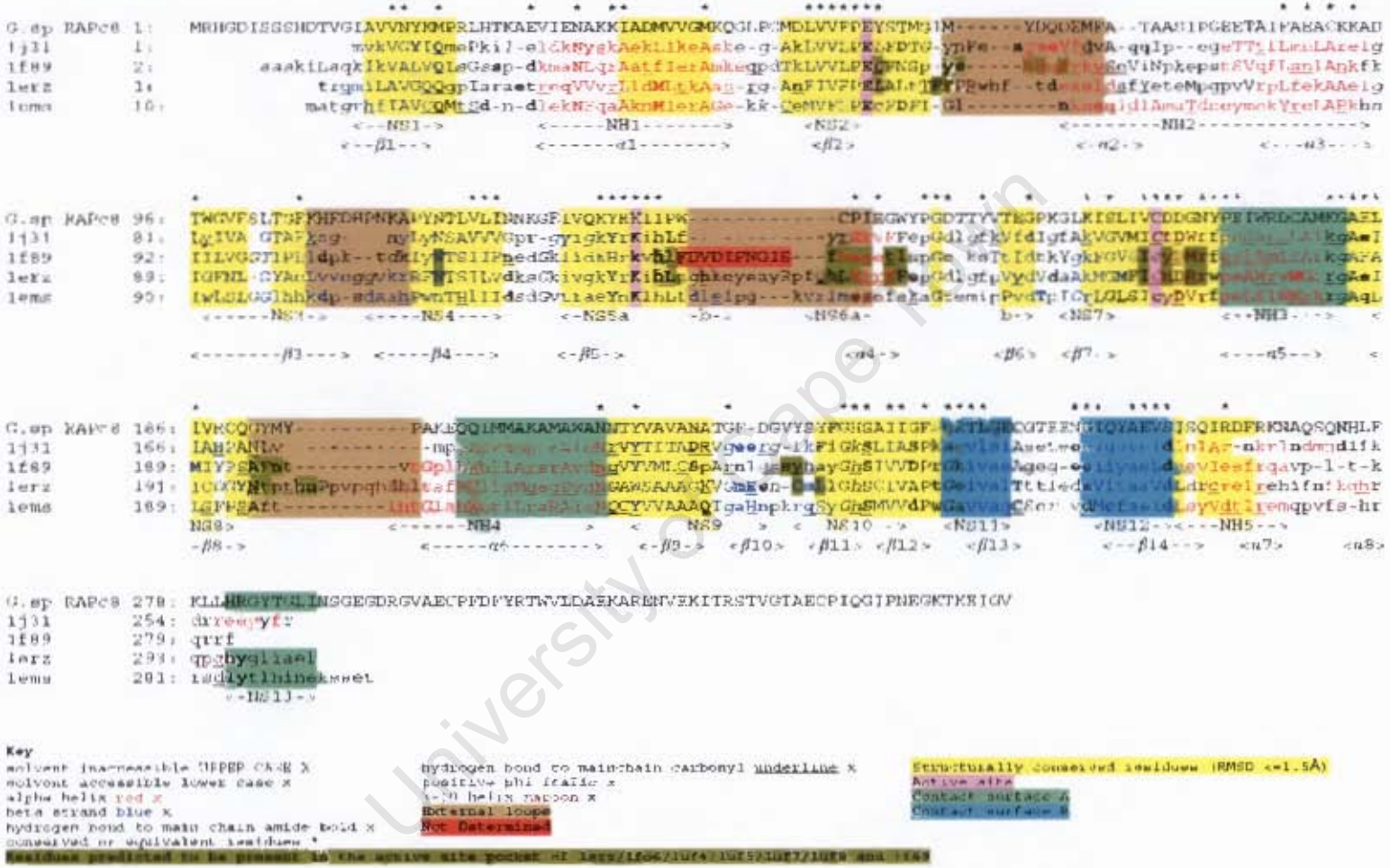


Figure 3-16: Structural alignment of *Geobacillus pallidus* RAPc8 aliphatic amidase against the four crystal structures of distant homologues in the nitrilase superfamily. The alignments are presented in JOY notation (Mizuguchi et al., 1998). The secondary structure elements of 1ems ((NH1-NH5 for helices; NS1-NS13 for β strands) (Pace et al., 2000) and 1erz (β1-β14 for β strands and α1-α8 for helices) (Nakai et al., 2000) are annotated below the alignment respectively. The conserved catalytic triad residues (Glu, Lys, Cys) are highlighted in pink. The alignment is based on predictions using *GenTHREADER* (Jones, 1999) and the pair-wise alignment of the atomic structures using *ALIGN* (Cohen, 1997).

3.6.2. Rotational and translational molecular replacement searches

The MR searches were likely to be very difficult and/or unsuccessful due to the following reasons:

1. Since the success of MR is largely dependent on the sequence identity between the search model and the target protein (Hazes, 2003), the low similarity (19.2% on average) between RAPc8 amidase and the nitrilase homologues was clearly a predicted bottle-neck. Experimental phasing would have been useful in this case, but attempts to soak the crystals or to co-crystallize with a number of heavy metals (mercury, platinum, uranium) introduced disorder in the crystals rendering them non-diffracting. Selenomethionine-containing RAPc8 amidase protein on the other hand did not crystallize. MR was therefore the only method of choice. It was hoped that, the high quality dataset, to 1.9 Å, would partly compensate for the low sequence identity.

2. The complexity provided by the high symmetry of the space group was another major problem. With a cubic unit cell, which has 24 copies of the asymmetric unit, a relatively complete model represents only $1/24^{th}$ of the atoms in the unit cell (Hazes, 2003). In addition to this, the number of unique reflections relative to the total number of reflections is small in a cubic space group. These two factors were likely to present difficulties in obtaining the rotation function. Further more, the rotation function is more compact in cases of high symmetry, resulting in increased noise levels in the rotational searches. Thus a problem of reduced signal-to-noise ratio was predicted for the RAPc8 amidase.

3. The extent to which the four homologous structures represented RAPc8 amidase (in terms of completeness) was likely to further complicate the MR searches. RAPc8 amidase had 348 residues whereas the homologues had 240 residues on average; representing only 70% of RAPc8 amidase sequence. Most of the missing residues were at the C-terminus.

Most of the existing MR programs perform both rotational and translational searches in an automated manner; but the parameter settings can also be changed in difficult cases to increase the sensitivity of MR. MR searches were attempted using the Patterson-based programs *MOLREP* (CCP4, 1994; Vagin, Teplyako,1997), *AMoRe* (CCP4, 1994; Navaza, 1994) and *REPLACE* (Tong, 1993), as well as the evolutionary 6D search program *EPMR* (Kissinger et al., 1999). No MR solution was obtained in any of the attempted runs (data not shown). Therefore a program with a different approach to MR searches was considered.

PHASER (CCP4, 1994; Read, 2001) is a general phasing program that performs likelihood-based MR utilizing the rotational likelihood function (Storoni et al., 2004) to orient the model and then the translational likelihood function (McCoy et al., 2005) to position the model in the unit cell. The maximum likelihood approach statistically measures the agreement between the model and the observed data, with the best model being the one with the highest probability of giving rise to the observed data (Read, 2001). The quality of the search model (as determined from the correlations between sequence identity and the coordinates' errors (Read, 2001)) is accounted for in the likelihood function, making likelihood-based scores more sensitive and accurate (with improved signal-to-noise ratio) than traditional Patterson-based targets, particularly in cases of low homology and multiple molecules in the unit cell. The increased sensitivity of MR using likelihood targets has been reported (Read, 2001; Yano et al., 2000) to allow the solution of difficult structures, which could not be solved by other methods. Due to these advantages, *PHASER* was considered to be an ideal program for attempting RAPc8 amidase MR searches, in the light of the low sequence identity, the incompleteness of the models and the high degree of symmetry in the unit cell; which may have presented difficulties for the other programs.

The following strategies were employed using the four nitrilase homologues to search for RAPc8 amidase MR solution using *PHASER* (CCP4, 1994; Read, 2001):

1. With the default settings of *PHASER* Automated Molecular Replacement mode (MR_AUTO), and with each of the four homologues having all their side chains and the external loops intact, MR was attempted with the RAPc8

amidase data to resolution limit (1.9 Å). Alteration of adjustable search parameters was not attempted at this stage as likelihood-based targets optimize the MR searches based on the structure factors and the quality of the models, which is deduced from the declared sequence identity (Read, 2001). None of the models was successful in determining the MR solution for RAPc8 crystal; all the searches were very noisy, yielding a series of possible 6D solutions that were indistinguishable (data not shown). The Log Likelihood Gain (LLG), which measures how well the data can be predicted with the model, was very low and in most cases even negative, pointing to the low sequence identity between RAPc8 amidase and the four homologues. Comparing the performance of the four homologues as judged from the LLG values, 1j31 was found to be slightly better than the other three models. This could be attributed to the slightly higher sequence identity (21.4%) between 1j31 and RAPc8 amidase.

2. All the search models (intact with side chains and loops) were used simultaneously (as an ensemble) for MR searches with the MR_AUTO mode in *PHASER*. This was to decrease the weighting of the least conserved regions between various models (Figure 3-15 above) while enhancing the conserved regions. This strategy was also not successful in obtaining a clear MR solution (data not shown).

3. As an alternative to strategy number 2 above, the least conserved external loops (Figure 3-14) were trimmed from each model, leaving only the conserved core. The models were used separately as well as all simultaneously as an ensemble for MR searches using MR_AUTO mode in *PHASER*. As the sensitivity of the MR runs may have been further affected by the reduced completeness of the models (due to the removal of residues in the least conserved regions), the searches were even noisier than in case 2 above, with no MR solution obtained (data not shown).

4. The side chains of all four homologues were mutated and replaced with those of RAPc8 amidase using *SCWRL* (Canutescu et al., 2003), so that the models had the same sequence as the target RAPc8 amidase protein. This was in an

attempt to minimize the overall differences between the homologues and RAPc8 amidase protein in the crystal, hence improving the quality of the searches. MR searches were attempted both with separate models as well as with the ensemble. No clear MR solution was obtained (data not shown) with any of the four models nor with the ensemble.

5. Mixed poly-alanine models (where all non-identical side chains longer than alanine were replaced with alanines) were generated for the four homologues, first by alignment of the RAPc8 amidase with the homologues using a sophisticated FFAS alignment protocol as suggested by Schwarzenbacher and colleagues (2004), followed by mutation of non-identical side chains to alanines using *CHAINSAW* (Schwarzenbacher et al., 2004). This was an attempt to alleviate the problem of low sequence identity by minimizing the differences between the target protein and the homologues. MR searches were performed with these mixed models from each homologue. None of the searches was successful in obtaining a clear MR solution (data not shown). Judging from the LLG values, mixed poly-Alanine models were better than any other models that were used in the searches.

6. Although in maximum likelihood MR approaches data at high resolution are automatically down-weighted where necessary (Read, 2001), it became necessary to optimize the MR searches by using data at different resolution ranges. Very low resolution data ($d > 20$) were omitted during the searches to minimize solvent-related systematic errors that are stronger at low resolution, due to the differences in solvent environment between the model (in a vacuum) and the crystal (in solvent) (Hazes, 2003). High resolution data ($d < 3$) were also omitted to minimize the differences between the search models and the target protein. MR searches were performed with different models (models intact with their side chains, conserved hydrophobic cores only, RAPc8 amidase side chains-substituted models and mixed poly-alanine models) with the observed data at varying resolution ranges. However, using data to different resolution cut-offs was not sufficient to allow successful determination of the RAPc8 amidase MR solution (data not shown).

7. The modes of *PHASER* were run separately with all the models from the four homologues (models intact with their side chains, conserved hydrophobic cores only, RAPc8 amidase side chains-substituted models and mixed poly-alanine models); i.e. the rotation function, followed by translation function, rigid body refinement and phasing. Both fast and brute force rotation and translation functions were attempted with several non-default inputs. This strategy was considered in order to control and enhance each step separately, and to monitor the exact stage at which MR was failing. In most cases, the rotation function was very noisy with a very flat peaks' distribution, resulting in numerous numbers of possible 3D solutions, which had extremely low rotation function Z-scores (the number of standard deviations over the mean). This partly demonstrated the effect of poor search models (significantly incomplete, and with coordinates errors arising from low similarities) on the rotation function – which makes worse the generally low signal-to-noise ratio that is associated with rotational searches owing to low correlations between mistranslated search model and the observed data (Kissinger et al., 1999). Performing exhaustive translation searches on all the low signal rotation function solutions gave no clear 6D solution (data not shown). The translation searches were noisy in all cases, with many 6D solutions being rejected by the packing tests. In many cases, the LLG value decreased from the rotational to translational searches; another indication of the poor quality of the search models. As observed in earlier trials, the mixed poly-alanine models were better than the other models.

8. To increase the sensitivity of the MR searches, automated MR runs were performed with all the available models and with different non-default parameters, including: change in resolution limits, change in peak selection criteria particularly during the rotation function searches, enabling or disabling the peaks clustering function before rotation function peaks are re-scored, choosing whether or not to re-score rotation function peaks before final peaks selection, change in the packing criteria to allow a few clashes between α-carbon atoms after the translation function step, and exhaustive searches in all the alternative space groups having the m-3m Laue group. Most of changed parameters were aimed at improving the signal-to-noise ratio in the rotation

function searches in an attempt to make up for the severe effects of the poor search models.

A clear single MR solution (Table 3-5) in the space group $P4_232$ was obtained using a mixed poly-alanine model (comprising 235 residues) of 1j31 with a combination of a number of non-default parameters in *PHASER* that included: (1) using data in the resolution range 3.5-20 Å, (2) generating many rotational orientations to allow for exhaustive searches in the subsequent translation function calculations; this was achieved through the selection of all high peaks without clustering, and by changing the peak selection criteria from the default 75% to 65% of peaks above the mean; and (3) carrying out exhaustive translational searches in all alternative space groups in m-3m Laue group (See section on space group determination).

Table 3-5: Details of RAPc8 amidase 6D MR solution from a successful MR_AUTO run in *PHASER*. RFZ is the RF Z-score, TFZ is the TF Z-score and LLG is the Log Likelihood Gain. A mixed 1j31 poly-alanine model was used with non-default MR search parameters in *PHASER* as explained above. A single solution in space group $P4_232$ was obtained, which had no alpha carbon clashes on unit cell packing, after the translation function step.

Rotation Function (RF)		**Translation Function (TF)**		
RFZ	**RF-LLG**	**TFZ**	**TF-LLG**	**Cα Clashes**
3.0	42	7.2	44	0
6-D solution:	**Orientation (Euler) (°)** : 334.955, 126.380, 8.007			
	Position (Fractional) (Å) : 0.19654, -0.08927, -0.55872			
Space Group:	$P4_232$			

3.6.2.1. Verifying the obtained molecular replacement solution

A repeat of the MR run in *PHASER* (CCP4, 1994; Read, 2001) with the same parameters that obtained the solution above (Table 3-5) but now using the correctly oriented and positioned mixed 1j31 poly-alanine model as a search probe, gave a solution (Table 3-6) that had a translation function Z-score of 8.4. The rotational Euler angles were related by 90° and negligible translations in all the three directions were observed, with zero clashes on the unit cell packing tests. This was a confirmation that the MR solution was correct.

Table 3-6: MR 6-D solution of RAPc8 amidase from a repeat of MR_AUTO *PHASER* run, with the parameters that generated the first solution in Table 3.6 above and the correctly oriented and positioned mixed 1j31 poly-alanine model.

Rotation Function (RF)		**Translation Function (TF)**		
RFZ	**RF-LLG**	**TFZ**	**TF-LLG**	**Cα Clashes**
2.8	42	8.4	44	0
6-D Solution:	**Orientation (Euler) (°)** : 359.944, 89.909, 180.202			
	Position (Fractional) (Å): 0.00018, 0.00018, -0.00012			
Space Group:	$P4_232$			

The MR solution, particularly the rotation function, was tested rigorously using the program *GLRF* of the *REPLACE* suite (Tong, 1993). *GLRF* uses the Patterson method to calculate the rotation function, but has the advantage of allowing the user to choose all the MR search parameters. It provides the option of refining the rotation function solutions using a Patterson correlation (PC) refinement method. The rotation angles from the *PHASER* MR solution were applied to the starting search model, structure factors (F_{calc}) were calculated in a P1 cell, and the fast cross rotation function was re-calculated in *GLRF* using the appropriate search parameters. A single peak (not

shown) was observed at the origin of the β=0 map sections, indicating that the rotation function was correct.

The lack of α-carbon clashes on packing the model in the unit cell was a further confirmation that the translation function solution was correct.

3.6.2.2. Packing of the $P4_232$ unit cell with the MR solution

The ultimate test for the correctness of the obtained MR solution was the consistency of the packed unit cell with the RAPc8 amidase homohexameric complex features as revealed by EM analysis. A packed $P4_232$ unit cell (Figure 3-17(A)) had four hexamers arranged in a tetrahedral manner. The conserved dimer interface that is found in the four nitrilase superfamily distant homologues was indeed preserved by the crystallographic 2-fold symmetry in the MR solution (Figure 3-17(B)). In addition, three dimers packed around the $P4_232$ crystallographic 3-fold axis (Figure 3-17(C)), with D3 point group symmetry. Projection of the model was remarkably similar to the EM 3-fold symmetry projections.

3.6.3. Rigid body refinement and phase angles calculation

Although the MR solution model is normally optimised and the phases calculated using the *PHASER* (CCP4, 1994; Read, 2001) MR_AUTO mode, an independent rigid body refinement and phase calculation step was performed to ensure and confirm the accuracy of the MR solution before an electron density map was calculated. The MR_RNP mode in *PHASER* was used to perform rigid body refinement of the correctly-oriented and positioned mixed 1j31 poly-alanine model against the RAPc8 amidase data to 3.5 Å resolution. The model was found to be optimal as no changes were observed in either its orientation (rotational Euler angles) or its position (translation) in the unit cell. The calculated phases however were extremely poor, with an overall phasing Figure of Merit (FOM, a measure of effectiveness, efficiency and performance) of 0.191.

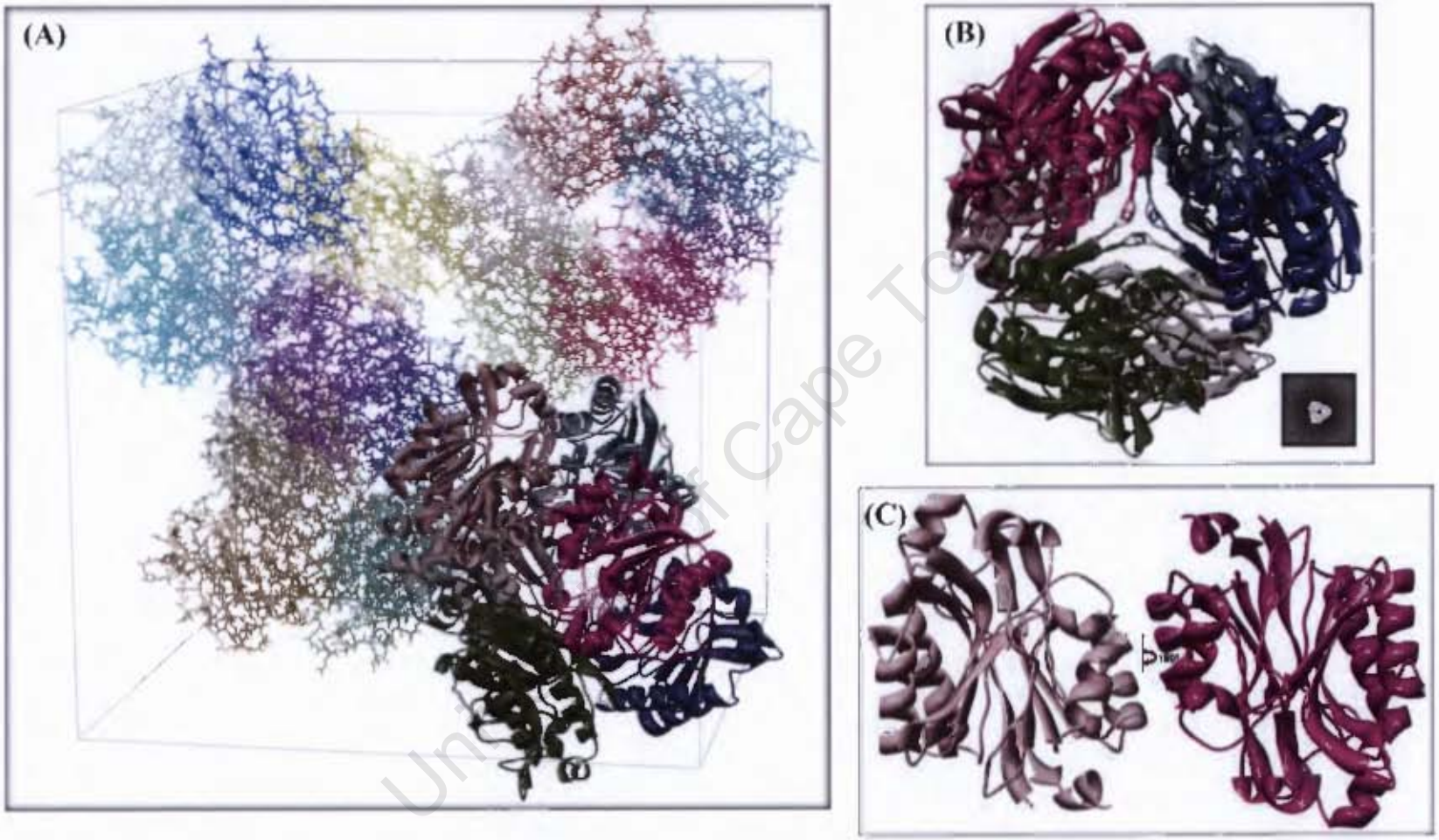


Figure 3-17: Packing of the 1j31 starting model in the $P4_232$ unit cell of RAPc8 amidase: **(A)** $P4_232$ cell with 24 molecules (shown in different colours); one of the hexamers is rendered as cartoons. The four hexamers are arranged in a tetrahedral manner. **(B)** A view of a hexamer along the 3-fold axis; this arrangement and close packing is very similar to the triangular EM projections (inset). **(C)** A view of the conserved dimer interface; monomers are depicted in different colours. These packing features confirm the MR solution to be correct. The images were generated using *UCSF Chimera* (Pettersen *et al.*, 2004).

3.6.3.1. Initial electron density map calculation

A sigma-A weighted $2F_{obs}$-F_{calc} electron density map was generated using the calculated phases (from the correctly oriented and positioned mixed 1j31 poly-alanine model) and structure factor amplitudes from the observed RAPc8 amidase data to 3.5 Å resolution. The resultant map, although noisy, was interpretable in most regions; the density for some of the side chains that had been replaced with alanines in the model had re-appeared (Figure 3-18). Continuous density with considerable amount of noisy was also observed on the C-terminal side where over 80 residues were missing in the model. Despite the poor phase information, the emergence of features missing in the model, in the electron density map was a further confirmation that the MR solution was correct.

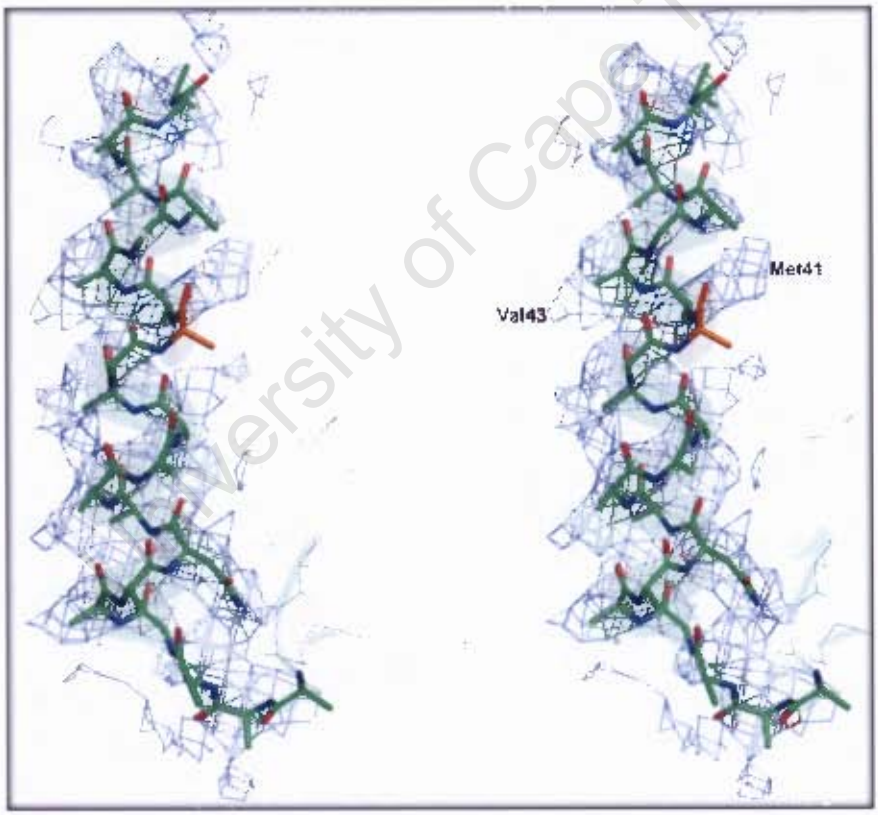


Figure 3-18: A stereo view of a portion of the initial $2F_{obs}$ - F_{calc} electron density map generated with mixed 1j31-poly-alanine phases and RAPc8 amidase structure factor amplitudes, calculated at 3.5 Å and contoured at 0.8σ. The mixed 1j31 poly-alanine model is displayed in the density. Though the map has a considerable amount of noise, density for some of the side chains that had been replaced with alanines is visible. An example is the Met41 side chain (orange, with labelled density). The picture was rendered using *PyMOL* (Delano, 2004).

3.7. Model rebuilding and refinement

Model rebuilding and refinement is an interactive and iterative process that aims at improving the agreement between the atomic model and the observed data (Pannu, Read, 1996). While model rebuilding focuses on fitting the model to the electron density map, either manually or in a semi-automated manner, refinement minimizes the difference between the atomic model and the observed data by making small movements of the model atoms, which ultimately result in the improvement of the fit between the model and the electron density.

Major differences existed between the starting atomic model and the observed RAPc8 amidase data; aside from being only 68% complete, the mixed 1j31 poly-alanine model had over 75% of all the side chains mutated to alanines. Although the initial electron density map was interpretable in most regions, a significant amount of bias towards the starting model bias could not be ruled out. The process of model rebuilding and refinement was therefore bound to be difficult and complicated with model-bias creating a bottle-neck. The graphical program *O* (Jones et al., 1990) was used for manual rebuilding and fitting of the model to the electron density while *REFMAC5* (CCP4, 1994; Murshudov et al., 1997) was used for restrained refinement. *REFMAC5* uses maximum likelihood targets that take into account errors and uncertainties (standard deviations) in both the experimental data and the atomic model, as well any prior phase information (Murshudov et al., 1997), to minimize the differences between the observed data and the atomic model. The following strategies were employed in the refinement of RAPc8 amidase structure.

1. The alanines in the mixed 1j31 poly-alanine MR solution were replaced with RAPc8 amidase side chains using *SCWRL* (Canutescu et al., 2003), to generate a model that had the same sequence as the targeted RAPc8 amidase structure. The resultant model was used to calculate phases which together with observed RAPc8 amidase data were used to generate a sigma-A weighted $2F_{obs}–F_{calc}$ electron density map to 3.5 Å resolution. The generated map was heavily biased towards the model; the electron density almost perfectly

matched the model, with no new features observed. The model-biased map could not allow for further model rebuilding and refinement. This clearly emphasises the fact that phases have more structural information than the structure factor amplitudes (Taylor, 2003).

2. A sigma-A weighted $2F_{obs}–F_{calc}$ electron density map calculated with the phases from the mixed 1j31 poly-alanine model and the measured RAPc8 amidase amplitudes to 3.5 Å resolution, was used for model fitting. The MR solution model whose poly-alanines had been replaced with RAPc8 amidase side chains (as described in number 1 above) was manually rebuilt and fitted to the interpretable regions of the density using *O*, alternating with a few cycles of restrained refinement in *REFMAC5* after each rebuilding cycle. Although the R-factor (measure of the agreement between the model and the observed data) and phasing FOM showed slight improvements with each cycle of rebuilding and refinement, the electron density map did not improve sufficiently to reveal clear density for most of the features that were noisy and less determined in the initial map. With more cycles of rebuilding and refinement, the model's structural elements slowly started to disintegrate (Figure 3-19). This was attributed to the possibility of over-fitting the model to electron density features that were not necessarily true (due to starting poor phases), eventually leading to the disruption of the hydrogen-bonding network and other interactions that hold secondary structure elements together. The electron density thus became more and more biased towards the model, reducing the possibility of building a correct model and making further model refinement impossible.

3. Density modification (DM) through solvent flattening using *RESOLVE* (Terwilliger, 2004a) was performed in an attempt to improve the initial electron density map prior to rebuilding. *RESOLVE* uses the principle of maximum likelihood to maximize the total probability of the phases (calculated from the map-probability function as described by Terwilliger in 2001), by combining information from the experimental data and the expected consistency of the map with the flat solvent region (Terwilliger, 1999). Solvent flattening of the initial electron density map (from mixed 1j31 poly-

alanine model phases) was attempted with 50% solvent content (determined from the calculated Matthew's coefficient). Although the phasing FOM showed slight improvement after solvent flattening, the bias ratio (representing the extent to which the map is biased towards the starting model phases) was significantly high (3.91) and the map was even noisier. More runs were performed with gradually increasing levels (in steps of 5%) of solvent content in an attempt to account for regions that were likely to be highly disordered in the crystal. DM did not seem to improve the maps even with as much as 70% solvent. This was most likely an indication that, although the initial phases were relatively accurate, they were likely to be heavily model-biased.

4. In an attempt to reduce the bias introduced by using a single model, phases were calculated from an ensemble of the four homologues; prepared by superimposing (using *ALIGN* (Cohen et al., 1997)) the homologues (intact with their side chains) onto the mixed poly-alanine MR solution from *PHASER*. This strategy was aimed at enhancing the conserved regions in the four homologues while down-weighting the variable loop regions. Rigid body refinement of the ensemble against RAPc8 amidase data to 3.5 Å, resulted in an increase in LLG value from 44 to 56, and very minimal changes (almost negligible) on the rotational and translational parameters. The increase in LLG value was an indication that the ensemble agreed better with the data than a single model. The overall phasing FOM improved from 0.191 (with single mixed poly-alanine model) to 0.306. The resultant electron density map was much improved, and new features could clearly be observed. These new phases were used for model rebuilding using the prime-and-switch phasing technique (Terwilliger, 2004b) as implemented in *RESOLVE* (Terwilliger, 2004a), within *PHENIX* (Adams et al., 2002). In prime-and-switch phasing, the crystallographic phases (which might be model-biased) calculated from the atomic model are used to 'prime' the DM process, which then 'switches' to a map-probability function (Terwilliger, 2001) in order to iteratively adjust and improve the phases so that they are more correct and devoid of the memory of the initial model (Terwilliger, 2004b). The process of phase improvement involves cross-validation procedures (that remove model bias. In the model rebuilding step, the phases calculated from the ensemble as described above,

were used together with the measured amplitudes (between 45 and 1.9 Å) from the RAPe8 amidase crystal and a 1j31 model that contained RAPe8 amidase side chains. 20 iterative cycles of model rebuilding (in-place) with three refinement cycles after each rebuilding cycle and 50% solvent content were performed using prime-and-switch phasing with *PHENIX*. Over 94% of all the residues in the model were rebuilt. Prime-and-switch phasing was successful in significantly reducing model bias; the rebuilt model had atoms shifted significantly (Figure 3-20) to their correct positions, the phases had dramatically improved with an increase in phasing FOM from 0.306 to 0.552 and the electron density map (Figure 3-21) was clear with interpretable density for most of the residues that were absent in the starting model.

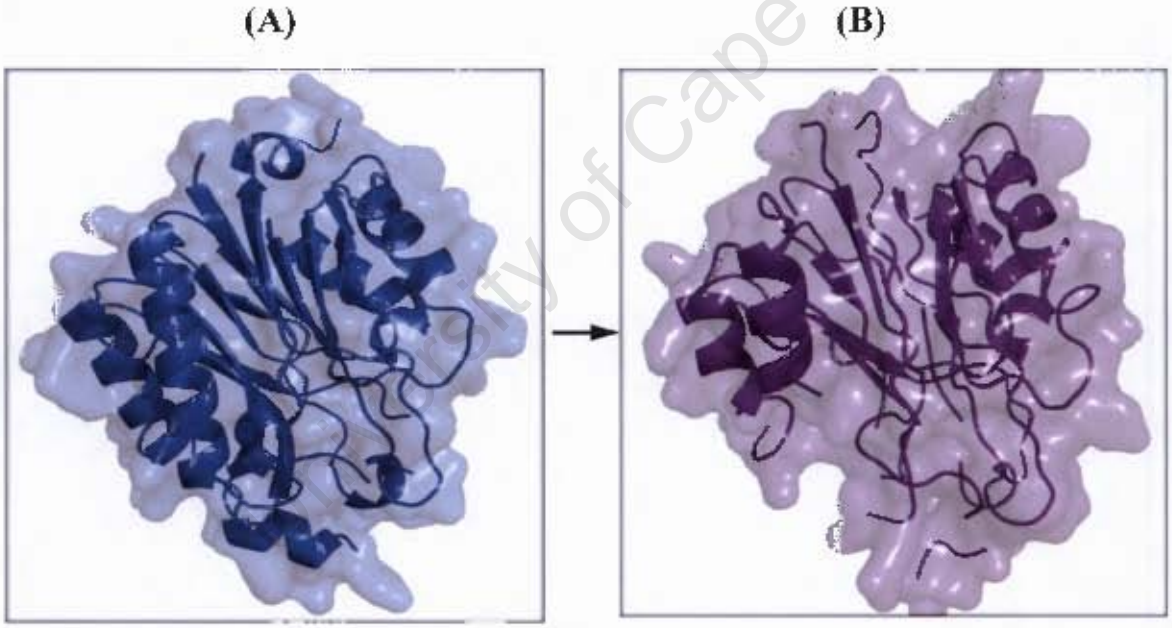


Figure 3-19: Cartoon and transparent surface representations of the models: (A), before rebuilding and refinement and (B), after several cycles of rebuilding in *O* and restrained refinement in *REFMAC5*. The model structural elements slowly disintegrated (helices and beta sheets breaking up) with continued fitting, rebuilding and refinement. Images rendered in *PyMOL* (Delano, 2004).

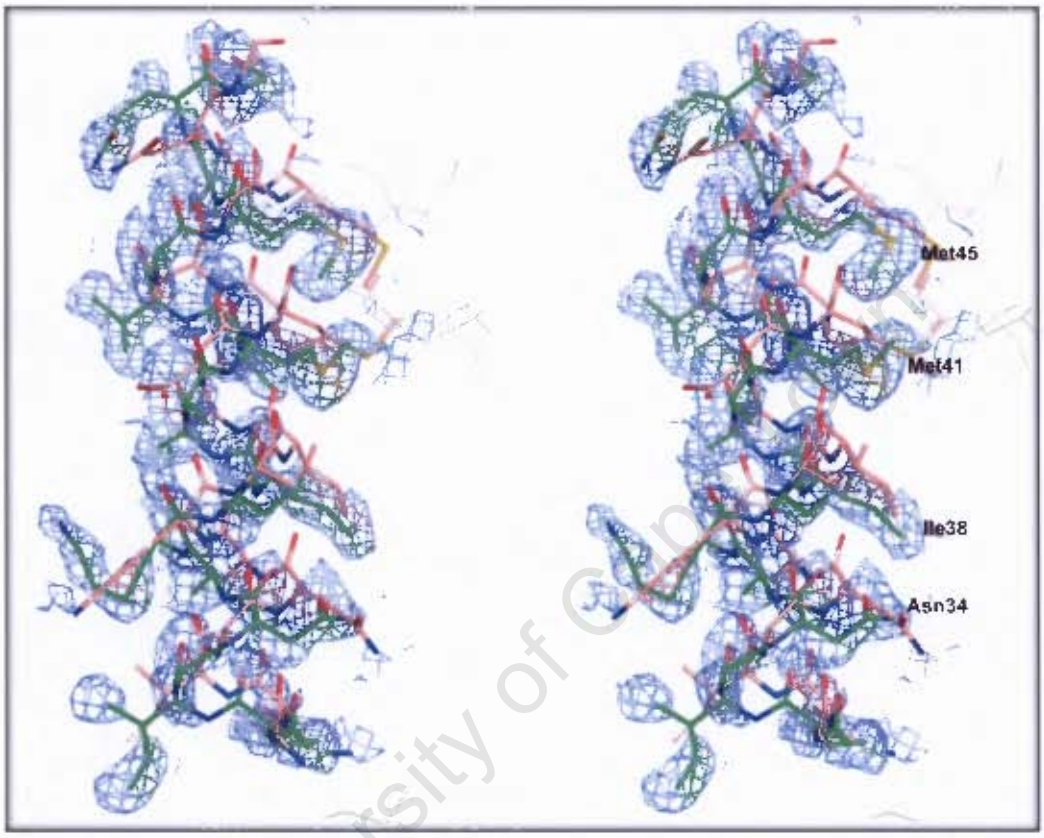


Figure 3-21: A stereo view of a portion of a simple map generated with phases after the rebuilding step with *PHENIX*, calculated at 1.9 Å and contoured at 1.2 σ. The 1j31 model substituted with RAPc8 amidase side chains before rebuilding is shown in pink sticks while in green sticks is the model after rebuilding with *PHENIX*. The rebuilt model is greatly shifted and the map features are quite clear, with reduced bias towards the starting model. Met41 (labelled) is now in well-defined density; and so are the other amino acids in that portion of the map. The picture was rendered using *PyMOL* (Delano, 2004).

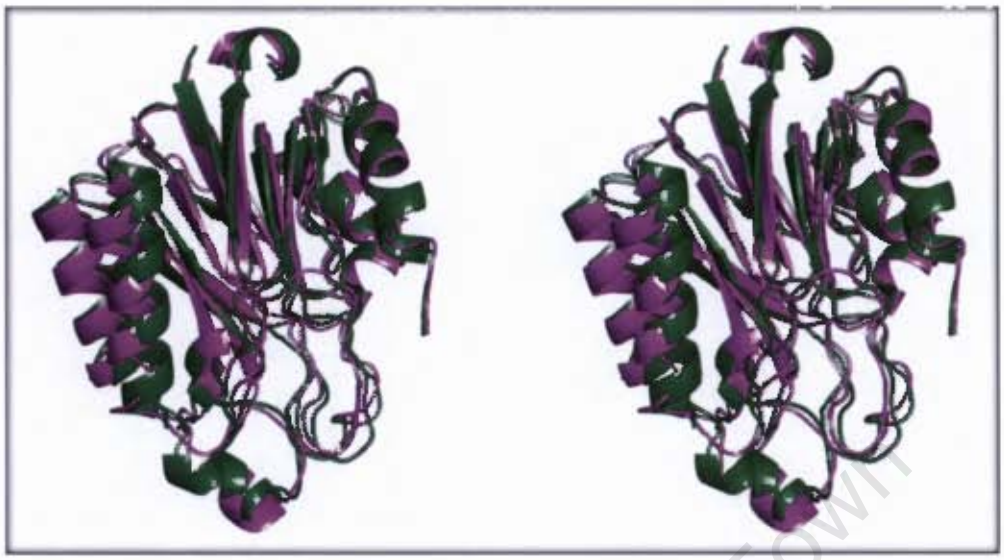


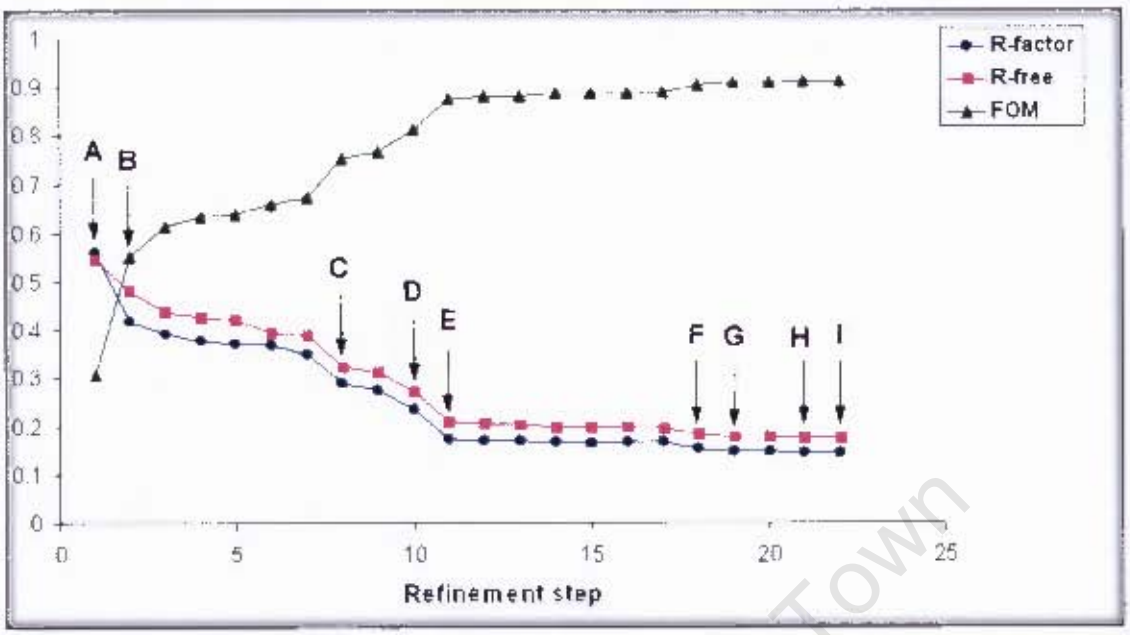
Figure 3-20: Stereo view of cartoon representation of superimposed models: In green, before and in purple, after prime-and-switch rebuilding in *PHENIX*. The atoms of the model were shifted significantly to their correct positions during the rebuilding step, helping to reduce bias from the starting model. The structural fold was preserved. The image was rendered using *PyMOL* (Delano, 2004).

Following successful model rebuilding, reduction of initial model bias and improvement of starting phase information, the iterative process of manual building of over a hundred missing residues, rebuilding and refinement proceeded smoothly without further complications. As the model became more and more accurate and complete, the phase accuracy improved dramatically, resulting in improved electron density maps and better fit. Difference maps (F_{obs}-F_{calc}), which highlight the differences between the current model and the actual structure) allowed the regions that had errors to be identified and rectified by rebuilding in *O* (Jones et al., 1990). To improve the accuracy of the regions that had some level of uncertainty (including regions where hydrophilic residues were facing bulk solvent and those with alternative locations or conformations), the affected residues were omitted during the refinement, resulting in improved 'omit' maps; allowing unbiased fitting and rebuilding of problematic residues. Water molecules peaks were searched automatically using the program *ARP/WARP* (CCP4, 1994; Lamzin et al., 2001) and

rationalised manually to ensure availability of possible hydrogen bonding network. As model refinement proceeded, more water molecules were incorporated in the structure, improving the quality and accuracy of the model each time. The model was manually inspected and analysed after every cycle of refinement.

The progress of the refinement process as well as the overall accuracy of the atomic model was judged through the crystallographic R-factor (which represents the average fractional error in the model of the molecular structure compared to the observed diffraction data from which it has been derived from). A free R-factor (R-free) was calculated from 5% cross-validation data (test set) that was omitted from the refinement process. R-free was introduced by Brünger (1992) and it represents a statistical measure that is essentially free from the artificial lowering that can be caused by over-fitting. Figure 3-22 below shows the behaviour of crystallographic R-factor, R-free and phasing FOM during the rebuilding and refinement of RAPc8 amidase structure model. Since the crystallographic R-factor is only a global indicator of model quality (Kleywegt, 2000) (and does not highlight local errors in the structure), structure validation programs as well as manual evaluation of the fit of the model to the electron density were used as ultimate checks for problematic local regions in the model, especially towards the end of the refinement. Refinement was assumed to have converged globally when all the density belonging to the structural features had been accounted for, the difference maps were featureless and parameter changes in the model were negligibly small in further refinement cycles.

A portion of the electron density map at the end of the refinement, with the refined atomic model fitted in is shown in Figure 3-23 below.



Key:

Step	Action
A	1j31 MR solution - substituted with RAPc8 amidase side chains, before rebuilding in *PHENIX*
B	After prime-and-switch phasing and rebuilding in *PHENIX*
C	On correction of major errors as well as building and rebuilding residues 1-308
D	All 340 residues built
E	137 water molecules added automatically with *ARP/WARP*
F	100 more water molecules added manually; most errors corrected
G	Residues with alternative locations were built and 34 water molecules added
H	Most of the detectable errors corrected, and all justifiable waters added
I	Refinement converges

Figure 3-22: Progress in rebuilding and restrained refinement of RAPc8 amidase structure. R-factor is shown in blue circles, R-free in purple squares and phasing FOM in green triangles. As the model agreed better with the observed data, the crystallographic R-factor and R-free dropped consistently, while the phasing FOM increased in a similar but opposite trend. Refinement convergence was reached when the model was not improving much further and the refinement statistics had reached a plateau. Rebuilding the model in *PHENIX* caused a large increase in the phasing FOM as a result of improved accuracy in the phases.

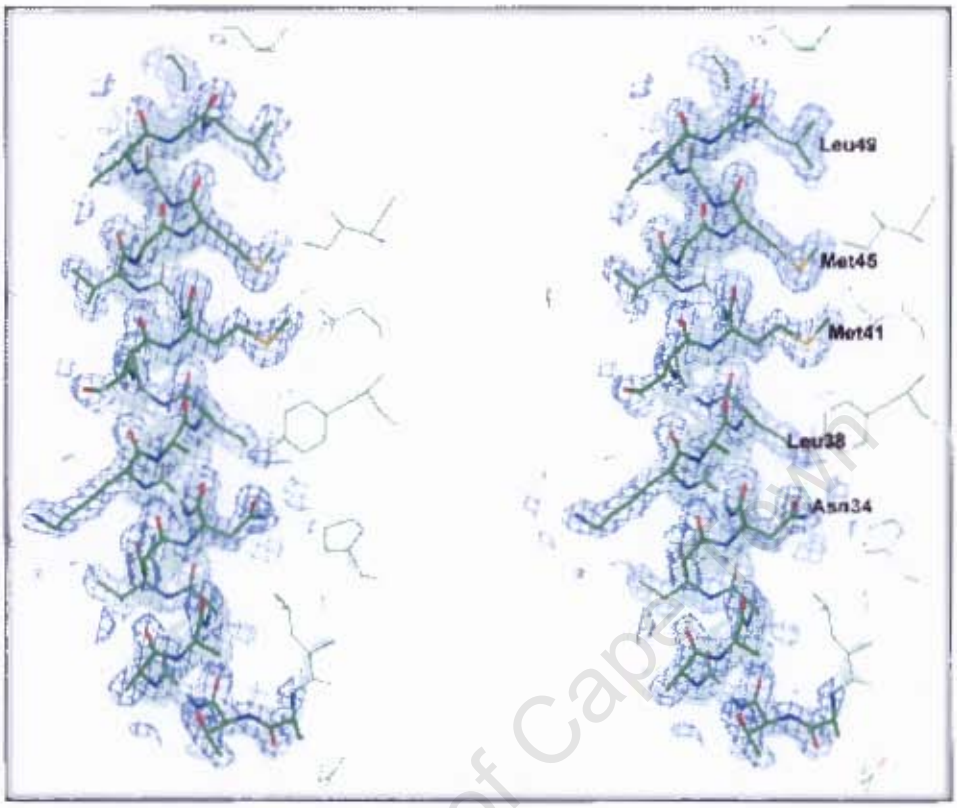


Figure 3-23: A stereo view of a portion of a $2F_{obs}$-F_{calc} map at the end of the refinement calculated at 1.9 Å and contoured at 1.3 σ. The refined model is displayed in the density. The picture was rendered using *PyMOL* (Delano, 2004).

The final model had 1-340 out of 348 (from the theoretically translated sequence) residues. The 8 residues that were missing on the C-terminus were likely to have been absent from the protein during crystallization, as opposed to them being disordered; as there was no density at all for these residues. The model was of good quality as can be observed from the refinement statistics presented in Table 3-7 below. The relatively high resolution data (to 1.9 Å) allowed for the location of most of the protein chain-associated water molecules, resulting in a total of 304 waters in the final structure.

Table 3-7: Refinement statistics of RAPc8 amidase model.

Resolution range (Å)	46.13 – 1.90
Number of non-hydrogen atoms	
Protein	2673
Water	304
Number of reflections	
Working set	28752
Test set (free)	1501
R-factor[§]/R-free[†] (%)	14.50/17.50
Rms deviations from ideality	
Bond lengths (Å)	0.01
Bond angles (°)	1.31
Average B value (Å^2)	10.56

[§]R-factor = $\sum||F_o| - |F_c|| / \sum|F_o|$ where $|F_o|$ and $|F_c|$ are observed and calculated structure factor amplitudes respectively. [†]R-free is similar to R-factor, but it was calculated using a random set containing 5% of observations, which were omitted during refinement.

A small difference between R-factor and R-free (2.5%) with R-factor/R-free ratio (Tickle et al., 1998) of 0.83 is indicative of a good agreement between the model and the observed data and a sign of very minimal over-fitting of the data.

Analysis of average B values, residue by residue (Figure 3-24) using *BAVERAGE* (CCP4, 1994) showed no large variations in B values, which otherwise would have pointed to either highly flexible regions or to error-prone regions of the model. The RAPc8 amidase structure is relatively stable overall, with the highly flexible loop regions being well fixed by crystal packing.

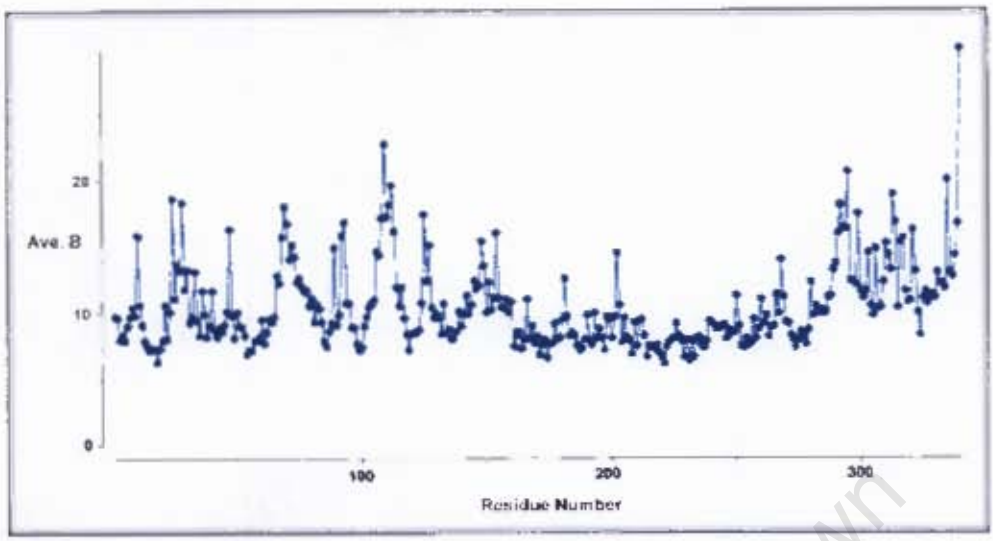


Figure 3-24: An all atom plot of average B value per residue. Apart from the C-terminal Glu340, which had an average B value of 29.5 $Å^2$, the rest of the residues had relatively low average B factors, and indication of a relatively stable overall structure.

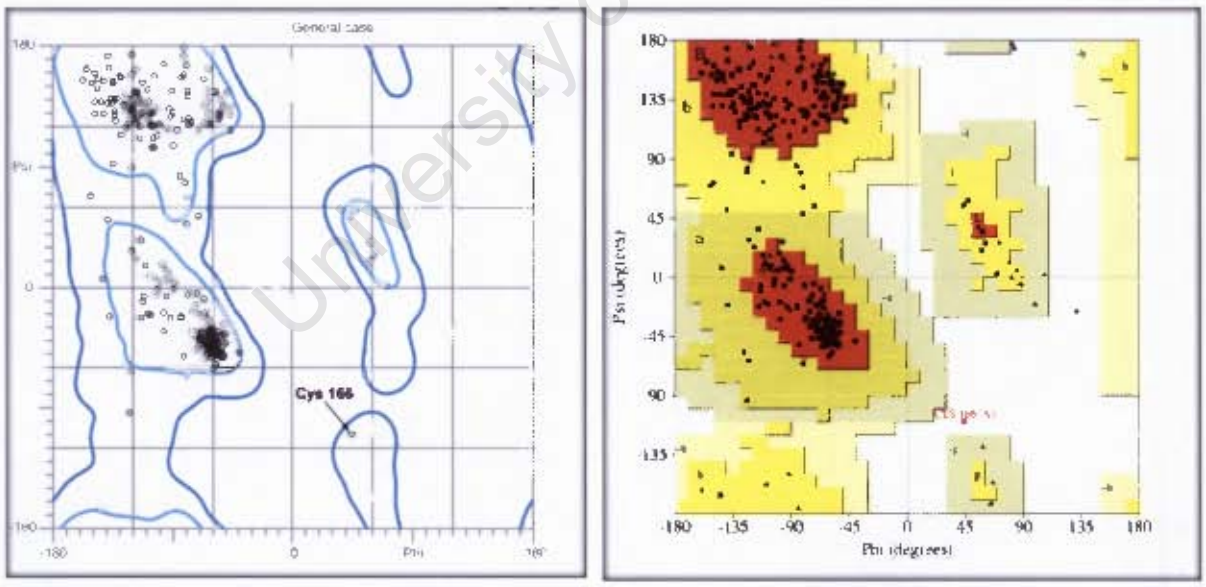
3.8. Final Model Validation

Aside from the global indicators of the quality model indicators; R-factor and R-free (that measure the agreement between the model and the data), the overall and local accuracy of the final model is usually assessed by various validation methods. These methods utilize aspects of the model that are independent of the information that is optimised during model rebuilding and refinement (Kleywegt, 2000; Richardson, 2003). Torsion angles for instance, are not optimised by the refinement process, thus they form the basis for most geometric validation methods. In most cases, validation approaches are entirely model-based and they utilize information derived from highly accurate structures from the databases, to identify local outlier regions that deviate from ideality (Richardson, 2003).

The program *PROCHECK* (CCP4, 1994; Laskowski et al., 1993) was used to assess the molecular geometry, particularly the backbone (φ and ψ angles) and the side chains (mainly side chain torsion angles) conformations. The program *WHATCHECK* (Hooft et al., 1996) was used for a number of geometric checks, analysis of residues environment and assessment of hydrogen bonding network. *MOLPROBITY* (Lovell et al., 2003) was used for geometric validation, particularly in analysing the complete α-carbon geometry (Lovell et al., 2003), including backbone φ and ψ angles and the deviation of the β-carbon positions (that provide information on the backbone-side chain incompatibilities). *MOLPROBITY* generates accurate φ, ψ Ramachandran plots (Ramachandran et al., 1963; Laskowski et al., 1993; Lovell et al., 2003), with redefined allowed and disallowed regions, based on current information from highly accurate structures (Lovell et al., 2003). It performs all-atom contacts analysis (Word et al., 1999), providing information on hydrogen bonding network, van der Waals contacts, and unfavourable atomic clashes thus highlighting local structural inaccuracies (Lovell et al., 2003). Table 3-8 provides a summary of the identified outliers in the RAPc8 amidase final model.

Table 3-8: Outliers in the refined RAPc8 amidase structure

Pairs of residues with atomic clashes	(12; 1 clash per pair)
Ramachandran outliers	0
Rotamer outliers	6
Cβ outliers	0
Residues with distorted geometry	1
Residues with unusual backbone conformations	30
Residues in abnormal packing environment	9
Buried unsatisfied hydrogen bond donors and acceptors	4



(A) **(B)**

Figure 3-25: Backbone φ, ψ angles distribution for the refined RAPc8 amidase structure: **(A)** A general case for non-glysine, non-proline and non-preproline residues, generated using *MOLPROBITY*. **(B)** An all residues case; generated with *PROCHECK*.

The Ramachandran plot (Figure 3-25) statistics from *PROCHECK* (CCP4, 1994; Laskowski et al., 1993) had 99.6% of all residues in the favoured (90%) and allowed (9.6%) regions, and 0.3% (1 residue, cysteine 166) in the disallowed region. Cysteine 166 is found in the allowed but rare region (type II'-turn conformation) of the *MOLPROBITY* (Lovell et al., 2003) Ramachandran plot. In the recently defined type II'-tight turns conformation, the second of the four residues in a turn adopts φ, ψ angles near 50°, -125°, resulting in steric overlap between the β-carbon and carbonyl oxygen. This conformation is common in the active sites of various protein families, including the serine of the catalytic triad of all α/β hydrolases and lipases (Uppenberg et al., 1994). The cysteine of the Glu, Lys, Cys catalytic triad in the four nitrilase superfamily structures is also found in the type II'-turn conformation region of the Ramachandran plot, indicating a functional role in mediating activity in these enzymes.

No β-carbon position outliers (Table 3-8) were found in the RAPc8 amidase structure, which indicates good compatibility between the backbone and the side chains. However, some unusual features were observed (Table 3-8) despite clear interpretable density in the highlighted regions. The 12 atomic clashes involved hydrogen atoms of neighbouring residues. Most of the overlapping hydrogen atoms are found in the tightly constrained active site, while a number of clashes involve residues in sharp bends. Of the 6 residues that had unusual side chain rotamers, 2 (Tyr192 and Met193) are in the highly constrained active site, 3 are in the tightly packed environment while the last one is the Glu340 residue that is facing the bulk solvent.

Only one residue had distorted geometry; a cis-conformation proline 195, that is found in a tight bend. Although the environment could partly explain the constraint on this residue, any efforts to improve the geometry (including simulated annealing and molecular dynamics) were unsuccessful. Of the four buried residues that had deficiencies in hydrogen-bonding network, two are neighbouring cysteine residues possibly forming weak cysteine hydrogen bonds, one is in the active site; while one is in a bend, where it is likely to be stabilised by neighbouring hydrogen bonding environment. The residues in the abnormal environment and those with unusual backbone conformations, are mostly found in the bends, some are on the edges, a

number are in the highly constrained active site and a few are on the interfaces, interacting with symmetry-related molecules.

Overall, the refined RAPc8 amidase model was of high quality, showing reasonably good agreement with the high quality diffraction data and minimal errors.

3.9. Analysis of the RAPc8 amidase structure

(A) - *Geobacillus pallidus* RAPc8 amidase has an α-β-β-α sandwich fold

The monomer of RAPc8 amidase is an α/β globular protein that consists of 11 helices and 14 β strands, interconnected with external loops. A cartoon representation of a RAPc8 amidase monomer, and the topology of β sheets and α helices are shown in Figure 3-26 below. Helices α1, α2 and α3 form one outer helical layer while helices α4, α5 and α6 form the other. β strands β1, β2, β3, β4, β5, and β14 form one β sheet while β strands β6, β7, β8, β9, β12 and β13 form the other sheet. The two six-stranded β sheets are sandwiched between the two helical layers to form a four-layer α-β-β-α sandwich architecture (Orengo et al., 1997). The two β sheet layers have a similar alternation of parallel and anti-parallel β strands; the three central β strands of each β sheet (β1, β2, β3 and β7, β8, β9) form a parallel β sheet, while the rest of the strands are arranged in an anti-parallel configuration. The interface between the two β sheets is tightly packed with hydrophobic residues. Helices α8, α9, α10, and α11 are found on the extended C-terminal tail, while β strands β10 and β11 are on a long external loop that links β9 and β12. Four residues form a 3-10 helix (shown in cyan on the topology diagram) at the active site, between β7 and α5. The two α/β halves are connected by two loops, one at each end; a 20 residues loop between β5 and β6 on one end and an 8 residues loop between β13 and β14 on the other end.

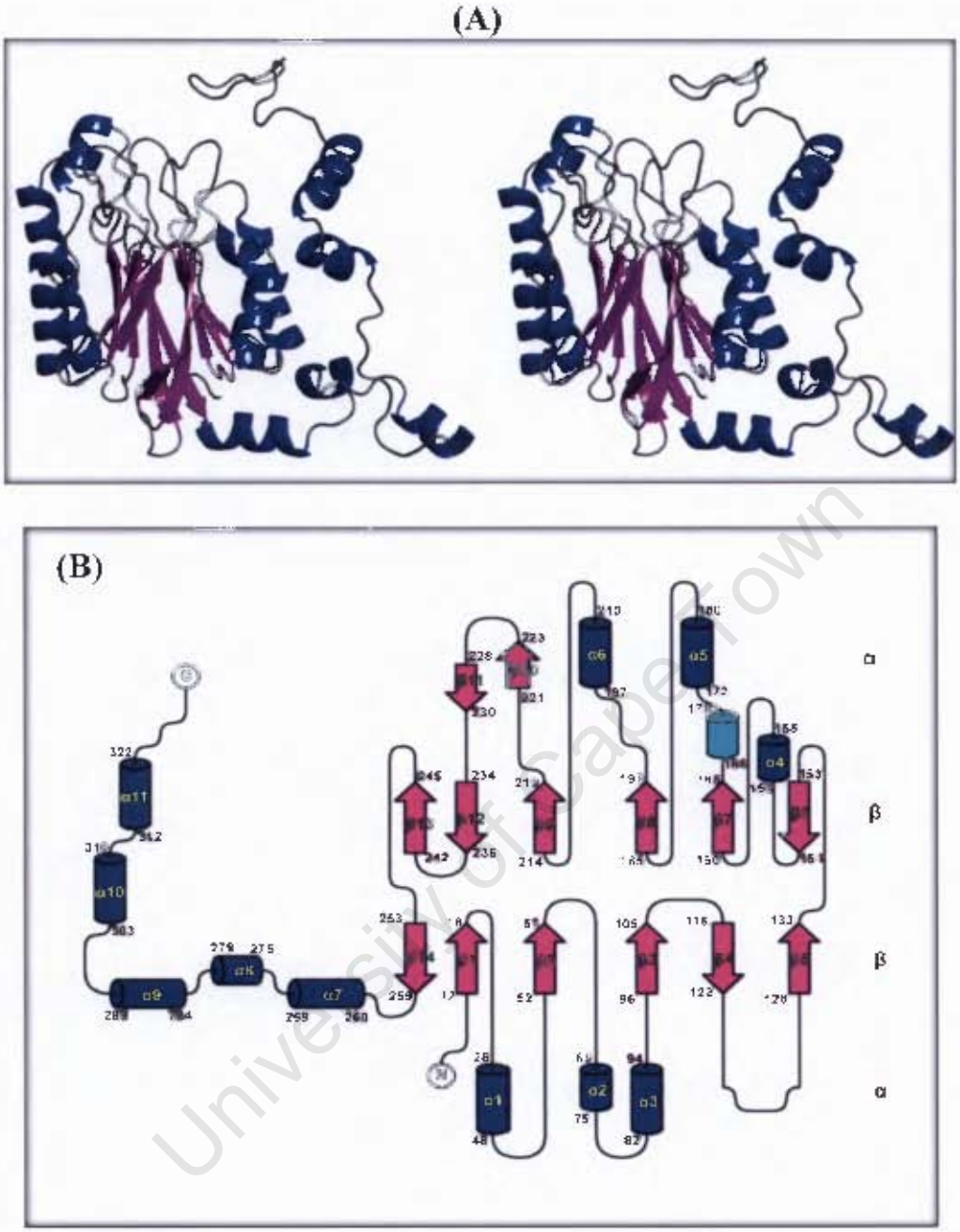


Figure 3-26: The structural fold of RAPc8 amidase: (A) A stereo view of a cartoon representation of RAPc8 amidase monomer. (B) α-helix and β-sheet topology of RAPc8 amidase. β-sheets are shown in purple, while α-helices are in blue, in both pictures. The cyan cylinder in the topology diagram is a 3-10 helix. The secondary structure elements are labelled and numbered accordingly from the N-terminal side in the topology diagram. RAPc8 amidase topology was generated in the program *TOPS* (Westhead *et. al.*, 1999) and manually redrawn in *TopDraw* (Bond, 2003) for simplification. The cartoon was rendered using *PyMOL* (Delano, 2004).

(B) - *Geobacillus pallidus* RAPc8 amidase is a member of the nitrilase superfamily

There are 14 existing topology classes under the α/β class with four-layer sandwich architecture (3.6.0) in the *CATH* (class, architecture, topology, homologous superfamily) database (Orengo et al., 1997). These topologies are found in a number of enzyme families, including hydrolases, lyases, ligases, transferases, oxidoreductases, and isomerases. A structure similarity search of the *DALI* database (Holm & Sander, 1993) with RAPc8 amidase identified the four nitrilase superfamily structures (PDB IDs; 1j31, 1f89, 1erz & 1ems) with high structural similarity (average Z-score of 32.0). The structural similarity with other reported four-layer sandwich proteins was poor. This was a further confirmation that RAPc8 amidase is a C-N hydrolase belonging to the nitrilase superfamily. The fold similarity between RAPc8 amidase and putative C-N Hydrolase from yeast (1f89; Kumaran et al., 2003) is shown in Figure 3-27 below. Although the two structures are only 20.7% similar in sequence, they share a common four-layer sandwich architecture; the secondary structure element topology is highly similar in the conserved core with only a few minor differences such as the position of α4 helix (which is between β5 and β6 in 1f89 and links the two α/β halves while in RAPc8 amidase, α4 helix is between β6 and β7) and the insertion of the two small β strands (β10 and β11) on the external loop of RAPc8 amidase. Outside the α-β-β-α conserved core, RAPc8 has an extended C-terminal tail that is completely helical. The same topology similarity and structural fold are shared with 1j31, 1erz and the Nit domain of 1ems as shown in Figure 3-28 below.

(C) - *Geobacillus pallidus* RAPc8 amidase is homohexameric is solution

Gel filtration chromatography and electron microscopy (EM) had indicated that RAPc8 amidase is homohexameric in solution. These findings have been confirmed by the crystal structure, where RAPc8 amidase exists as an associated homohexamer (Figure 3-29), comprised of trimers of dimers.

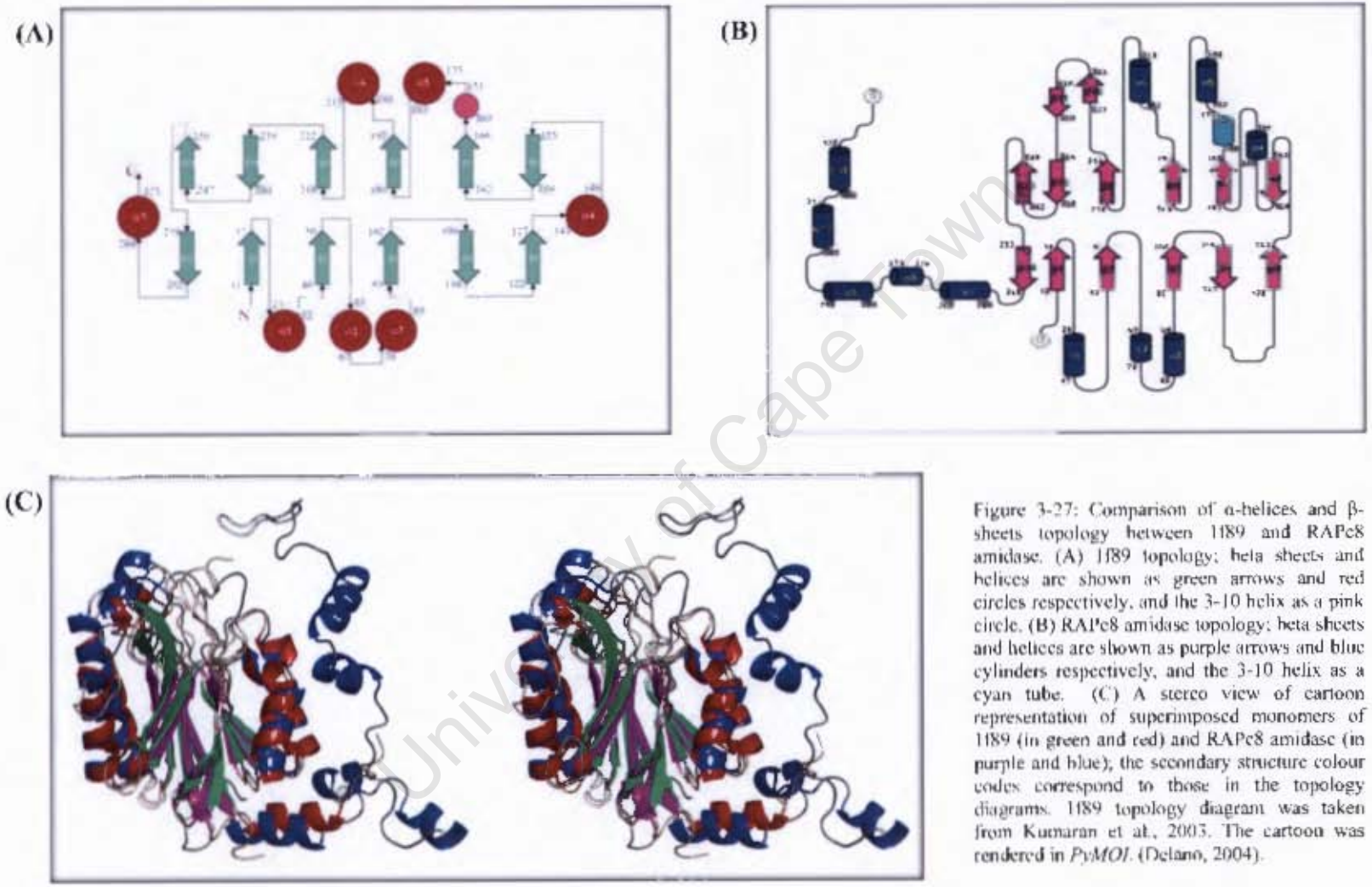


Figure 3-27: Comparison of α-helices and β-sheets topology between 1I89 and RAPc8 amidase. (A) 1I89 topology; beta sheets and helices are shown as green arrows and red circles respectively, and the 3-10 helix as a pink circle. (B) RAPc8 amidase topology; beta sheets and helices are shown as purple arrows and blue cylinders respectively, and the 3-10 helix as a cyan tube. (C) A stereo view of cartoon representation of superimposed monomers of 1I89 (in green and red) and RAPc8 amidase (in purple and blue); the secondary structure colour codes correspond to those in the topology diagrams. 1I89 topology diagram was taken from Kumaran et al., 2003. The cartoon was rendered in *PyMOl*. (Delano, 2004).

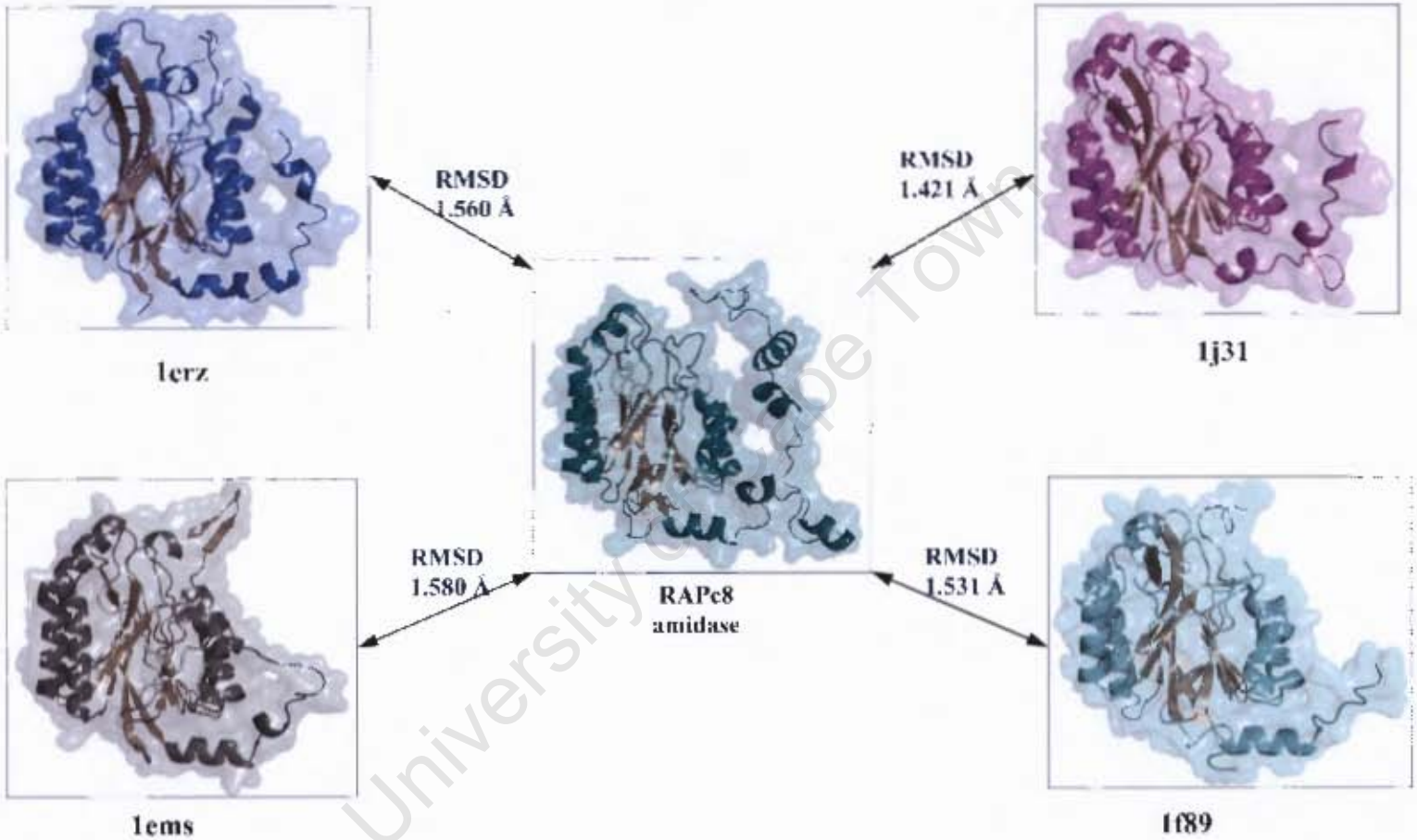


Figure 3-28: Transparent surface and cartoon representation of structural similarity between RAPc8 amidase and the four structures in the nitrilase superfamily. The structures are labelled with their PDB IDs. The average Root Mean Square Deviation (RMSD) over the superimposed C-alpha carbons between RAPc8 amidase and the homologous is 1.523 Å. The secondary structure elements are depicted in different colours to highlight the conserved α-β-β-α sandwich fold. The figures were generated in *PyMOL* (Delano, 2004).

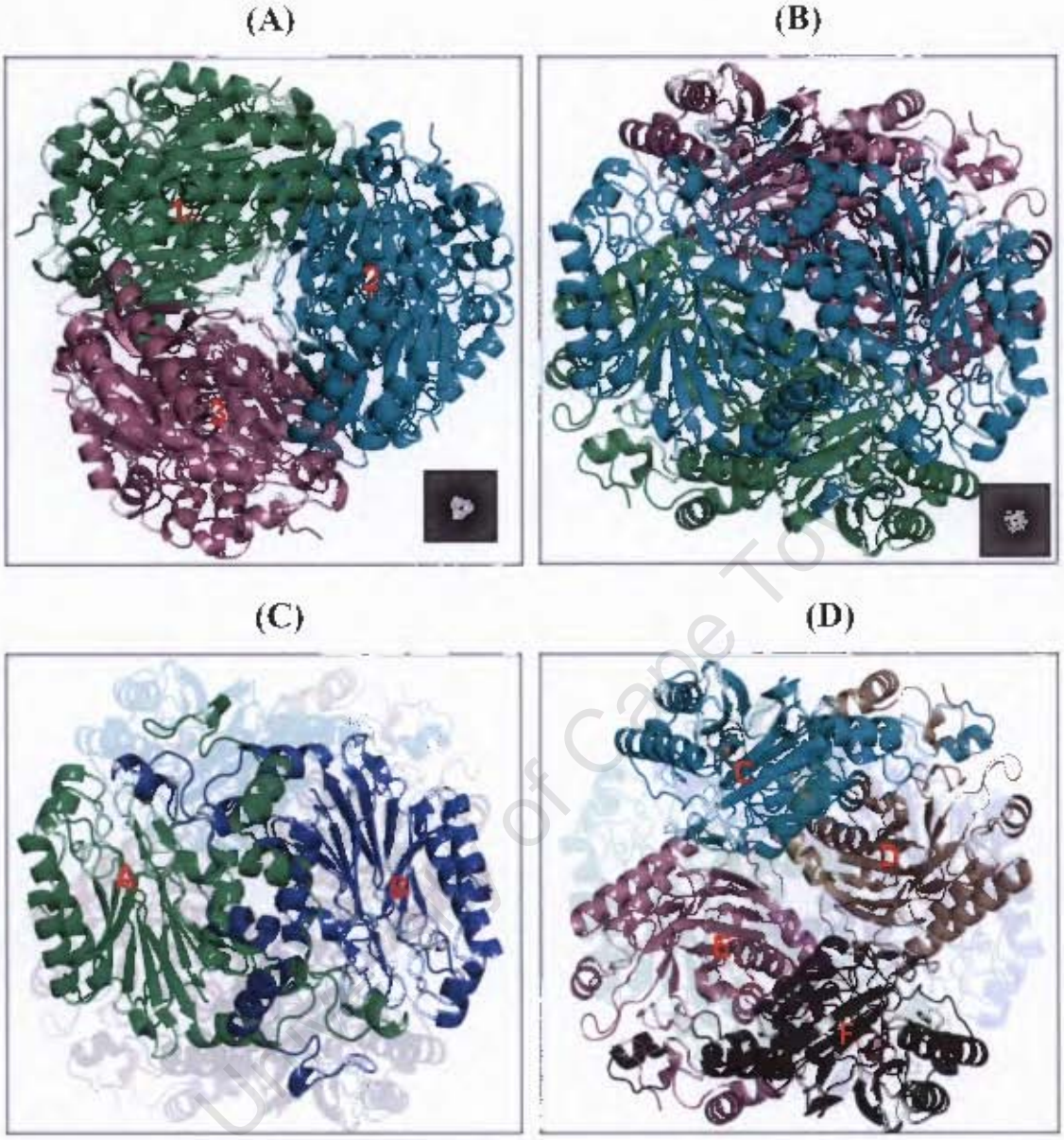


Figure 3-29: A cartoon representation of RAPc8 amidase homohexameric structure. (A) An associated hexamer viewed down the crystallographic three-fold axis; the hexamer is composed of trimers of dimers. The three crystallographic two-fold axes are perpendicular to the three-fold axis, giving rise to D3 point group symmetry that was observed in the EM projection (inset). The three dimers are labelled (1, 2, & 3). (B) A hexamer of RAPc8 amidase as viewed down the crystallographic two-fold axis (a two-fold view EM projection is inset). (C) A two-fold axis view of the hexamer, with an emphasis on one of the dimers (dimer 1). Monomers A and B of dimer 1 are shown in green and blue respectively. The other 2 dimers are shown in transparent cartoons. (D) The same two-fold view of the hexamer with dimer 1 in transparent cartoons. Dimer 2 is consists of monomers C (in cyan) and D (in wheat) while monomers E (in pink) and F (in grey) make up dimer 3. The cartoons were rendered using *PyMOL* (Delano, 2004).

The interaction of two monomers along the crystallographic two-fold axis to form a dimer in the crystal structure, results in an eight-layered α-β-β-α:α-β-β-α sandwich structure (Figure 3-30 (A)). This dimer interface (also denoted as A-surface by Sewell and colleagues (2003)) is conserved in the four nitrilase superfamily crystal structures (PDB IDs: 1j31, 1f89, 1erz, & 1ems; Sakai et al., 2004, Kumaran et al., 2003, Nakai et al., 2000 & Pace et al., 2000, respectively). Similar to the homologous structures, the conserved dimeric interface in RAPc8 amidase is formed by helices α5 (residues 166-186) and α6 (residues 195-213) from both monomers, as well as the C-terminal tails (residues 273-340) that fold over the entire A-surface. The interfacing helices (α5 and α6) are anti-parallel and they form compact interactions that are mainly hydrophobic, in addition to 4 salt bridges involving Glu173 and Arg176 of each monomer (Figure 3-30(C)). Two hydrogen bonds also exist between Trp209 and Asn170. Some of the residues (atoms) interacting on the dimer interface are listed in table 3-9 below

Table 3-9: Interactions at the conserved dimer interface.

Monomer A	**Monomer B**	**Distance (Å)**
Arg176 NH2	Glu173 OE1	2.89
Arg176 NE	Glu173 OE2	2.80
Trp209 NE1	Asn170 OD1	2.88
Met202 SD	Lys205 Cβ	3.40

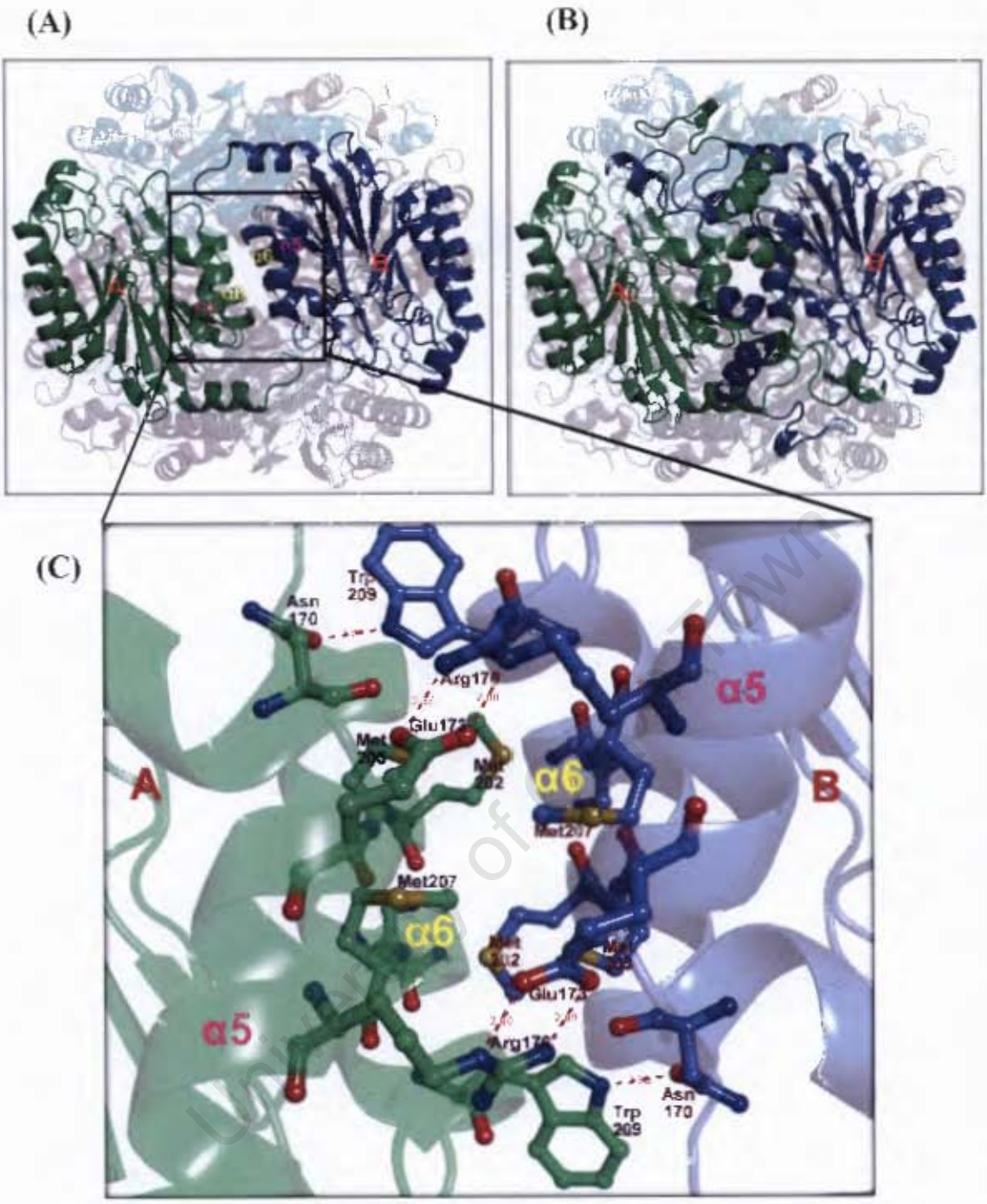


Figure 3-30: Cartoon representation of the RAPc8 amidase hexamer with an emphasis on the dimer interface (A-surface) from one of the dimers. (A) Dimer 1 in the hexamer, viewed down the conserved two-fold axes; the extended C-terminal tail is not shown. The monomers are labelled A and B, the interface helices are also labelled. The monomers associate to form an eight-layered α-β-β-α:α-β-β-α sandwich structure. (B) The same dimer interface; the C-terminal tail that is folded over the A-surface is now shown. (C) Transparent cartoon and ball-and-stick representation of the detailed interaction of helices α5 and α6 from both monomers; methionines (202, 203 and 207) from both monomers contribute to hydrophobic interactions along the interface, which are further strengthened by four salt bridges involving Arg176 and Glu173. Trp209 and Asn170 also contribute two hydrogen bonds in the interface. Hydrogen bonds and salt bridges are shown in red dashes. The C-terminal tail folds over the A-surface, to further strengthen the interface as described later in the chapter. The cartoons were rendered in *PyMOL* (DeLano, 2004).

The interaction of helices α6 on the conserved dimer interface shows a class II helix-helix packing pattern (Chothia et al., 1977) where the angle between the packing helices is approximately -60°. This packing is illustrated in Figure 3-31 below.

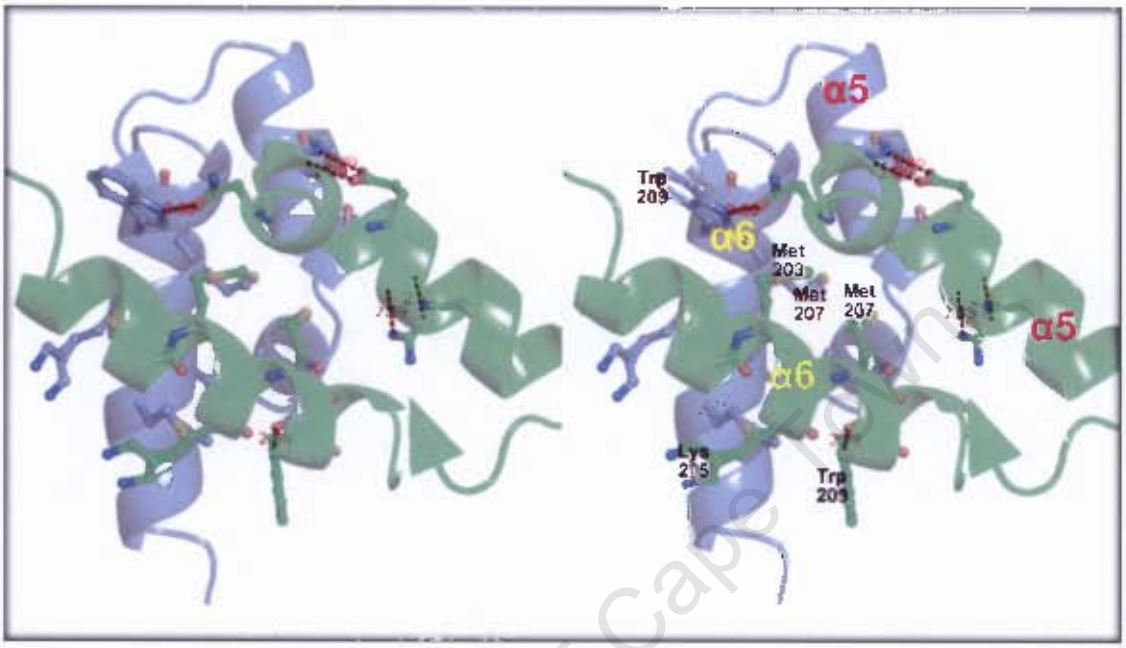


Figure 3-31: A stereo diagram of a cartoon representation of the helix-helix packing pattern at the conserved dimer interface in RAPc8 amidase. Helices from chain A are coloured green while those from Chain B are blue, they are all labelled. The helices are viewed from a direction perpendicular to the crystallographic two-fold axes. Helices α6 from both monomers pack against one another by crossing over to form a class II helix-helix pattern. The image was generated using *PyMOL* (Delano, 2004).

In the hexameric complex (Figure 3-29), one dimer interacts with the other two dimers at two-fold interfaces (which we have termed F-surfaces) that are opposite the conserved dimer interfaces (A-surface). Specifically, chain A of dimer 1 interacts with chain F of dimer 3, chain B of dimer 1 with chain C of dimer 2 while chain D of dimer 2 interacts with chain E of dimer 3, so that interface A-B is opposite D-E, C-D is opposite F-A and E-F is opposite B-C. The F-surface is formed by helices α7 and the N-terminal loops of the two interacting monomers. These interactions are predominantly electrostatic with a number of salt bridges and a few hydrophobic interactions. Figure 3-32 shows a 2D cylindrical projection of the RAPc8 amidase hexamer density, which makes it easy to visualize the interacting subunits. Table 3-10

shows a list of the interacting residues on the F-surface while Figure 3-33 shows the details of these interactions.

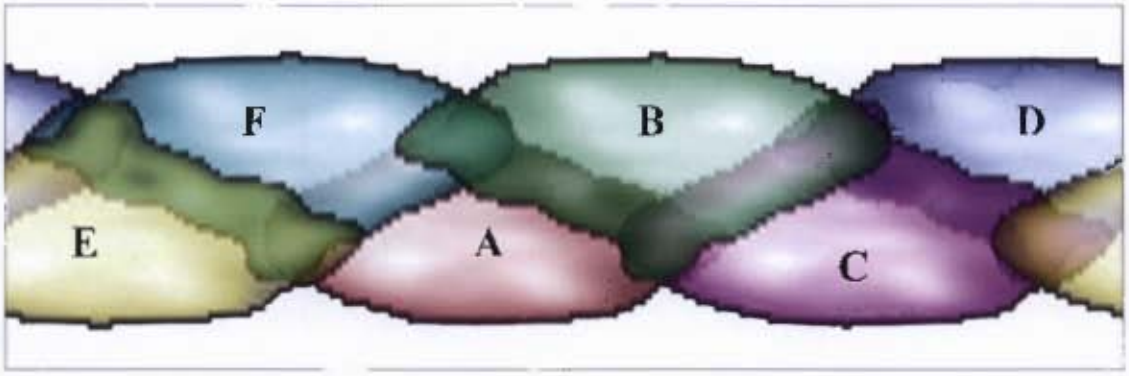


Figure 3-32: A 2D cylindrical projection of the density of the RAPc8 amidase hexamer, with the cylinder axis and the three-fold axis aligned. The density of each subunit was projected separately, coloured and then combined to form the composite image. The subunits are labeled. The conserved A-surfaces exist between A-B, C-D and E-F, which constitute dimers 1, 2, and 3 respectively, while the novel F-surfaces links A-F, B-C and D-E. The figure was generated using UCSF Chimera (Pettersen *et al.*, 2004)

Table 3-10: Interactions at the two-fold interface (F-surface) that links one dimer to another in RAPc8 amidase hexameric complex.

Monomer B	Monomer C	Distance (Å)
Arg2 NH1	Gln262 OE1	2.94
Arg2 NH2	Asp265 OD2	2.95
His3 N	Asp239 OD2	2.94
Asp5 OD1	Gln273 N	2.71
Asp5 OD2	Gln271 N	2.85
Ser7 OG	Asn269 O	2.98
Ser9 OG	Asp265 OD2	2.67
Arg241 NH1	Thr242 O	2.98

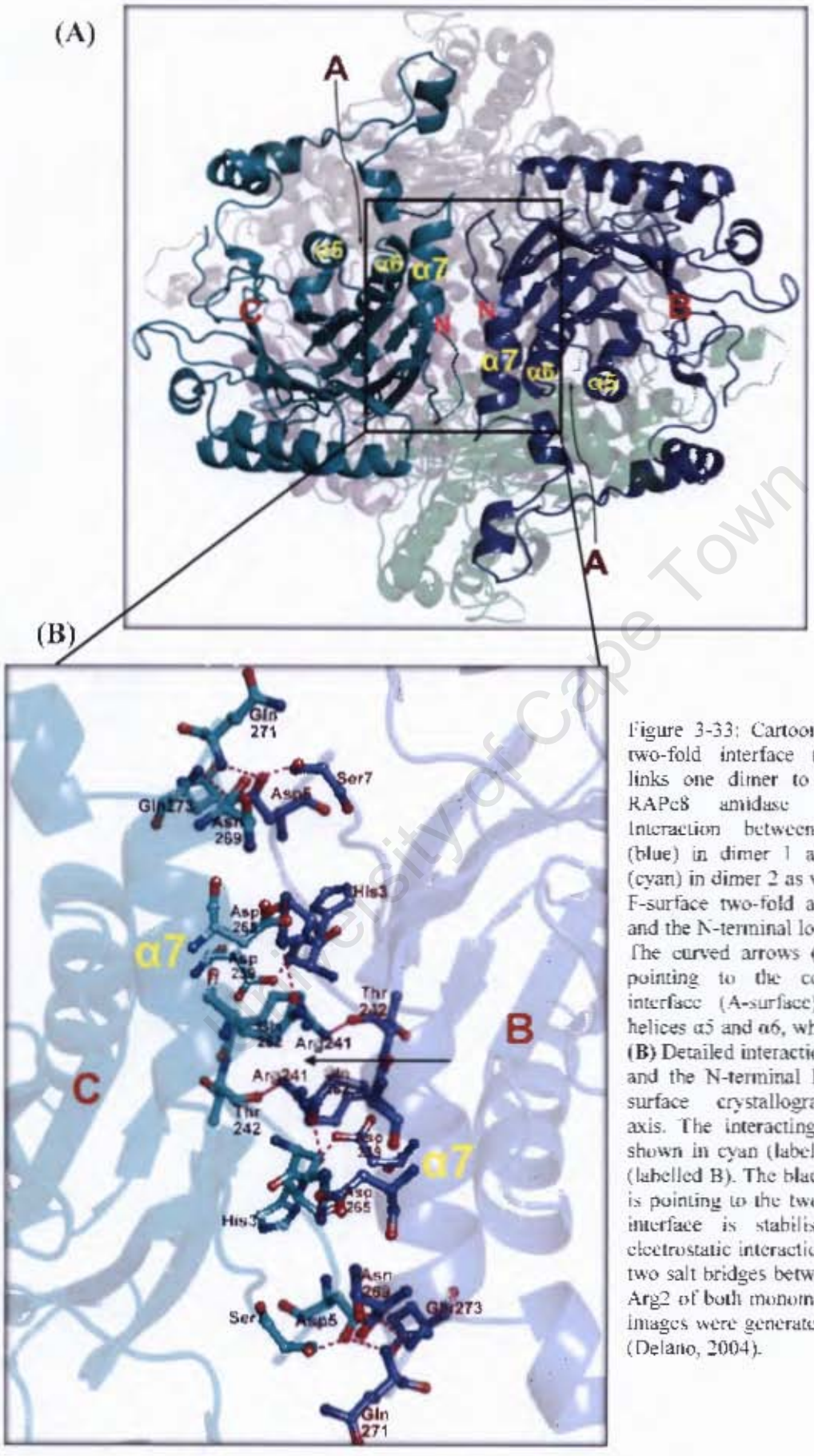


Figure 3-33: Cartoon images of the two-fold interface (F-surface) that links one dimer to another in the RAPs8 amidase hexamer: **(A)** Interaction between monomer B (blue) in dimer 1 and monomer C (cyan) in dimer 2 as viewed down the F-surface two-fold axes. Helices α7 and the N-terminal loops are labelled. The curved arrows (labelled A) are pointing to the conserved dimer interface (A-surface) that involves helices α5 and α6, which are labelled. **(B)** Detailed interactions of helices α7 and the N-terminal loops on the F-surface crystallographic two-fold axis. The interacting monomers are shown in cyan (labelled C) and blue (labelled B). The black straight arrow is pointing to the two-fold axis. This interface is stabilised mainly by electrostatic interactions, with at least two salt bridges between Asp265 and Arg2 of both monomers. The cartoon images were generated using *PyMOL* (Delano, 2004).

(D) - The extended C-terminal tail in RAPc8 amidase has a structural function

Residues 272 to 340 form an extended C-terminal tail that is folded into 4 helices; α8, α9, α10 and α11 (see topology diagram in Figure 3-26 above). The C-terminal tail from one monomer is interlocked with that from the second monomer. The two termini fold over (Figure 3-34 (A)) the conserved dimer interface (A-surface) further stabilizing it. The interaction between the two tails is mainly electrostatic, with two salt bridges between Arg295 and Glu320. Electrostatic contacts also exist between the tails and the core structures, where a tail from monomer A interacts with the residues along the dimer interface on monomer B. These interactions are illustrated in Figure 3-34 (B) below, while table 3-11 lists the interacting residues.

Table 3-11: Interactions between the C-terminal tails and also between the tails and the core structure in RAPc8 amidase. The residues in bold characters are involved in interactions between the tails and the core structure while those in plain text are interactions between the C-terminal tails from the two monomers.

Monomer A	Monomer B	Distance (Å)
Lys113 NZ	**Glu292 OE2**	**2.72**
Ala114 N	**Ser290 OG**	**2.93**
Try138 NE1	**Gln273 OE1**	**2.97**
Tyr145 OH	**Glu292 OE1**	**2.61**
Asp167 OD2	**Lys278 NZ**	**2.70**
Asp177 OD1	**Phe304 N**	**2.80**
Tyr192 OH	**Gln273 OE1**	**2.69**
Ala270 O	Arg324 NH1	2.85
Gln273 OE1	Tyr192 OH	2.69
His281 CE1	Glu320 OE2	2.62
Arg295 NH2	Glu320 EO2	2.68
Ala298 O	Trp308 NE1	2.88
Pro140 O	Thr285 OG1	2.65
Asn269 OD1	Thr329 OG1	2.70

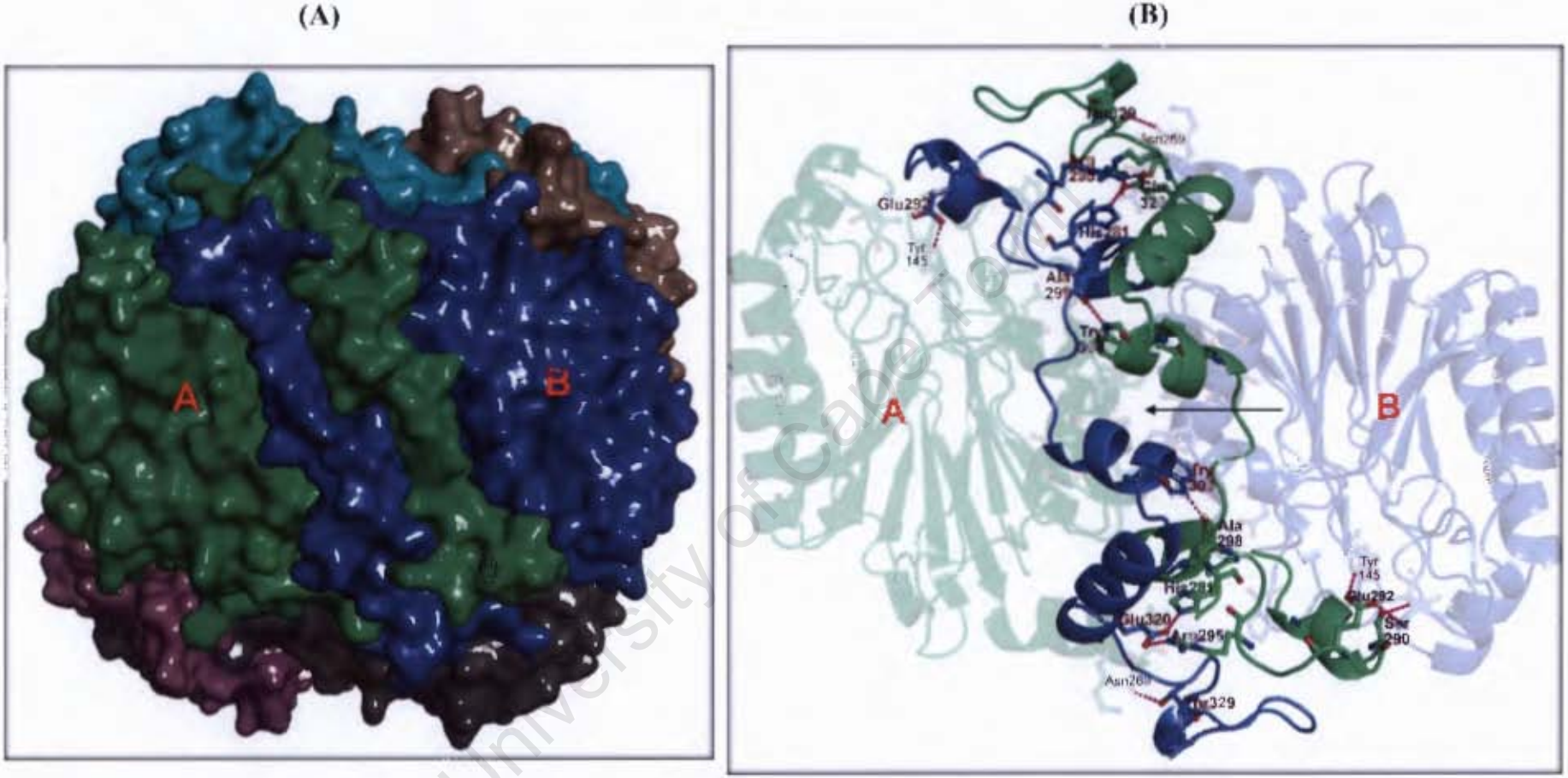


Figure 3-34: Illustration of the structural role of the extended C-terminal tails in the RAPc8 amidase hexameric complex. (A) A molecular surface representation of a RAPc8 amidase hexamer as viewed down the crystallographic two-fold axes of dimer 1. Monomers A (in green) and B (in blue) are labelled. Domain swapping or interlocking, involving C-terminal tails from both monomers is clearly visible. (B) A detailed cartoon illustration of the interactions between the C-terminal tails in the dimer interface and the contact of the tails with the core structure, as viewed down the two-fold crystallographic axis that is indicated by the black arrow. Chains A (in green) and B (in blue) are labelled. Some of the residues that are involved in hydrogen bonding network and salt bridge interactions between the two tails are shown in ball-and-stick representation, and labelled in bold text. Residues on the core structure that interact with the tails labelled in smaller text. The figures were generated in *PyMOL* (Delano, 2004).

Aside from the overall strengthening of the A-surface, the domain swapping of the C-terminal tails could also play the role of linking the monomers in the process of dimer formation. This role was first proposed by Nakai and colleagues (2000), in the structure of N-carbamyl-D-amino acid amidohydrolase (DCase: PDB ID, 1erz). RAPc8 amidase C-terminal tails however are unique in that they are much longer which probably emphasizes their role in driving the process of hexamer formation, as well as in holding the hexameric complex together.

(E) - The RAPc8 amidase active site framework is similar to nitrilase superfamily enzymes

RAPc8 amidase shares structural homology to the members of the nitrilase superfamily. One of the conserved motifs in the nitrilase enzymes is the cysteine, glutamate and lysine (Glu, Lys, Cys) catalytic triad (Novo et al., 1995; Pace, Brenner, 2001) that is involved in the hydrolysis of the carbon-nitrogen bond. In RAPc8 amidase, the catalytic triad is composed of Glu59, Lys143 and Cys166. The coordination of these residues is similar to that of the four nitrilase superfamily structural homologues, as shown in Figure 3-35 below.

A buried cleft corresponding to the binding pocket was identified and characterised in each monomer of the RAPc8 amidase structure using *CASTp* (Binkowski et al., 2003) online server. The pocket is formed by residues in the loops between β-strand β2 and the small helix α2, strands β5 and β6, strand β7 and helix α5, as well as strand β8 and helix α6. It is located near the edge of the dimer interface. This active site location is similar to all α/β four-layer sandwich structures including nitrilase superfamily enzymes. Figure 3-36 below shows the location of active site clefts of both RAPc8 amidase and N-carbamyl-D-amino acid amidohydrolase (DCase) (Nakai et al., 2000).

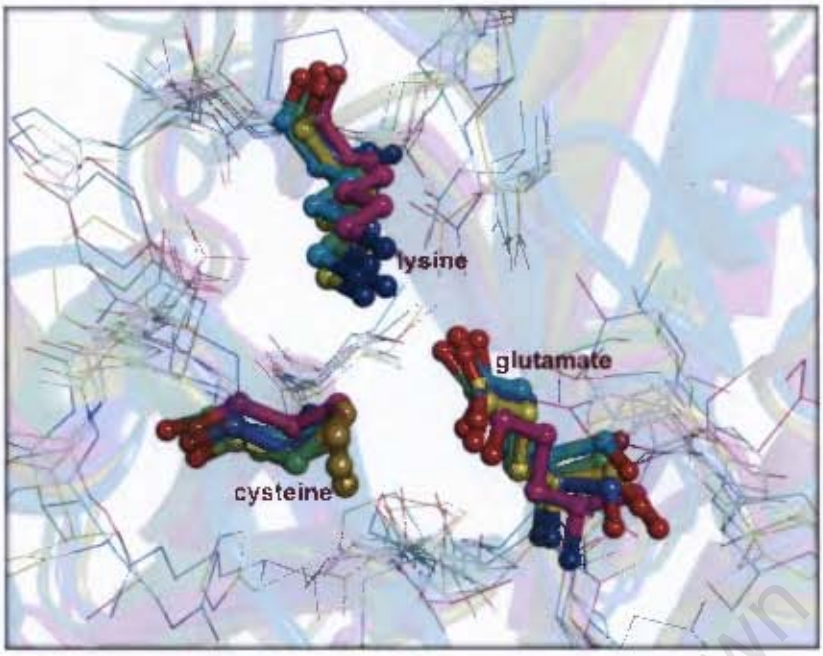


Figure 3-35: A transparent cartoon image showing the coordination of the catalytic triad residues in RAPc8 amidase as compared to the nitrilase superfamily homologous structures (PDB IDs, 1j31, 1f89, 1erz, and 1ems). 1j31 is in yellow, 1ems in green, 1erz in cyan, 1f89 in magenta and RAPc8 amidase in blue. The catalytic triad residues (Glu, Lys, Cys) have similar coordination and position in all five structures indicating their role in driving the catalysis in the superfamily. This further confirms the structural relationship between RAPc8 amidase structure and the existing nitrilase superfamily structures. The five structures were superimposed with *ALIGN* (Cohen, 1997) and the figure was rendered in *PyMOL* (Delano, 2004).

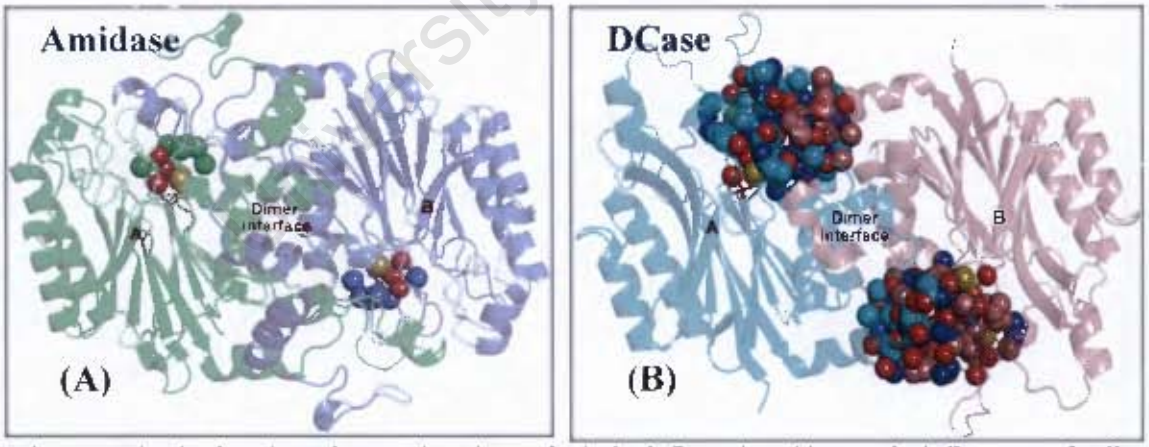
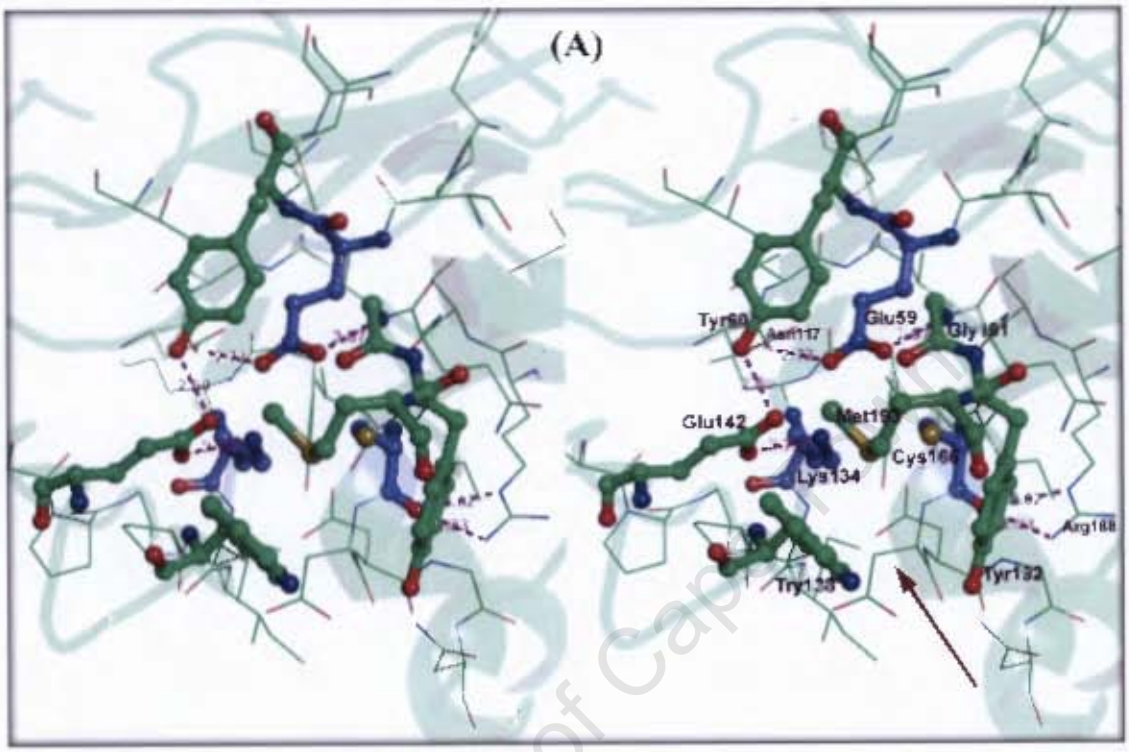


Figure 3-36: The location of the active site clefts in both RAPc8 amidase and nitrilase superfamily enzyme, DCase (Nakai *et al.*, 2000): (A) Cartoon representation of a RAPc8 amidase dimer; the monomers are shown in green (monomer A) and blue (monomer B). (B) A cartoon representation of a DCase dimer; monomers A (in cyan) and B (in brown) are labelled. The dimer interfaces are labelled and the atoms lining the binding sites are rendered as spheres. The active-site pocket exists in each monomer in both cases with the pockets having similar locations. DCase pocket is however many times larger and deeper than that of RAPc8 amidase. The figures were generated using *PyMOL* (Delano, 2004).

The RAPc8 amidase active site cleft is lined with residues Glu59, Tyr60, Trp138, Glu142, Cys166, Gly191, Tyr192 and Met193 as depicted in Figure 3-37 below. A hydrogen bonding network exists that plays a role in maintaining a highly constrained configuration of the residues in the substrate binding site. The catalytic residue Cys166 is located on a β strand-turn-helix structural motif, which is formed by strand β7 and helix α5 (Figure 3-37 (B)). This motif is termed a nucleophile elbow (Kumaran et al., 2003) and is found in many α/β four-layer sandwich hydrolases (Schrag & Cygler, 1997). Similar to the catalytic cysteines in the other C-N hydrolase structures, Cys166 has a rare but allowed combination of peptide bond dihedral angles, falling in the recently described type II'-turn conformation (Lovell et al., 2003) region of the Ramachandran plot. The carbonyl oxygen of Cys166 forms hydrogen bonds with the NE (3.02 Å) and NH2 (2.83 Å) atoms of Arg188; this hydrogen bonding network may play a role in maintaining the orientation of Cys166 side chain in the active site.

A second glutamate residue (Glu142) is present in the active site. This may play a role in making Glu59 a better nucleophile, which together with an electrophilic Lys134, assists Cys166 in attacking the carbonyl carbon of an amide substrate. Figure 3-38 presents a catalytic mechanism of RAPc8 amidase that was proposed in our research group. Upon the nucleophilic attack of the carbonyl group by Cys166, an unstable tetrahedral intermediate (number [1] in the diagram) is formed. With the removal of ammonia, a nucleophilic Lys134 and a water molecule assist in the generation of a second tetrahedral intermediate [3], which then decomposes into an acidic product and a regenerated amidase active site environment.



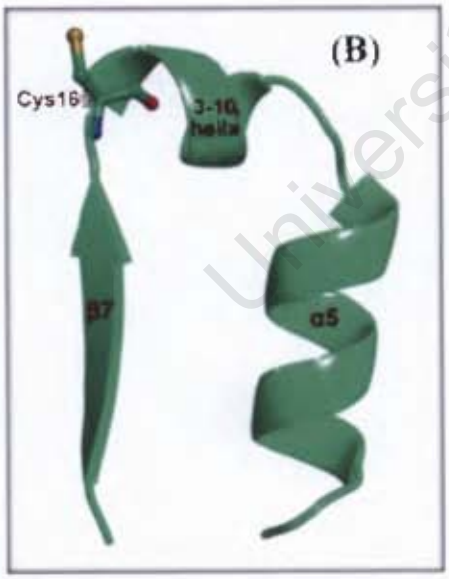


Figure 3-37: RAPc8 amidase active site structure. (B) A stereo view of the residues lining the active site pocket; Glu59, Lys134 and Cys166 are the catalytic triad residues, and they are shown in blue ball-and-sticks. The hydrogen bonds between atoms are shown in purple dashes. The brown arrow points to the possible route into the cleft. (B) A cartoon diagram of the nucleophile elbow where the catalytic Cys166 is located. Cys166 is shown as ball-and-stick. Similar to other nitrilase superfamily structures, Cys166 backbone adopts type II'-turn conformation with a positive phi angle. This is important in maintaining a favourable side chain orientation for catalysis. The images were rendered using *PyMOL* (Delano, 2004).

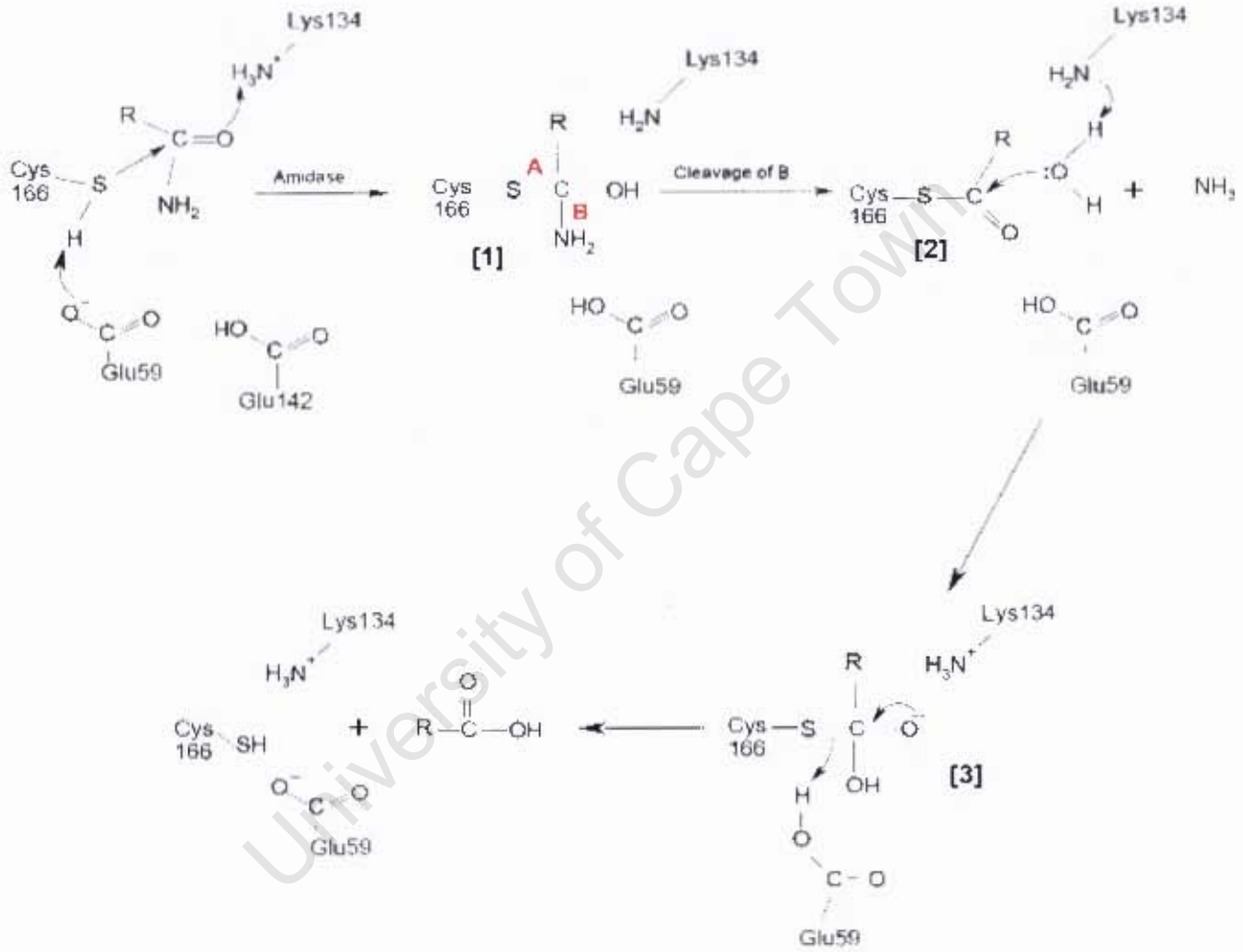


Figure 3-38: The proposed catalytic mechanism of *Geobacillus pallidus* RAPc8 amidase

Unlike in the other nitrilase superfamily structures where the active site clefts are reported to be large and deep (Kumaran et al., 2003; Nakai et al., 2000; Sakai et al., 2004), the RAPc8 amidase binding pocket is relatively shallow, with a molecular volume of 53.59 $Å^3$ and a solvent accessible area of 4.683 $Å^2$. The DCase substrate-binding pocket for instance has a molecular volume of 1134.7 $Å^3$ and a solvent accessible area of 425.8 $Å^2$, making it over 20 times bigger than RAPc8 amidase pocket. The small size of the active site pocket may explain the limited specificity of RAPc8 amidase for short acyl-chained aliphatic amides as revealed by recent biochemical characterization (Makhongela et al., in press). A comparison of the RAPc8 amidase pocket (Figure 3-39) with that of a substrate-bound DCase mutant structure (PDB ID: 1fu5, Hashimoto et al., unpublished data) rationalized the specificity of RAPc8 amidase for small un-branched amide compounds and also revealed the role of some active site residues in restricting the size of amide substrates. The bulky side chain from Trp138 partly covers the binding pocket from the outside, hence acting as a 'filter', that only allows small un-branched substrates to access the active site. The presence of the Trp138 side chain at the gateway to the substrate binding pocket may also explain why *CASTp* (Binkowski et al., 2003) could not locate any mouth openings for the RAPc8 amidase binding pocket.

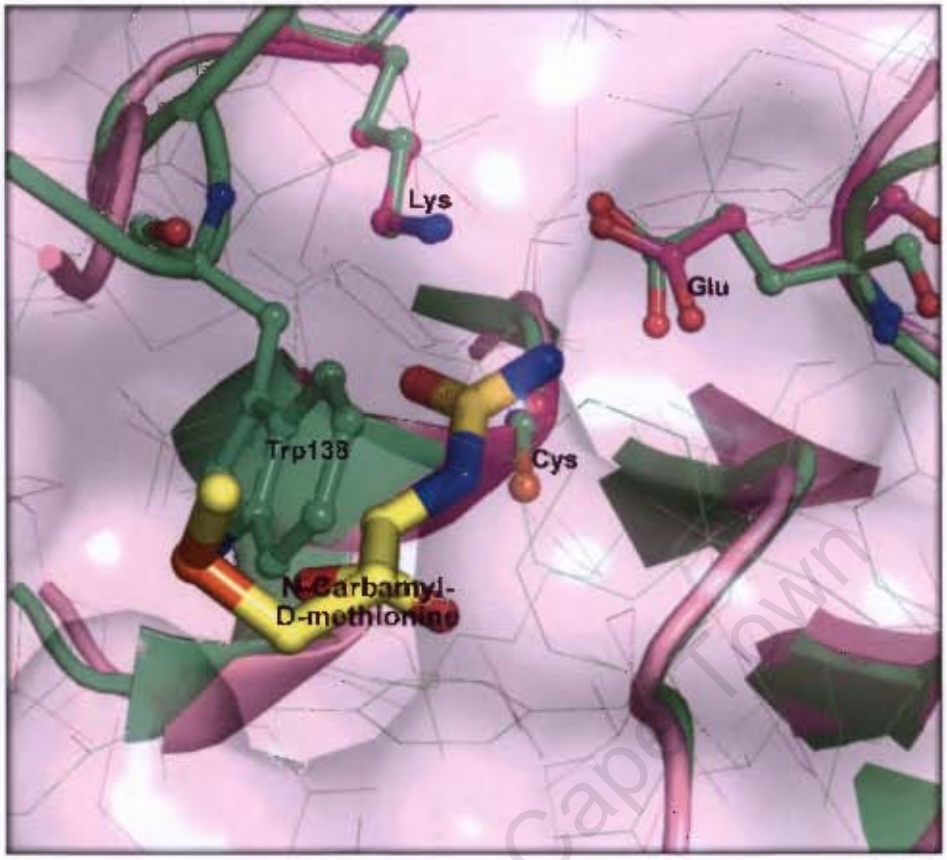


Figure 3-39: Surface and cartoon representation of a comparison between RAPc8 amidase active site pocket and that of a substrate-bound DCase mutant (PDB ID: 1fu5, Hashimoto *et al.*, unpublished data). RAPc8 amidase is shown in green while DCase is in violet. The N-carbamyl-D-methionine substrate bound in the active site pocket of DCase is shown in yellow sticks and labelled. The catalytic triad residues (shown in ball-and-stick) are in similar positions in the two structures. The Trp138 side chain of RAPc8 amidase clashes with the bulky substrate of DCase. This means that the amide substrates of RAPc8 amidase would have to be of a smaller size in order to access the pocket for hydrolysis. The two structures were superimposed using *ALIGN* (Cohen, 1997) and the image was rendered with *PyMOL* (Delano, 2004).

(F) - The catalytic Cys166 is modified

Unlike other nitrilase superfamily structures in which a number of bound water molecules, believed to participate in catalysis, are present in the active sites, no waters were found in the active site pocket of RAPc8 amidase. However, both $2F_{obs}–F_{calc}$ and positive difference electron density maps indicated that Cys166 was modified, possibly by oxidation in the crystal (Figure 3-40). Attempts to fit this density either as

a sulfinic acid or a sulfenic acid were unsuccessful. We therefore suggest that the Cys166 oxidative modification is a mixture of species. The importance (if any) of this modification is uncertain, but any form of oxidation would decrease the nucleophilicity of Sγ atom in the modified cysteine residue; this partly explains why RAPc8 amidase enzyme progressively lost activity in the absence of reducing agents during purification (Makhongela et al. in press).

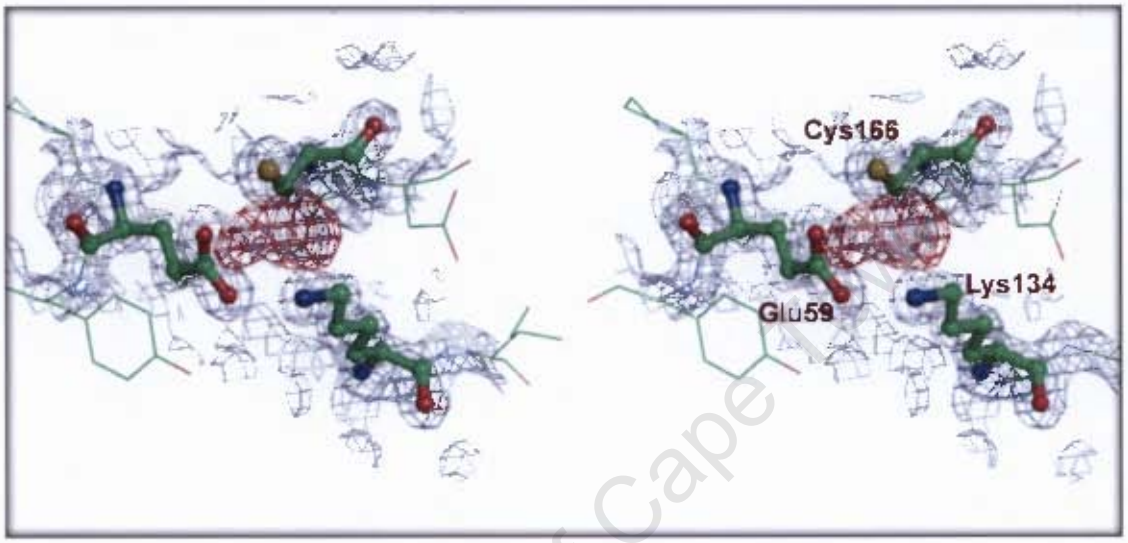


Figure 3-40: A stereo view of the electron density maps around the active site residues Glu59, Lys 134, and Cys166. A sigma-A weighted $2F_{obs}-F_{calc}$ map contoured at 1.3 σ is shown in blue while a sigma-A weighted F_{obs} F_{calc} difference map contoured at 3.0 σ and is shown in red. The positive difference electron density around Sγ atom of Cys166 suggests oxidative modification probably to a mixture of species, including sulfinic acid. The image was rendered using *PyMOL* (Delano, 2004).

CHAPTER 4: CONCLUSIONS

The successful determination of the *Geobacillus pallidus* RAPc8 amidase crystal structure was a combination of skills and techniques from various fields, each of which contributed unique information to the process.

Gel filtration analysis was the first to shed light on the possible biological unit of the RAPc8 amidase enzyme. A 220 kDa peak on a preparative gel exclusion chromatography column indicated an approximate hexameric complex, of 38 kDa molecular weight per subunit. Electron microscopy analysis allowed for the visualization of these complexes, and provided crucial information on the symmetry of the molecule, which strongly guided the X-ray crystallography work. The prediction of the RAPc8 amidase secondary structure elements revealed the structural relationship between this enzyme and the members of the nitrilase superfamily, despite the low sequence homologies that exist between these enzymes.

High protein purity levels as well as homogeneity were important in allowing crystallization, leading to highly ordered crystals that diffracted well, and that generated a high quality dataset, which led to the determination of RAPc8 amidase structure. The power of maximum likelihood-based crystallographic software cannot be under-estimated in this work. A new generation of software is now emerging that is allowing determination of difficult structures. The high symmetry of RAPc8 amidase crystal and the low sequence identity between RAPc8 amidase and the nitrilase superfamily distant homologues, presented huge obstacles in the search for the molecular replacement solution, and in the model refinement. These were however overcome by a combination of maximum likelihood crystallographic programs including *POINTLESS*, *PHASER*, *PHENIX* and *REFMAC5* among others, that allowed assignment of the correct space group, determination of the molecular replacement solution, and rebuilding of the starting model prior to model fitting and refinement.

The crystal of RAPc8 amidase has revealed high structural homology between this enzyme and the determined structures of other nitrilase superfamily enzymes (PDB IDs, 1j31, 1f89, 1erz and 1f89). Apart from the highly conserved four-layer α–β–β–α sandwich fold, RAPc8 amidase also has Glu, Lys, Cys catalytic triad that has a similar

framework to other nitrilase superfamily enzymes. This is despite that fact that, enzymes in the family have varying substrate specificities. Aside from the conserved dimer interface that is observed across the members in the family, a novel dimer interface (termed as F-surface in this work) has been identified in RAPc8 amidase structure that allows the association of dimers to form a hexamer. RAPc8 amidase has a longer C-terminal sequence than the other solved structures in the superfamily that forms an interlock by domain swapping. This, apart from strengthening the dimer interface, is thought to play the role of a linker in the process of dimer formation, and to hold to the active hexameric enzyme in the catalytic conformation.

A pocket corresponding to the active site cleft of RAPc8 amidase has been identified, that is small enough to explain the specificity of this enzyme for short aliphatic amides.

Future work

Biochemical work that is underway has revealed specificity of RAPc8 amidase for a number of short acyl-chained substrates. Future work will examine the active site cleft to rationalize this specificity. It will also focus on defining the catalytic mechanism that allows amide hydrolysis, acyl transfer and D-lactamide enantioselectivity. To achieve this, a number of active site mutants will need to be prepared, and structures of these mutants in complex with substrates will need to be determined.

REFERENCES

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. 2002. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr 58*:1948-1954.

Agarkar VB, Kimani SW, Cowan DA, Sayed MF-R, Sewell BT. 2006. The quaternary structure of the amidase from *Geobacillus pallidus* sp. RAPc8 is revealed by its crystal packing. *Acta Cryst. F62*:1174-1178.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res 25*:3389-3402.

Ambler RP, Auffret AD, Clarke PH. 1987. The amino acid sequence of the aliphatic amidase from *Pseudomonas aeruginosa*. *FEBS Lett 215*:285-290.

Asano Y, Yasuda T, Tani Y, Yamada H. 1982. Microbial-degradation of nitrile compounds .7. A new enzymatic method of acrylamide production. *Agricultural and Biological Chemistry 46*:1183-1189.

Baker RT, 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, Banerjee UC. 2002. The nitrile-degrading enzymes: current status and future prospects. *Appl Microbiol Biotechnol 60*:33-44.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Research 28*: 235-242.

Binkowski TA, Naghibzadeg S, and Liang J. 2003. CASTp: computed atlas of surface topography of proteins. *Nucleic Acid Research 31*:3352-3355.

Boger DL, Fecik RA, Patterson JE, Miyauchi H, Patricelli MP, Cravatt BF. 2000. Fatty acid amide hydrolase substrate specificity. *Bioorg Med Chem Lett 10*:2613-2616.

Bond CS. 2003. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics 19*:311-312.

Bork P, Koonin EV. 1994. A new family of carbon-nitrogen hydrolases. *Protein Sci 3*:1344-1346.

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

Bretaudiere JP, Frank J. 1986. Reconstitution of molecule images analysed by correspondence analysis: a tool for structural interpretation. *J Microsc 144*:1-14.

Brünger AT. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature 355*: 472-475.

Cameron RA, Sayed M, Cowan DA. 2005. Molecular analysis of the nitrile catabolism operon of the thermophile *Bacillus pallidus* RAPc8. *Biochim Biophys Acta 1725*:35-46.

Canutescu AA, Shelenkov AA, Dunbrack RL, Jr. 2003. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci 12*:2001-2014.

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

Chen CY, Chiu WC, Liu JS, Hsu WH, Wang WC. 2003. Structural basis for catalysis and substrate specificity of *Agrobacterium radiobacter* N-carbamoyl-D-amino acid amidohydrolase. *J Biol Chem 278*:26194-26201.

Cheong TK, Oriel PJ. 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.

Chothia C, Levitt M, Richardson D. 1977. Structure of proteins: Packing of α-helices and pleated sheets. *Chemistry 74(10)*:4130-4134.

Cilia E, Fabbri A, Uriani M, Scialdone GG, Ammendola S. 2005. The signature amidase from *Sulfolobus solfataricus* belongs to the CX3C subgroup of enzymes

cleaving both amides and nitriles - Ser195 and Cys145 are predicted to be the active site nucleophiles. *Febs Journal 272*:4716-4724.

Ciskanik LM, Wilczek JM, Fallon RD. 1995. Purification and Characterization of an Enantioselective Amidase from Pseudomonas chlororaphis B23. *Appl Environ Microbiol 61*:998-1003.

Cohen GH. 1997. ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Cryst. 30*:1160-1161.

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

Cravatt BF, Giang DK, Mayfield SP, Boger DL, Lerner RA, Gilula NB. 1996. Molecular characterization of an enzyme that degrades neuromodulatory fatty-acid amides. *Nature 384*:83-87.

Crowther RA, Henderson R, Smith JM. 1996. MRC image processing programs. *J Struct Biol 116*:9-16.

Curnow AW, Hong K, Yuan R, Kim S, Martins O, Winkler W, Henkin TM, Soll D. 1997. Glu-tRNAGln amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc Natl Acad Sci U S A 94*:11819-11826.

Dale Minerals International (Online) available: http://webmineral.com/crystal/Iso_432.shtml

Delano WL. 2004. The PyMOL Molecular Graphics System. *Delano Scientific*, San Carlos CA, USA. http://www.pymol.org.

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

Farnaud S, Tata R, Sohi MK, Wan T, Brown PR, Sutton BJ. 1999. Evidence that cysteine-166 is the active-site nucleophile of *Pseudomonas aeruginosa* amidase:

crystallization and preliminary X-ray diffraction analysis of the enzyme. *Biochem J 340 (Pt 3)*:711-714.

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

Fournand D, Arnaud A, 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, 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.

Fournand D, Bigey F, Ratomahenina R, Arnaud A, Galzy P. 1997. Biocatalyst improvement for the production of short chain hydroxamic acids. *Enzyme and Microbial Technology 20*:424-431.

Frank J. 2006. Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State, *Oxford University Press*, New York.

Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A. 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol 116*:190-199.

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

Hashimoto H, Aoki M, Shimizu T, Nakai T, Morikawa H, Ikenaka Y, Takahashi, S, Sato M. (unpublished data). Crystal structures of substrate-complexed N-carbamyl-D-amino acid amidohydrolase active site mutants (1uf4, 1uf5, 1uf7 & 1uf8).

Hashimoto Y, Nishiyama M, Ikehata O, Horinouchi S, Beppu T. 1991. Cloning and characterization of an amidase gene from *Rhodococcus* species N-774 and its expression in Escherichia coli. *Biochim Biophys Acta 1088*:225-233.

Hazes B. 2003. Biochem609: Molecular replacement class. (Online) Available: http://eagle.mmid.med.ualberta.ca/tutorials/MR/MR.html

Hirrlinger B, Stolz A, Knackmuss HJ. 1996. Purification and properties of an amidase from *Rhodococcus erythropolis* MP50 which enantioselectively hydrolyzes 2-arylpropionamides. *J Bacteriol 178*:3501-3507.

Holm L and Sander C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol. 233*:123-138.

Hooft RW, Vriend G, Sander C, and Abola EE. 1996. Errors in protein structures. *Nature. 381*:272-272.

Huang W, Jia J, Cummungs J, Nelson M, Schneder G, Lindqvist Y. 1997. Crystal structure of nitrile hydratase reveals a novel iron center in a novel fold. *Structure 5*:691-699.

Hughes J, Armitage YC, Symes KC. 1998. Application of whole cell r*hodococcal* biocatalysts in acrylic polymer manufacture. *Antonie Van Leeuwenhoek 74*:107-118.

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

Jones DT. 1999a. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol 287*:797-815.

Jones DT. 1999b. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol 292*:195-202.

Jones T, Bergdoll M and Kjeldgaard M. 1990. O: A macromolecular modelling environment. In: *Crystallographic and Modeling Methods in Molecular Design.* Eds.: C. Bugg & S. Ealick. Springer-Verlag Press: 189-195.

Karmali A, Tata R, Brown PR. 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, Asano Y. 2000. Distribution of aldoxime dehydratase in microorganisms. *Appl Environ Microbiol 66*:2290-2296.

Kato Y, Tsuda T, Asano Y. 1999. Nitrile hydratase involved in aldoxime metabolism from *Rhodococcus* sp. strain YH3-3 purification and characterization. *Eur J Biochem 263*:662-670.

Keating TA, Marshall CG, Walsh CT, Keating AE. 2002. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol 9*:522-526.

Kim S, Oriel P. 2000. Cloning and expression of the nitrile hydratase and amidase genes from *Bacillus* sp. BR449 into Escherichia coli. *Enzyme Microb Technol 27*:492-501.

Kissinger CR, Gehlhaar DK, Fogel DB. 1999. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr D Biol Crystallogr 55*:484-491.

Kleywegt GJ. 2000. Validation of protein crystal structures. *Acta Crystallogr D Biol Crystallogr 56*:249-265.

Kobayashi M, Fujiwara Y, Goda M, Komeda H, Shimizu S. 1997. Identification of active sites in amidase: evolutionary relationship between amide bond- and peptide bond-cleaving enzymes. *Proc Natl Acad Sci U S A 94*:11986-11991.

Kobayashi M, Goda M, Shimizu S. 1998. The catalytic mechanism of amidase also involves nitrile hydrolysis. *FEBS Lett 439*:325-328.

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

Kotlova EK, Chestukhina GG, Astaurova OB, Leonova TE, Yanenko AS, Debabov VG. 1999. Isolation and primary characterization of an amidase from Rhodococcus rhodochrous. *Biochemistry (Mosc) 64*:384-389.

Kumaran D, Eswaramoorthy S, Gerchman SE, Kycia H, Studier FW, Swaminathan S. 2003. Crystal structure of a putative CN hydrolase from yeast. *Proteins 52*:283-291.

Kvalnes-Krick KL, Traut TW. 1993. Cloning, sequencing, and expression of a cDNA encoding beta-alanine synthase from rat liver. *J Biol Chem 268*:5686-5693.

Labahn J, Neumann S, Buldt G, Kula MR, Granzin J. 2002. An alternative mechanism for amidase signature enzymes. *J Mol Biol 322*:1053-1064.

Lamzin VS, Perrakis A and Wilson KS. 2001. The ARP/WARP suite for automated construction and refinement of protein models. In *Int. Tables for Crystallography.* Vol. F: Crystallography of biological macromolecules (Rossmann, M.G. & Arnold, E. eds.), Dordrecht, Kluwer Academic Publishers, The Netherlands, pp. 720-722.

Laskowski RA, MacArthur MW, Moss DS and Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst. 26*:283-291.

Liu JK, Liu CH, Lin CS. 1997. The role of nitrogenase in a cyanide-degrading *Klebsiella oxytoca* strain. *Proc Natl Sci Counc Repub China B 21*:37-42.

Lovell SC, Davis IW, Arendall WB, III, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. 2003. Structure validation by C-alpha geometry: phi,psi and C-beta deviation. *Proteins 50*:437-450.

Maestracci M, Thiery A, Arnaud A, Galzy P. 1986. A study of the mechanism of the reactions catalyzed by the amidase from *Brevibaterium* sp. R312. *Agricultural and biological chemistry. 50*:2237-2241.

Makhongela HS, Glowacka AE, Agarkar VB, Sewell BT, Weber B, Cowan D & Burton SG. (in press). A novel thermostable nitrilase superfamily amidase from *Geobacillus pallidus* with acyl transfer activity. *Applied Microbiology and Biotechnology.*

Matthews BW. 1968. Solvent content of protein crystals. *J Mol Biol 33*:491-497.

Mayaux JF, Cerbelaud E, Soubrier F, Yeh P, Blanche F, Petre D. 1991. Purification, cloning, and primary structure of a new enantiomer-selective amidase from a

Rhodococcus strain: structural evidence for a conserved genetic coupling with nitrile hydratase. *J Bacteriol 173*:6694-6704.

Mayaux JF, Cerebelaud E, Soubrier F, Faucher D, Petre D. 1990. Purification, cloning, and primary structure of an enantiomer-selective amidase from *Brevibacterium* sp. strain R312: structural evidence for genetic coupling with nitrile hydratase. *J Bacteriol 172*:6764-6773.

McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. 2005. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr 61*:458-464.

McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server. *Bioinformatics 16*:404-405.

Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP. 1998. JOY: protein sequence-structure representation and analysis. *Bioinformatics 14*:617-623.

Moore PB, Steitz TA. 2002. The involvement of RNA in ribosome function. *Nature 418*:229-235.

Moreau JL, Bernet N, Arnaud A, 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.

Mueller P, Egorova K, Vorgias CE, Boutou E, Trauthwein H, Verseck S, Antranikian G. 2006. Cloning, overexpression, and characterization of a thermoactive nitrilase from the hyperthermophilic archaeon Pyrococcus abyssi. *Protein Expr Purif. 47(2)*:672-81.

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

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr 53*:240-255.

Mylerova V & Martinkova L. 2003. Synthetic applications of nitrile-converting enzymes. *Current Organic Chemistry 7*: 1-17.

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

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

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*-carbamoyl-D-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases. *Structure 8*:729–739.

Navaza J. 1994. AMoRe: an Automated Package for Molecular Replacement. *Acta Cryst. A50*:157-163.

Nawaz MS, Khan AA, Bhattacharayya D, Siitonen PH, Cerniglia CE. 1996. Physical, biochemical, and immunological characterization of a thermostable amidase from *Klebsiella pneumoniae* NCTR 1. *J Bacteriol 178*:2397-2401.

Novo C, Farnaud S, Tata R, Clemente A, Brown PR. 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, Brown PR. 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.

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. 1997. CATH: A hierarchic classification of protein domain structures. *Structure 5*:1093-1108.

O'Reilly C, Turner PD. 2003. The nitrilase family of CN hydrolysing enzymes - a comparative study. *J Appl Microbiol 95*:1161-1174.

Otwinowski Z and Minor W. 1997. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* Volume 276: Macromolecular Crystallography, part A, p.307-326,C.W. Carter, Jr. & R.M. Sweet, Eds., Academic Press.

Pace HC, Brenner C. 2001. The nitrilase superfamily: classification, structure and function. *Genome Biol 2:REVIEWS* 0001.1-0001.9.

Pace HC, Hodawadekar SC, Draganescu A, Huang J, Bieganowski P, Pekarsky Y, Croce CM, 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.

Pollak P, Romender G, Hagedorn F, Gelbke H-P. 1991. In: Elvers B, Hawkins S, Schulz G (eds) *Ullman's encyclopedia of industrial chemistry*, 5th edn, vol A17. Wiley-VCH, Weinheim. pp 363–376.

Pannu NS and Read RJ. 1996. Improved structure refinement through maximum likelihood. *Acta Cryst. A52*:659-668.

Patricelli MP, Cravatt BF. 2000. Clarifying the catalytic roles of conserved residues in the amidase signature family. *J Biol Chem 275*:19177-19184.

Pereira RA, Graham D, Rainey FA, Cowan DA. 1998. A novel thermostable nitrile hydratase. *Extremophiles 2*:347-357.

Pertsovich SI, Guranda DT, Podchernyaev DA, Yanenko AS, Svedas VK. 2005. Aliphatic amidase from *Rhodococcus rhodochrous* M8 is related to the nitrilase/cyanide hydratase family. *Biochemistry (Mosc) 70*:1280-1287.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera: a visualization system for exploratory research and analysis. *J Comput Chem 25*:1605-1612.

Piotrowski M, Schonfelder S, Weiler EW. 2001. The Arabidopsis thaliana isogene NIT4 and its orthologs in tobacco encode beta-cyano-L-alanine hydratase/nitrilase. *J Biol Chem 276*:2616-2621.

Ramachandran GN, Ramakrishnan C, Sasisekharan V. 1983. Stereochemistry of polypeptide chain configurations. *J Mol Biol 7*:95-99.

Rawlings ND, O'Brien E, Barrett AJ. 2002. MEROPS: the protease database. *Nucleic Acids Res 30*:343-346.

Read RJ. 1999. Detecting outliers in non-redundant diffraction data. *Acta Crystallogr D Biol Crystallogr 55*:1759-1764.

Read RJ. 2001. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr D Biol Crystallogr 57*:1373-1382.

Richardson JS. 2003. All-atoms contacts: A new approach to structure validation. In *Structural Bioinformatics* (Philip EB & Weissig H. Eds.). Wiley-Liss, Inc. pp 303-317.

Rossmann MG. 1990. The molecular replacement method. *Acta Crystallogr A 46 (Pt 2)*:73-82.

Sakai N, Tajika Y, Yao M, Watanabe N, Tanaka I. 2004. Crystal structure of hypothetical protein PH0642 from *Pyrococcus horikoshii* at 1.6A resolution. *Proteins 57*:869-873.

Sawyer DT, Sugimoto H, Calderwood TS. 1984. Base ($O^{\cdot-2}$, 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.

Schrag JD, Cygler M. 1997. Lipases and α/β hydrolase fold. In: Rubin B, Dennis EA, editors. *Methods Enzymol*, vol 284. New York: Academic Press; pp 85–107.

Schwarzenbacher R, Godzik A, Grzechnik SK, Jaroszewski L. 2004. The importance of alignment accuracy for molecular replacement. *Acta Crystallogr D Biol Crystallogr 60*:1229-1236.

Suh I-H, Kim K-J, Choo G-H and Lee J-H. 1963. The asymmetric unit of X-ray intensity data of the seven crystal systems. *Acta Cryst. A49*:369-371.

Sewell BT, Thuku RN, Zhang X, Benedik MJ. 2005. Oligomeric structure of nitrilases: effect of mutating interfacial residues on activity. *Ann N Y Acad Sci 1056*:153-159.

Skouloubris S, Labigne A, De Reuse H. 1997. Identification and characterization of an aliphatic amidase in Helicobacter pylori. *Mol Microbiol 25*:989-998.

Soubrier F, Levy-Schil S, Mayaux JF, Petre D, Arnaud A, 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.

Stelkes-Ritter U, Wyzgol K, Kula MR. 1995. Purification and characterization of a newly screened microbial peptide amidase. *Appl Microbiol Biotechnol 44*:393-398.

Stevenson DE, Feng R, Storer AC. 1990. Detection of covalent enzyme-substrate complexes of nitrilase by ion-spray mass spectroscopy. *FEBS Lett 277*:112-114.

Storoni LC, McCoy AJ, Read RJ. 2004. Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr 60*:432-438.

Taylor G. 2003. The phase problem. *Acta Crystallogr D Biol Crystallogr 59*:1881-1890.

Terwilliger T. 2004a. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat 11*:49-52.

Terwilliger TC. 2004b. Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Crystallogr D Biol Crystallogr 60*:2144-2149.

Terwilliger TC. 1999. Reciprocal-space solvent flattening. *Acta Crystallogr D Biol Crystallogr 55*:1863-1871.

Terwilliger TC. 2001. Map-likelihood phasing. *Acta Crystallogr D Biol Crystallogr 57*:1763-1775.

Thiery A, Maestracci M, Arnaud A, Galzy P. 1986. Acyltransferase activity of the wide spectrum amidase of *Brevibacterium* sp R312. *Journal of General Microbiology 132*:2205-2208.

Tickle IJ, Laskowski RA, Moss DS. 1998. Rfree and the Rfree ratio. I. Derivation of expected values of cross-validation residuals used in macromolecular least-squares refinement. *Acta Crystallogr D Biol Crystallogr 54*:547-557.

Tong L. 1993. *REPLACE*, a suite of computer programs for molecular-replacement calculations. *J. Appl. Cryst. 26*:748-751.

Uppernberg J, Hansen MT, Patkar S, Jones TA. 1994. The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida Antarctica*. Structure 2: 293-308.

Vagin A and Teplyakov A. 1997. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst. 30:*1022-1025.

Weiss MS. 2001. Global indicators of X-ray data quality. *J. Appl. Cryst. 34*:130-135.

Westhead DR, Slidel TW, Flores TP and Thornton JM. 1999. Protein structural topology: automated analysis and diagrammatic representation. *Protein Sci. 8:* 897-904.

Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC. 1999. Visualization and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogens. *J Mol Biol 285*:1711-1733.

Wyatt J, 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, 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, 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, Komatsu K. 1990. Production of S-(+)-ibuprofen from a nitrile compound by Acinetobacter sp. strain AK226. *Appl Environ Microbiol 56*:3125-3129.

Yano JK, Koo LS, Schuller DJ, Li H, Ortiz de Montellano PR, Poulos TL. 2000. Crystal structure of a thermophilic cytochrome P450 from the archaeon Sulfolobus solfataricus. *J Biol Chem 275*:31086-31092.

Zalkin H, Smith JL. 1998. Enzymes utilizing glutamine as an amide donor. *Adv Enzymol Relat Areas Mol Biol 72*:87-144.