Helical structures of the cyanide degrading enzymes from *Gloeocercospora sorghi* and *Bacillus pumilus* providing insights into nitrilase quaternary interactions

Margot Petra Scheffer

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

Supervisor:

Associate Professor Trevor Sewell

February 2006

KEYWORDS

Bacillus pumilus

Cyanide dihydratase

Cyanide hydratase

Docking

Electron microscopy

Gloeocercospora sorghi

Iterative Helical Real Space Reconstruction (IHRSR)

Nitrilase

3D reconstruction

Protein

Structure

ABSTRACT

Nitrilases (EC 3.5.5.1) are important industrial enzymes which convert nitriles into their corresponding acids or amides through a common (Glu/Asp)-Lys-Cys catalytic triad. The fungal cyanide hydratases (CHT) are a subgroup of this family converting cyanide to formamide whereas the bacterial cyanide dihydratases (CynD) convert cyanide to formic acid. The nitrilases from *Bacillus pumilus* and *Gloeocercospora sorghi* share only 30% sequence identity, but both enzymes are found in subfamily two with the fungal CHT's forming a group deeply nested within the clade of bacterial nitrilases. Atomic resolution structures of several members of the nitrilase superfamily have been determined but these differ from microbial and fungal nitrilases in that they do not form typical large homo-oligomeric complexes. Although sequence analysis reveals that the central core of these proteins is conserved, the spiral forming nitrilases differ from their crystallized homologues by two insertions of between 12 and 14 amino acids and a C-terminal extension of up to 35 amino acids, and thus these residues are implicated in forming the interfaces responsible for oligomerisation.

In this study, the helical structures of the CynD from *Bacillus pumilus* and the CHT from *Gloeocercospora sorghi* were solved using the method of Iterative Helical Real Space Reconstruction (IHRSR). High-resolution unidirectional metal shadowing showed that the extended fibres of CHT and CynD are unequivocally left handed helices. This was useful in assigning the handedness of the three-dimensional structures, and allowed the $\Delta\varphi$ (rotation per subunit) parameter to be constrained to a negative value. Two negative stain reconstructions of wild type *G. sorghi* CHT converged to the same helical parameters ($\Delta\varphi$ = -77, and Δz = 15) from different data sets, one taken on film and the other taken with a CCD.

The FSCs of the two structures were shown to correlate to 16Å at the fixed-value threshold of 0.5. An active mutant of *G. sorghi* CHT, namely R87Q was also solved and was shown to correlate to a lesser degree (33Å) with the WT structure, indicating inherent differences between the structures. A cryo reconstruction was also performed but a resolution of only 29Å was achieved due to the large amount of images that were excluded from back projection and the low signal-to-noise ration of the images. The structure of the CynD from *Bacillus pumilus* converged on similar helical parameters but by visual inspection appeared to have different features to the *G. sorghi* structure.

Even though the negative stain *G. sorghi* CHT structures converged to the same solution, it was impossible to reconcile the diffraction pattern of the negative stain structure with the information obtained in the shadowing experiment and the cryo reconstruction due to an anomalous layer line present in the power spectrum of Bessel order four (layer line two). The average power spectrum of the rotationally aligned images was however identical to the power spectrum generated from the reconstructed volume, indicating that the structure of the *G. sorghi* CHT was correct as far as the available data was concerned. Nevertheless, caution was taken upon interpretation of the structure since the available shadowing and cryo data did not agree with that of negative stain. Through closer inspection and analysis of the layer line data it was possible to put forward a 'flattening in negative stain' hypothesis to explain the origin of the layer line. It is possible that the larger diameter of the *G. sorghi* fibre caused the structure to collapse under the pressure of the negative stain while the more rigid structure of the *B. pumilus* fibre with its smaller diameter was able to withstand the dehydration forces.

Both the *G. sorghi* CHT and *B. pumilus* CynD structures were docked with homology models of dimers with the A surface preserved. Models were sometimes docked ambiguously but it was premised that the carboxy-terminus should face inwards since unoccupied densities filled by insertions and available knowledge of ion-pair interactions could be more easily rationalized in this manner. Docking thereby allowed the interpretation of biological interactions and provided justifications for various biochemical data. The interacting residues leading to spiral formation were speculated upon and the C and D surfaces involved in oligomerisation were identified.

DECLARATION

I declare that ***Helical structures of the cyanide degrading enzymes from Gloeocercospora sorghi and Bacillus pumilus providing insights into nitrilase quaternary interactions*** is my own work, that it has not been submitted for any degree or examination in any other university, and that all sources I have used or quoted have been indicated and acknowledged by complete references.

Margot Petra Scheffer 17 February 2006

Signed: Signed by candidate

Signature removed

ACKNOWLEDGEMENTS

I praise you, Father, Lord of heaven and earth for all that you have done for me, a pitiful sinner. I'm sorry Christ Jesus for the things I had prioritized, when I knew that it should have been all about you, Lord. You took upon my burdens Lord, and you gave me rest so that I could be set free. To God be the grace, honour and the power for all eternity.

To my parents, thank you for the support and prayers that you have tirelessly given me. Knowing that you are behind me has guided me through all that I have accomplished in life. You are the wind beneath my wings.

To my supervisor and second dad, Trevor Sewell, thank you for your great kindness. The concern that you have for the wellbeing of your students has made me feel like I could always turn to you when things went wrong. Thank you for the immense amount that you have taught me in the last two years.

To my colleagues, Jason van Rooyen, Ndoria Thuku, Itai Chitapi, Serah Kimani, Samuel Kwofie, thank you all for your guidance and friendship throughout this period. Thank you to the staff at the electron microscope unit, Mohammed Jaffer, Miranda Waldron, Sean Karrim and Brendan Price for their willing assistance. Many thanks to Brandon Walker for his collaboration, this project could not have been completed without him. May God bless you all and may you prosper in all your future endeavours.

I especially thank the caffeine source in my life Tim Frouws whose knowledge and friendship during coffee times was invaluable.

LIST OF ABBREVIATIONS:

ATCC	American Type Culture Collection
B. pumilus	*Bacillus pumilus*
CC	score cross-correlation score
CCD	charge coupled device
CHT	cyanide hydratase
CynD	cyanide dihydratase
DCase	*N*-carbamyl-*D*-amino acid amidohydrolase
DPR	differential phase residual
EM	electron microscopy/microscope
FFT	Fast Fourier transform
FSC	Fourier shell correlation
G. sorghi	*Gloeocercospora sorghi*
HGT	horizontal gene transfer
IHRSR	Iterative helical real space reconstruction
KCN^-	potassium cyanide
LB	Luria broth
MW	molecular weight
$NAD^{(+)}$	Nicotinamide Adenine Dinucleotide
ORF	open reading frame
PDB	protein databank
P. stutzeri	*Pseudomonas stutzeri*
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SNR	signal-to-noise ratio
TEM	TEM transmission electron microscopy/microscope
WT	wild type

TABLE OF CONTENTS

SECTION 1 – INTRODUCTION

CHAPTER 1.1

SUBJECT REVIEW

1.1.1 Nitrilases and their industrial applications

Hydrolysis and condensation reactions of carbon-nitrogen bonds are fundamental to biology. Formation and cleavage of the peptide bond is the most described reaction of this class but there are other equally important carbon-nitrogen bond reactions which are not as well-understood. These types of linkages include organic cyanides, known as nitriles (R-C≡N), acid amides (R-C(=O)-NH_2), secondary amides (R-NH-C(=O)-NH_2) and *N*-carbamyl amides (R-NH-C(=O)-NH_2) (Brenner, 2002).

Metal-cyanide complexes such as sodium cyanide (NaCN) and potassium cyanide are used in many industrial processes which leads to large amounts of toxic cyanide waste. In mining, cyanide solutions are used to separate gold and silver from their ores. A large amount of cyanide waste is also produced from electroplating plants and coal-fueled gas facilities (Knowles, 1988). Cyanide is a potent cellular toxin that inhibits the activity of cytochrome-c-oxidase, catalase and a number of other enzymes. It is therefore important to develop economically viable means to decontaminate cyanide waste resulting from industrial processes (Banerjee et al., 2002).

Biological remediation has the advantage over conventional techniques (such as alkaline chlorination, electrolytic decomposition and ozonation) in that it does not produce large quantities of chemicals that adversely affect the environment. Remediation plants utilizing cyanide degrading enzymes can be used to remove concentrated or very dilute cyanide wastewaters, and are relatively inexpensive to maintain once the infrastructure is in place (Banerjee et al., 2002; Knowles, 1988).

Current technology uses immobilized cell-free fraction packed-cell columns to continuously detoxify cyanide. ICI Biological Products markets a preparation known as CYCLEAR® consisting of spray-dried fungal mycelia from *Fusarium lateritium* (Knowles, 1988; Cluness et al., 1993). Novo Industries produce CYANIDASE® which is a granular

form of Alcaligenes xylosoxidans subsp. denitrificans used in a recirculation fixed bed reactor (Basheer et al., 1992). The potential exists for the use of purified enzymes for on site cyanide remediation since they may be engineered to tolerate varying levels of pH and temperature which the organisms themselves may be sensitive to.

Many opportunities exist for the use of nitrilases in commercial processes which is why many companies such as DuPont, Lonza, Dow, Diversa, BASF, and DSM have been extensively investigating nitrile biocatalysis and its industrial applications (Brady et al., 2004). For example, the nitrilase of *Comomonas testosteroni* sp. converts the dinitrile adiponitrile into cyanovaleric acid and adipic acid, which are intermediates in the production of nylon. Also, a number of acids are produced from corresponding nitriles by the nitrilase of *R. rhodocrous* J1 including acrylic acid, which is used in the production of acrylic fibres (Brady et al., 2004), the vitamin niacin (nicotinate) and pyrazinoic acid, an antibiotic used in the treatment of tuberculosis (Nagasawa et al., 2000). A particular application which is under investigation, is the use nitrilases to selectively hydrolyze nitriles with labile functional groups which cannot be hydrolyzed by conventional strong acid or alkaline methods (Brady et al., 2004). The nitrilase of *Klebsiella ozaenae* is specific for the herbicide bromoxynil and has been expressed in transgenic plants to confer resistance against this herbicide (Duke, 2005; Banerjee et al., 2002).

Nitrilases, like most other biological catalysts are enantioselective which makes them attractive as industrial enzymes. Currently, nitrilases are used to manufacture active enantiomers of (R)-mandelic acid, (R)-3-chloromandelic acid, (Brady et al., 2004) (S)-phenyllactic acid and (R)-3-hydroxy-4-cyano-butyric acid which is an important intermediate in the synthesis of the cholesterol-lowering drug atorvastatin calcium(Banerjee et al., 2002; O'Reilly and Turner, 2003). On the subject of pharmaceutical design, it is known that some types of nitrilase-like proteins are only found in prokaryotes, thus making them suitable targets for antibiotic development (Pace and Brenner, 2001).

In this study the structure of two nitrilases (a cyanide hydratase from *Gloeocercospora sorghi* and a cyanide dihydratase from *Bacillus pumilus*) are elucidated in the hope of gaining further insights into the quaternary interactions of these cyanide degrading enzymes. Through analysis of the structures a better understanding of their function and possible biomolecular engineering prospects will be achieved.

1.1.2 The nitrilase superfamily (The C-N hydrolases)

Nitrilases are a family of proteins forming one of the thirteen branches of the carbon-nitrogen hydrolases. This superfamily is sub-divided into groups based on similarity in sequence and the occurrence of fusion domains (Pace and Brenner, 2001). The families are further categorized by bond specificity and reaction mechanism which is known or deduced for nine of the thirteen branches. The nitrilase branch is the only group with members that perform nitrile hydrolysis - the other eight are reported to have amidase activity or condense acyl chains to amino groups (Robertson et al., 2004; Pace and Brenner, 2001). Nitrile hydratases, which are nitrile degrading enzymes dependant on a metal co-factor (Huang et al., 1997), are not members of the nitrilase superfamily. These enzymes, which are not to be confused with the fungal cyanide hydratases, have a different catalytic mechanism for converting nitriles to their corresponding amides (Pace and Brenner, 2001; Sewell et al., 2003).

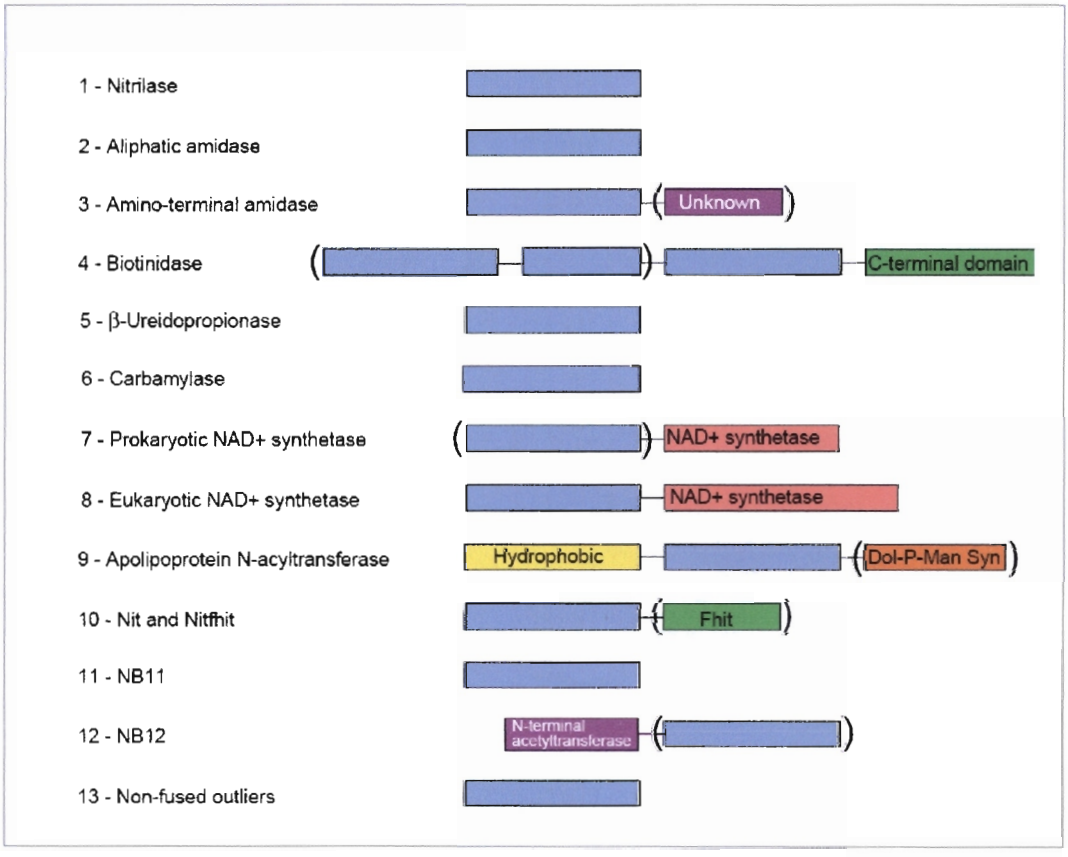


Figure 1.1.1: Domain structures for the 13 branches of the nitrilase superfamily. Additional domains are found in members of seven branches. Parentheses denote domains found in only some members of the branch. Reproduced from (Pace and Brenner, 2001)

Seven of these branches contain nitrilase-related proteins with fusion domains which may either serve to change the localization of the protein and possibly link proteins involved in cellular signalling or they may be involved in ammonia consumption and production. Some of the fusion events are considered as "Rosetta Stone" fusions - meaning that separate polypeptides are fused in order to co-ordinate biochemical reactions or cellular functions (Pace and Brenner, 2001).

Branches two and three contain the aliphatic amidases, the major difference between them being that branch two amidases hydrolyse small-molecule substrates such as the carboxamide sidechain of asparagine, whereas branch three enzymes hydrolyze asparagine or glutamine residues at the N-terminus of a polypeptide chain (Brenner, 2002).

Branch four of the superfamily consists of the biotinidases, as well as the vanins. These two clades have a unique carboxyterminal domain which is exclusively found in this branch (Pace and Brenner, 2001). Biotinidase is the enzyme responsible for recycling biotin (vitamin H) from lysine-conjugated biotin (biocytin) and from lysyl biotinylated proteins (Swango et al., 2000). Other possibilities for the role of biotinidase in biotin metabolism have been discovered; it was found that biotinidases have biotinyl-transferase activity, in addition to biotinidase hydrolase activity. The enzyme is implicated in the regulation of protein transcription since it is known to transfer biotin to histones (Hymes and Wolf, 1999; Stanley et al., 2004). The C-terminal of the biotinidase sequence is lacking in nitrilases and amidases which suggests that the biocytin-biotin-binding site of biotinidase is to be found in this domain (Wolf and Jensen, 2005). Vanins are related to the biotinidases which were first characterized as antigens in thymic homing and adhesion of hematopoietic precursor cells (Brenner, 2002).

The β-ureidopropionases of branch five are similar to branch six enzymes in that they catalyse the double hydrolysis of *N*-carbamyl compounds to produce the corresponding primary amine, carbon dioxide and ammonia. The branch five enzymes are responsible for pyrimidine base catabolism. Branch six contains bacterial *N*-carbamyl-D-amino acid amidohydrolases including the enzyme from *Agrobacterium*, namely the carbamylase *N*-carbamyl-D-amino acid amidohydrolase which is one of the four crystal structures available to date (Nakai et al., 2000). The other enzyme for which a crystal structure exists, which will also be discussed in further detail later, may be found in branch 10. This is the Nit

protein which is a fused Rosetta Stone protein with Fhit in invertebrates and is co-ordinately expressed with separate Fhit proteins in mammals (Pace et al., 2000). The hypothetical protein pH0642 from Pyrococcus Horikoshii which was crystallized in 2004 (Sakai et al., 2004), is also a member of branch 10.

In eukaryotes, NAD^+ synthetases (group 8) are always fused with a nitrilase-related domain, however in prokaryotes (group 7) this is only the case for some members of the family. Eukaryotic NAD synthetases can use glutamine as an ammonia source whereas some prokaryotic (bacterial) NAD^+ synthetases are able to use ammonia but not glutamine (Brenner, 2002). It is now known that glutamine amidotransferase activity of NAD^+ synthetases depends on an amino-terminal nitrilase domain (Bellinzoni et al., 2005; Pace and Brenner, 2001). The association of the nitrilase domain with the NAD^+ synthetases is thought to have arisen out of the need for a shuttling mechanism whereby the ammonia produced by the nitrilase domain is "couriered" to the NAD^+ synthetase active site (Brenner, 2002).

Branch nine contains the apolipoprotein N-acyltransferases which always have a hydrophobic amino-terminal domain. These enzymes catalyse the acyl condensation (reverse amidase) reactions *in vivo*. The nominal function of the branch nine enzymes is post-translational modification however this is only substantiated by one member which is fused to an apparent dolichol phosphate mannose synthetase. The basis reaction involves addition of a palmitoyl group to an N-terminal diacylglycerol-modified cysteine residue (Brenner, 2002). Branch 12 enzymes are also predicted to function in post-translational modification of proteins since they are sometimes fused to a homolog of an amino-terminal acetyltransferase, namely RimI (Pace and Brenner, 2001).

The most conserved motif is located toward the N-terminus of these proteins – it contains the active site glutamic acid that follows a very hydrophobic interior β-strand (β2). In fact, a feature of this super-family is that the conserved regions are particularly hydrophobic, due to the tightly packed β-sheet with the hydrophobic β-strands located in the core shielded from the protein surface (Bork and Koonin, 1994; Nakai et al., 2000).

1.1.3 The Nitrilase Family

Nitrilases are seemingly rare in bacteria since only 19 out of over 150 sequenced bacterial genomes contain nitrilase genes (Podar et al., 2005). Yet recently 137 unique nitrilases were identified through functional screening of over 600 biotope-specific environmental DNA libraries, for use in industrial biocatalysis applications. This indicates that it is more efficient to screen a large number of samples from different environments than deep coverage of one or a few environments in order to uncover the sequence space of a gene family (Robertson et al., 2004; DeSantis et al., 2002).

Through this screening and biochemical characterisation, Robertson *et al.* revealed the presence of six major sequence clades in the nitrilase family which they refer to as sub-families. Four of the sub-families have no representatives in known bacterial species. A number of enzymatic properties such as substrate specificity and enantioselectivity are specific to subfamilies and in some cases the enzyme activities correlate with the ecology of the environmental samples (Robertson et al., 2004).

Since a correlation exists between enzymatic properties and gene sequence clades or sub-families, it may be implied that nitrilases have diverged functionally in order to acquire a biological role necessary to allow that microbe to occupy a specific ecological niche (Podar et al., 2005). Some authors believe that the functional diversity within the superfamily is reflective of a process of convergent evolution towards a common mechanism (Bork and Koonin, 1994) but considering the lack of evidence it seems that the nitrilase fold arose only once. Not much is known about the role of gene duplication and natural selection of the nitrilases but some hypotheses have recently been put forward by Podar *et al.* which makes the prospect of a common ancestral fold which has undergone diversification and neofunctionalization all the more likely (Podar et al., 2005).

Resulting from selection pressure, functional divergence of an enzyme is often characterized by the fact that the gene is accompanied by an operon encoding enzymes with related metabolic activities. By analysis of the genes neighbouring nitrilase genes, patterns of conserved gene clustering in relation to biochemical and phylogenetic data were discovered. The phylogeny of genes from cultivated bacteria shows that 18 nitrilases belong to subfamilies one and two. The sequence similarity among the enzymes in

subfamily one is relatively high, ranging from 50-70% pairwise identity. Subfamily two has a range of 30-50% pairwise identity (Podar et al., 2005).

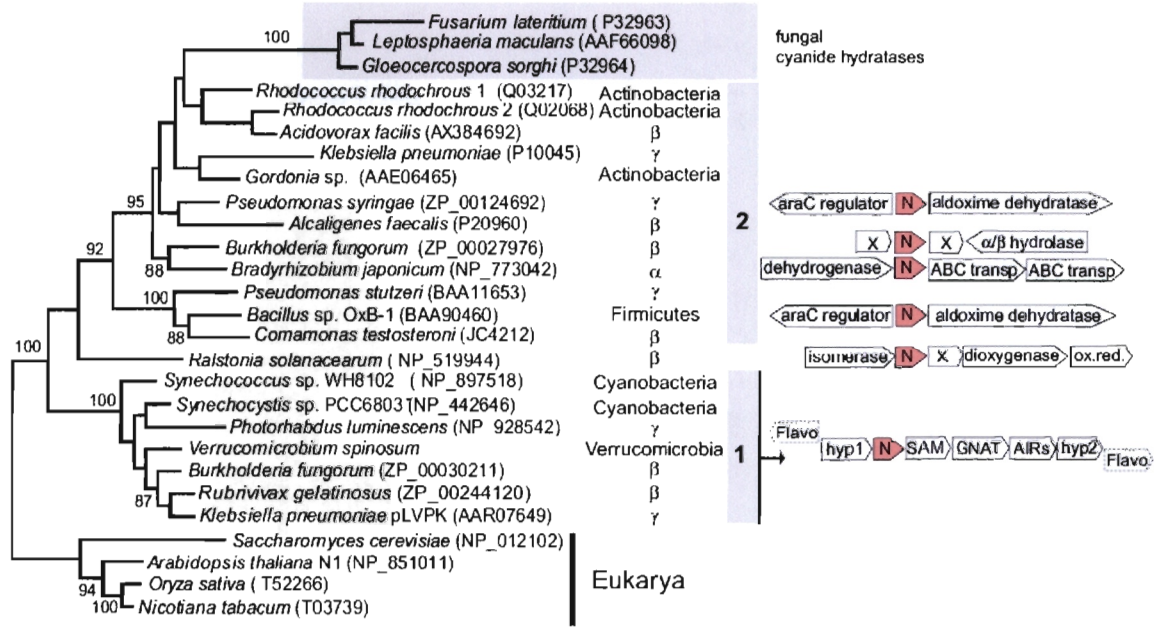


Figure 1.1.2: Maximum likelihood tree of nitrilases in subfamily one and two from known bacterial and fungal species. Where sequence information was available, the organization of gene clusters around the nitrilase ORF is shown on the right (Podar et al., 2005).

It has been suggested that the nitrilase superfamily emerged prior to the separation of plants, animals and fungi, radiated into families and thereafter spread laterally through horizontal gene transfer (HGT) to bacteria and archaea (Pace and Brenner, 2001). Several lines of evidence revealed in the nitrilase taxonomy support the theory of ancient gene duplication events or HGT for the acquisition of nitrilases. The first of which is that the classification of bacterial nitrilase genes does not reflect the phylogeny of their host organisms. Secondly, in several species that have two nitrilases (*Klebsiella pneumoniae*, *Burkholderia fungorum* and all *Pseudomonas* species), the genes are assigned to different clades. Lastly, in *Photorhabdus*, a large polyketide synthase pathway is interrupted by a nitrilase cluster and in other species the nitrilase ORF is associated with mobile DNA elements such as transposases. These facts all suggest ancient gene duplication or HGT (Podar et al., 2005).

According to the recent maximum likelihood classification of nitrilase ORFs (Fig 1.1.2), both the cyanide dihydratase from *Bacillus pumilus* and the cyanide hydratase from *Gloeocercospora sorghi* are found in subfamily two, which is characterized by (*R*)-

enantioselectivity, especially on the substrate mandelonitrile. The fungal CHT's form a group deeply nested within the clade of bacterial nitrilases, suggesting an ancient horizontal gene transfer event followed by neofunctionalization of the enzyme (Podar et al., 2005).

To test if positive selection could act as a driving force for nitrilase functional diversification, a maximum likelihood method based on codon substitution models was used to determine which sites are positively selected. Branch two is a basal lineage which has lost the gene cluster associations of subfamily one. It is the diversified branch which may represent the first of the neofunctionalized nitrilases. A positive selection pressure is detected in this lineage, particularly residues T41, Q157, Y184, N200, Q203 and R284. Residues T41, Q157, Y184 are found to be buried within the protein interior close to the catalytic triad. It is likely that these amino acids are involved in maintaining the active site conformation. Residues N200, Q203 and R284 are found on the dimer interface (A surface) which has been shown in D-NCase to form a hydrophobic pocket responsible for preserving the dimer interface (Chen et al., 2003; Wang et al., 2001).

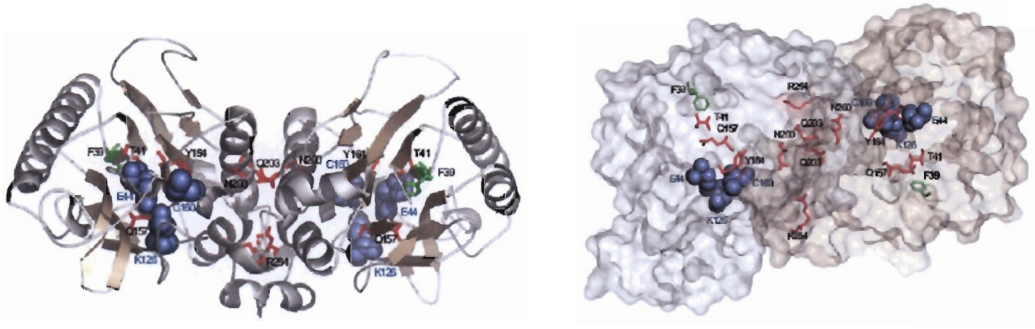


Figure 1.1.3 A three dimensional model of the nitrilase dimer of an environmental isolate known as 1A21. The catalytic triad is shown in blue spheres and the residues under positive selection are shown in red. Residue 39, which changes from a leucine to a phenylalanine outside the Nit1C cluster is shown in green *(Podar et al., 2005)*

1.1.4 Structure of the Nitrilase Dimer

Despite the intensive efforts of the Steve Almo in collaboration with the Diversa group (pers comm) to crystallize several hundred nitrilases, none have been susceptible to this process and hence the structure of a nitrilase has not yet been elucidated. Currently, only five crystal structures of the superfamily members exist, namely the worm NitFhit protein (Pace et al., 2000), the bacterial N-Carbamyl-D-Amino-Acid Amidohydrolase (Nakai et al.,

2000) and several structures of mutants and substrate bound bacterial *N*-carbamoyl-D-amino acid amidohydrolase (Wang et al., 2001; Chen et al., 2003). The last two structures to be solved, the yeast Nit3 protein (Kumaran et al., 2003) and a hypothetical protein pH0642 from *Pyrococcus Horikoshii* (Sakai et al., 2004), have been due to the efforts of the Structural Genomics Consortium.

Table 1.1: PDB accession codes for crystal structures of members of the Nitrilase superfamily

PDB accession number	Protein Name	Author	Date
1EMS	NitFhit from *Caenorhabditis elegans*	Pace, Hodawadekar, *et al.* (Pace et al., 2000)	2000
1ERZ	Dcase (N-Carbamyl-D-Amino-Acid Amidohydrolase) from *Agrobacterium sp.*	Nakai, Hasegawa, *et al.* (Nakai et al., 2000)	2000
1F06	N-Carbamoyl-D-Amino-Acid Amidohydrolase from *Agrobacterium radiobacter*	Wang, Hsu, *et al.* (Wang et al., 2001)	2001
1F89	Nit3 from *Saccharomyces Cerevisiae*	Kumaran, Eswaramoorthy, *et al.* (Kumaran et al., 2003)	2003
1J31	Hypothetical Protein pH0642 from *Pyrococcus Horikoshii*	Sakai, Tajika, *et al.* (Sakai et al., 2004)	2004

Using a BLAST protein sequence alignment, it is evident that there is only 20% identity between the DCase sequence and that of the nitrilases from *G. sorghi* and *B. pumilus*. The highest sequence identity is found between Nit3 and *G. sorghi* with a value of 28%. Sequence identity between *G. sorghi* and *B. pumilus* is 34% with 52% positives (amino acid substitution of a similar type of residue).

On the other hand, when searching for structural alignment using the *G. sorghi* and *B. pumilus* sequences in GenTHREADER (Jones, 1999), DCase and the Nit domain of the NitFit were found to have structural homology with a score of higher than 0.98 against both queries. The structural alignments of alpha carbon atoms are shown in figure 1.1.4. When the pdbs 1erz and 1ems are superimposed using ALIGN (Cohen, 1997), the central core of β-sheets and other regions including the three active site residues were well conserved. Using a cut-off RMSD value of 1.5Å, 111 alpha-carbon atoms were well matched.

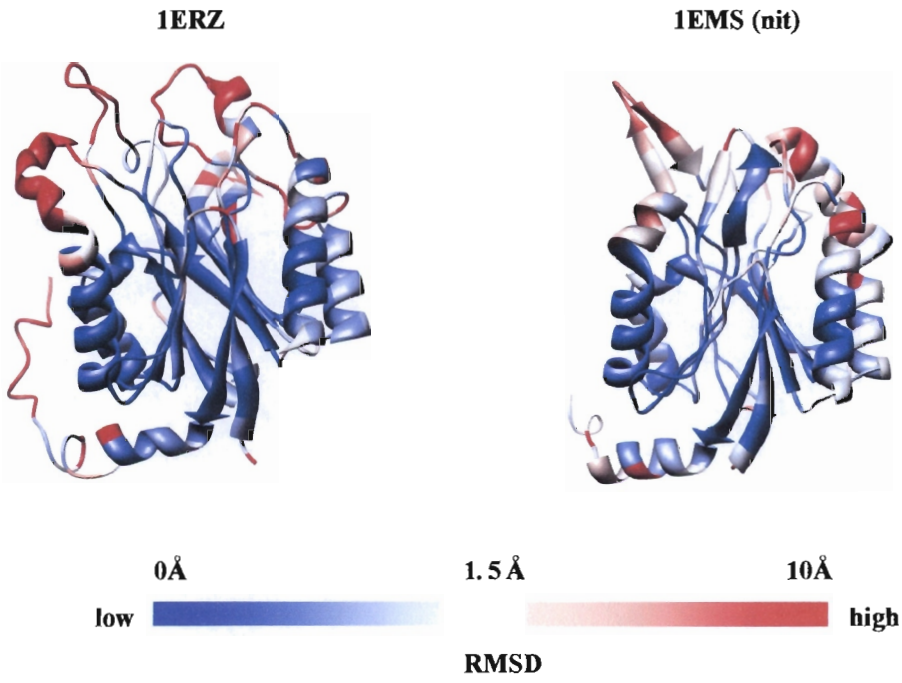


Figure 1.1.4: Values of RMSD (Root Mean Squared Deviation) of Cα carbons calculated using the program ALIGN (Cohen, 1997). Average RMSD between these two structures is 1.583Å.

Although no structural information is available for nitrilases yet, it is known that these enzymes share a common structural framework with their distant homologues in the nitrilase superfamily. Using three of the crystallographic structures, it is possible to generate a model of the fourth structure which concurs to within 1.5Å of the known crystal structure (pers comm. Ndoria Thuku), showing that the protein fold is conserved enough to be able to model the core structure to a reasonable degree of certainty. It is however, not possible to accurately model the two major insertions (one prior to the α-helix NH2 and the other which lengthens the β-loop between NS9 and NS10) and C-terminal extension of the *G. sorghi* and *B. pumilus* sequences relative to the Nit domain.

According to protein classification by the CATH database, a monomer of DCase has a four-layer sandwich architecture. The sandwich is made up of two six-stranded β-sheets flanked by helices α1 and α3 on one side and helices α5 and α6 on the other. The β-sheets have three parallel strands at the centre which are bordered by three anti-parallel strands. Predominantly hydrophobic interactions occur between the α-helices and β-sheets and compact packaging of other such hydrophobic residues are responsible for formation of the interface between β-sheets.

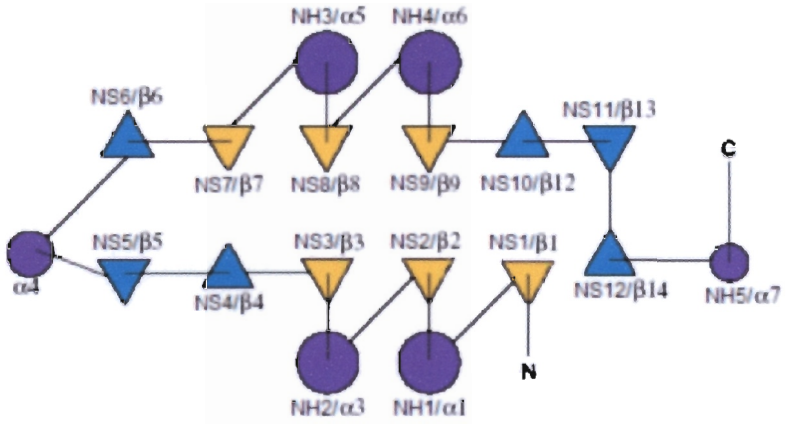


Figure 1.1.5: Topological diagram of the protein fold of a monomer showing corresponding numbering of secondary structural elements. Purple circles represent α-helices, yellow triangles represent parallel β sheet and blue triangles represent the anti-parallel β-sheets. Pace et al. (Pace et al., 2000) use the numbering scheme NSx, NHx whereas Nakai et al. (Nakai et al., 2000) implement the αx, βx system. In the DCase structure, amino acids 61-66 (between β2 and α3) are designated α2 and two very small β-sheets, consisting of three and two amino acids respectively, are designated β10 and β11. Also, in the Nit domain of the NitFhit structure, an α-helix corresponding to α4 does not exist. This accounts for the discrepancy in numbering between the Pace and Nakai structures.

Except for two monomeric exceptions, nitrilases are always found as dimers or oligomeric structures comprised of an even number of subunits (Banerjee et al., 2002). Evidence exists that spirals are made up of dimers since the dimers of certain members of the family are reported to form multimeric complexes upon incubation with substrate (Nagasawa et al., 2000; Banerjee et al., 2002). Although DCase crystallizes as a homotetramer, it appears that in vivo it is mostly seen as a dimer or seldomly a trimer consisting of identical subunits of 34kDa (Nakai et al., 2000).

In glutaraldehyde crosslinking experiments of *B.pumilus* cyanide dihydratase, a very faint band of trimer as well as a single band corresponding to the tetramer conformation was also observed at increasing concentrations of glutaraldehyde. However, three closely spaced bands of similar intensity were observed in the molecular weight region corresponding to that of a dimer which appear even at very low concentrations of glutaraldehyde indicating that this is the predominant form of subunit interaction (Berman, 2003).

It has been shown that the electron density maps of the terminating nitrilase spirals of *P. stutzeri* (Sewell et al., 2003) and *B.pumilus* (Jandhyala et al., 2005) have global 2-fold axes with the dyad axis of the complex coinciding with the dyad axis of the dimer. This suggested that a dimer model rather than a monomer model should be used to fit these maps and it was thus postulated that either one of the associating surfaces (A surface or B

surface) found in the crystallographic tetramers is preserved. The decision on how to construct the dimer was based on analysis of the DCase and NitFhit crystal structures.

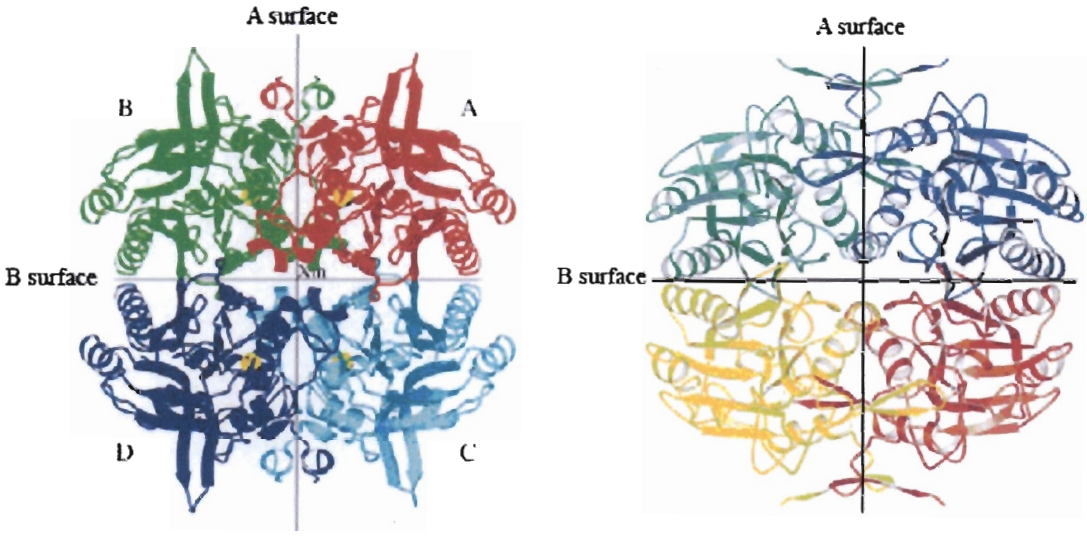


Figure 1.1.6: Schematic ribbon diagrams of the crystallized homotetramers of Dcase (Nakai et al., 2000) and the Nit domain of NitFhit (Pace et al., 2000).

The two contact surfaces of the crystallographic tetramer of the NitFhit structure, the Nit–Nit interfaces were termed north–south and east–west. These surfaces, which are almost perpendicular to one another, have previously been named A surface (east–west interface) and B surface (north–south interface) by Sewell *et al.* (Sewell et al., 2003), which is the nomenclature that shall henceforth be used.

Nakai *et al.* seem to accept the A surface as the dimer interface for several reasons. In DCase, the X_m and Y_m axes are non-crystallographic twofold axes whereas the Z_m axis, where the B surface is found, is a crystallographic twofold axis. Molecules A-C and B-D are therefore related by the crystallographic axis – the association of which is therefore much more likely to have been generated through crystal packing forces. Molecules A-B and C-D of the tetramer form tightly compacted dimers but the A-C and A-D molecules are packed loosely in both DCase and Nit tetramers (Wang et al., 2001; Nakai et al., 2000).

When considering each DCase monomer, the buried area between molecules A and B (2524Å) is much larger than that of the A-C (679Å) and A-D (377Å) interfaces. The majority of the interface is formed by helices α5 and α6 with their symmetry mates, and residues 280-303 of the C-terminus (α8) which surrounds α5 of the other monomer. It is

speculated by Nakai *et al.* that the C-terminus is involved in forming a linker between the monomers during dimer formation (Nakai et al., 2000). This dimer model with a preserved A surface is now accepted as the convention when modelling nitrilase structure based on sequence homology (Sewell et al., 2003; Podar et al., 2005).

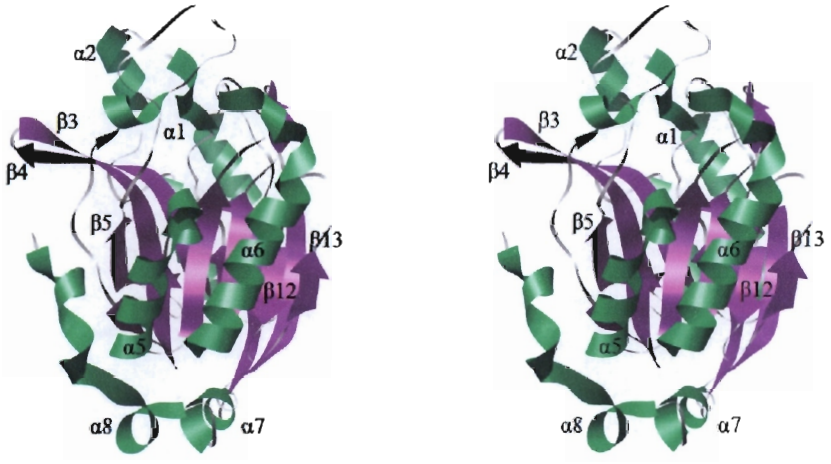


Figure 1.1.7: Stereoview of schematic ribbon diagrams depicting the protein fold of a monomer of DCase. Some secondary structure elements are labeled. α helices are shown in green and β strands are shown in purple.

Pace *et al.* (Pace et al., 2000) also favour the idea of a preserved A surface and provide several lines of evidence to support the idea. In the NitFhit structure, binding of heavy atoms locked the northern and southern hemispheres into different conformations but did not disturb the perfect east–west symmetry of the A surface. The Fhit domains also form dimers which are located at the north and south poles of the Nit tetramer, indicating that east-west dimers need to be preserved in order to retain the Fhit dimerization. The A surface of Nit, is formed by a four-helix bundle comprising of NH3 and NH4 with their symmetry mates (the equivalent of α5 and α6 in DCase). The final element of the A surface differs between DCase and the Nit domain of NitFhit. In Nit an anti-parallel β-interaction is formed by NS13 whereas in DCase the C-terminus forms a helix (α8) which is responsible for creating the dimer contacts.

As with DCase, the Nit dimer turns a four-layered α-β-β-α sandwich into an eight-layered α-β-β-α-α-β-β-α sandwich (Pace et al., 2000). Helices NH1 and NH2 are solvent-exposed on the external layers of the sandwich and are likely to be those mediating interactions in the process of oligomerization in multimeric nitrilases (Sewell et al., 2003; Sewell et al., 2005). Within each monomer, one edge of the β-sheet is enclosed by NH5 while the other

edge is exposed to solvent. Pace *et al.* detected substantial amino-acid conservation in their sequence alignment of several Nit domains and predicted that all Nit proteins would form dimers with the A surface preserved (Pace et al., 2000).

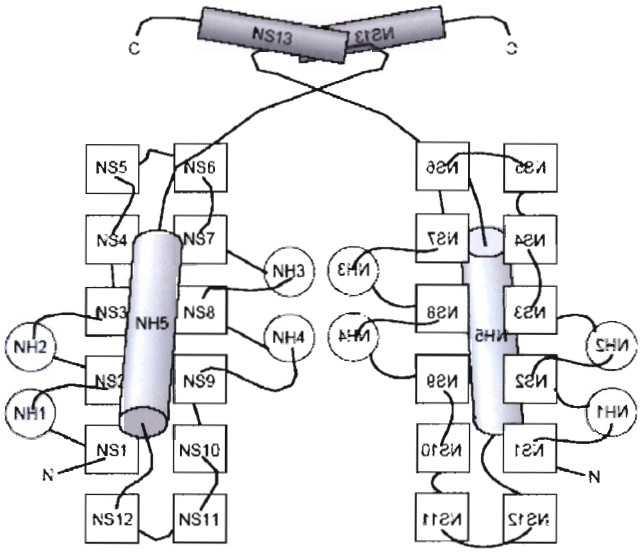


Figure 1.1.8: Topology diagram depicting the structure of the Nit domain dimer of NitFhit with the A surface preserved (Brenner, 2002; Pace et al., 2000). The symmetry axis is indicated by correspondingly flipped labels. N, amino terminus and C, carboxyl terminus. Surface A consists of the helices α5, α6 and α8 in DCase and helices NH3, NH4 and NS13 in Nit in such a way that an αββαααββα superstructure is formed from the two αββα sandwiches. Helices α6 cross near the centre at a ±70° angle and helices α5 cross each other near the end at a ±15° angle.

The B surface is formed by anti-parallel, β interactions involving strands NS11 and NS12 which doubles the width of the four β-sheets in the α-β-β-α-α-β-β-α sandwich from six strands to 12. Pace *et al.* express doubt over whether the north–south Nit interface (B surface) would be retained in all Nit proteins since some of the salt bridges that stabilize the B surface appear to be absent in homologous Nit sequences. They speculate that vertebrate and fungal Nit proteins may be dimers or tetramers, but not monomers (Pace et al., 2000).

The final pieces of evidence towards conservation of the A surface in the oligomeric nitrilases is presented by Sewell *et al* (Sewell et al., 2003). They rejected models based on preservation of the B surface for three reasons. They noted that upon docking an atomic model into the electron density map of *P. stutzeri* that there were large regions of unfilled density which could not be accounted for, particularly a bulge on the inner surface of the spiral. They noted a steric clash at the C-terminus involving NH5 of one dimer with NH3 and NS13 of a neighbouring dimer in the spiral. Also they noted some areas where the atomic model was not contained within the electron density map.

Besides the question of which surface is preserved, there is also the question of whether the dimer should face inwards or outwards. Sewell *et al.* argued that by retaining the A surface in an atomic model the curvature of the electron density including the bulge on the 2-fold axis on the inner surface of the spiral could be accounted for. This bulge was then occupied by the C-terminal structure corresponding to NS13 in the Nit domain of NitFhit. It was also argued that the oligomeric nitrilases have C-terminal extensions which could additionally contribute to the A surface interactions and could account for the large amount of density on the interior of the spiral. Secondary structure prediction (McGuffin et al., 2000) suggests that a model of the NS13 region should be modelled according to the β-sheet of Nit and not according to DCase where this region is helical. Secondary structure prediction of the C-terminus of *B. pumilus* and *G. sorghi* suggests that this extension is most likely to form a random coil or loop structure, but this is without taking dimerization interactions into effect.

The crystal structure of yeast Nit3 protein (1F89) was determined and shown to exist as a dimer both in the crystal and in solution. The protein fold of a monomer was shown to be similar to the other solved structures with a four-layer α-β-β-α sandwich. At the time when this structure was solved, it was only known as a yeast hypothetical protein with sequence similarity to CN hydrolases. It was selected in the New York Structural Genomics Consortium as representative of CN hydrolases at a time when no structures had been determined for members of the nitrilase family (Kumaran et al., 2003). Since then, other members of the family have been crystallized which also crystallize as dimers (Sakai et al., 2004). These proteins have similar topology and dimerization modes, which show that the dimers form an unusual eight-layer α-β-β-α-α-β-β-α structure (Vorwerk et al., 2001).

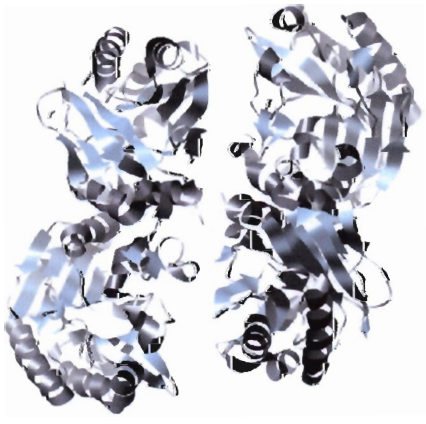


Figure 1.1.9: Crystallographically related dimers of the 1J31 hhypothetical Protein pH0642 from *Pyrococcus horikoshii.* These proteins do not crystallize as tetramers like 1ERZ and 1EMS.

There are lines of evidence, particularly from the DCase structure, that dimerization is required for enzyme function since residues involved in intermolecular interactions are close to the proposed active-site cleft. Mutagenesis showed and the crystal structure confirmed that strict geometric constraints of three histidine residues were required for D-NCAase catalysis. It was proposed that His129 and His144 are required to maintain a stable conformation of the catalytic cleft, whereas His215 was seen to be crucial for maintaining the tight dimer essential for the enzyme activity (Wang et al., 2001). This highlights the importance of the preservation of the dimeric interface 'A' in structure determination.

1.1.5 Oligomerisation

Members of the nitrilase superfamily which have been crystallized occur as homo-dimers or homo-tetramers but it is known that varying degrees of oligomerisation may occur. The microbial nitrilases exist as homo-oligomers of molecular weight greater than 300kDa (Berman, 2003). Low resolution single particle reconstructions showed that certain of these complexes exist as terminating spirals with a defined number of subunits. The cyanide dihydratase from *Pseudomonas stutzeri* ($CynD_{stu}$) for example exists as a 14 subunit structure (Sewell et al., 2005). On the other hand, the fungal nitrilases can form complexes of over a megadalton. The native CHT of *F. lateritium* was shown by gel-filtration and non-denaturing PAGE to form aggregates of up to 1217 kDa (Cluness et al., 1993) and the functional CHT from *G. sorghi* has similarly been shown to form large oligomeric complexes of comparable size. Preliminary negative stain electron microscopy studies have shown that *G. sorghi* CHT forms long extended fibres.

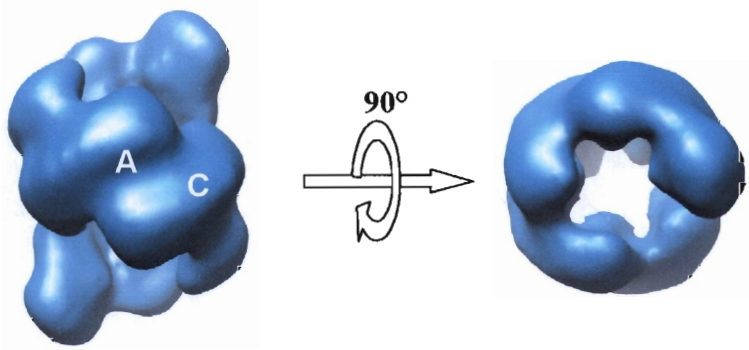


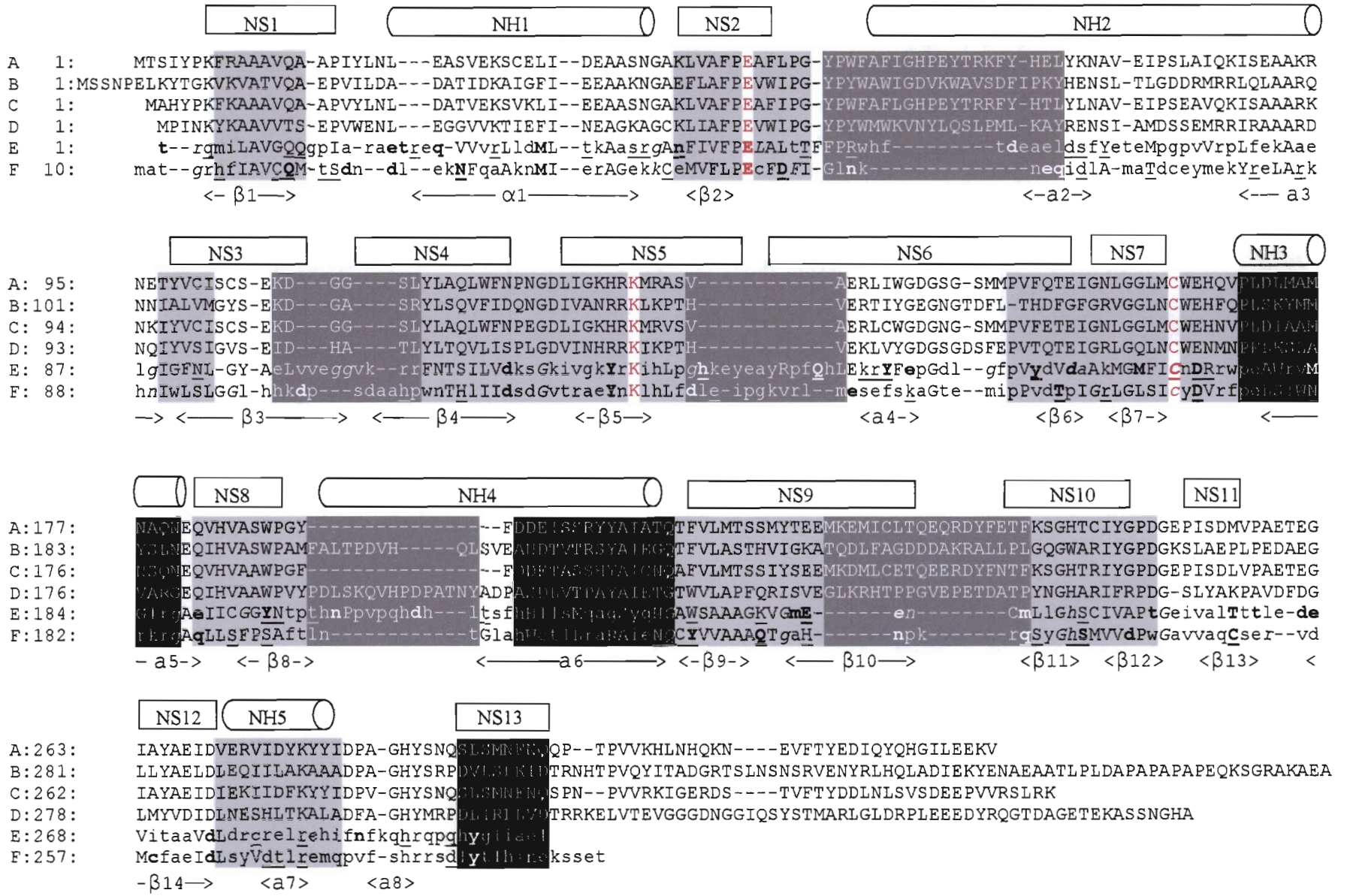
Figure 1.1.10: Side and top view of the electron density map of the cyanide dihydratase (CynD) from *Pseudomonas stutzeri* which exists as a left-handed 14 subunit structure. Surfaces A and C are shown.

The cyanide dihydratase from *Bacillus pumilus C1* ($CynD_{pum}$) forms a terminating spiral of 18 subunits at pH 7-8. This enzyme however is particularly interesting since it forms extended fibres below pH5.4, a property which will be analyzed in this study.

The terminating spiral formed by $CynD_{pum}$ has a global two-fold axis which coincides with the two-fold axis of the dimer formed by the A-surface. The same exists for structures of the extended fibres. The C-terminus of the Nit domain of NitFhit forms a helix which contributes to the formation of the A-surface. Although the structure is not known, it is thought that the C-terminal extensions of the oligomeric nitrilases are analogous to the C-terminus of the Nit domain in forming the interface.

Considering that the sequence is required to be aligned to a basal lineage before one can speak of "deletions" or "insertions" relative to the ancestral gene sequence, it is not really accurate to use these terms since they are relative. However, since this analysis is focused on the changes between the sequences of the known crystal structures and those of the oligomeric nitrilases being studied, the terms will be used in describing changes of sequence relative to the crystal structure sequences. Hence "insertion" is used in this context to mean addition of residues and "deletion" means removal of residues in relation to the crystal structure sequence in question (figure 1.1.11).

Interfaces other than the previously mentioned A surface have been implicated in the formation of the spiral. The variable loop regions of the insertions (white font on grey in figure 1.1.11) are implicated in the formation of these oligomeric interfaces, since they are not found in sequences of the dimeric and tetrameric crystallized proteins. The two main insertions are located between NS2 and NH2 and between NS9 and NS10. These insertions are thought to interact with the corresponding part of another subunit across a two-fold axis thus leading to spiral formation and consequently activation of the nitrilase enzyme. A deletion between NS5 and NS6 is also common to the oligomeric enzymes, indicating that these residues may prevent spiral formation due to steric hindrance. The surfaces involved in oligomerisation (C and D surface) are further described in the discussion (section 4.3).


NS1
NH1
NS2
NH2
NS3
NS4
NS5
NS6
NS7
NH3
NS8
NH4
NS9
NS10
NS11
NS12
NH5
NS13

... from page 19

Figure 1.1.11: Alignment of a nitrilase, two cyanide dihydratases (CDH) and a cyanide hydratase (CH) and with two homologous proteins, the structures of which have been determined by x-ray crystallography (Jandhyala et al., 2005). The conserved, glu-cys-lys catalytic site is in red font. The central core of the enzymes, the alpha carbons locations of which are less than 1.5A apart are shown in black text highlighted in pale grey. The loop regions in which most of the variability occurs, are shown in white font highlighted in grey. The major contact surface leading to dimerization of the enzymes, which is thought to be preserved in the entire family of enzymes, is shown as white text on black. Secondary structural elements according to the NitFhit numbering (Pace et al., 2000) are depicted above the sequence and the DCase structural elements are shown below the sequence.

KEY
A: Cyanide dihydratase from *Bacillus pumilus* C1
B: Nitrilase from *Rhodococcus rhodochrous*
C: Cyanide dihydratase from *Pseudomonas stutzeri* AK61
D: Cyanide hydratase from *Gloeocercospora sorghi*
E: D-amino acid carbamboylase from *Agrobacterium* sp. strain KNK712
F: Nit domain of the NitFhit fusion protein from *Caenorhabditis elegans*

Structurally conserved central core
Active site
Contact surface A
External loops

solvent inaccessible UPPER CASE X
solvent accessible lower case x

hydrogen bond to main chain amide **bold x**
hydrogen bond to mainchain carbonyl underline x
disulphide bond cedilla ç
positive phi *italic x*

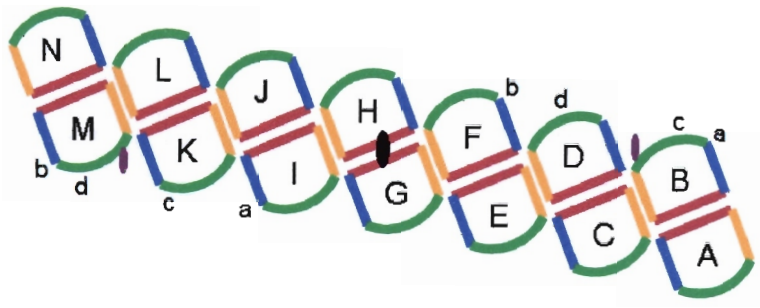


Figure 1.1.12: Schematic diagram of the subunit locations in the unwound spiral where G and H subunits are related by a global dyad. The A surface is shown in red and the C surface is shown in yellow. These subunit interactions result in local pseudo dyad symmetry at these interfaces which are labelled as a,b,c, and d according to their corresponding interacting surface across the groove.

There are two lines of evidence that suggest that oligomerisation is necessary for activity. In several cases, nitrilases isolated from Rhodococci species appear as inactive dimers which only become active as decamers once the substrate benzonitrile has been added (Nagasawa et al., 2000). As mentioned previously, *B. pumilus* CynD exists as an active octadecamer at neutral pH but has been shown to form extended fibres at pH5.4. Below optimum pH, the activity of most cyanide-degrading enzymes decreases monotonically, although with $CynD_{pum,}$ at the inception of fibre formation the activity increases slightly. This perceived increase in activity agrees with the model whereby terminal monomers of the short spirals become activated as additional dimers are added to either end in creating the extended fibre. A possibility is that oligomerisation is regulated *in vivo* and that this regulates enzymatic activity through the conformational changes which are allosterically transmitted to the active site.

1.1.6 Biological Relevance

It has been stated that members of the nitrilase superfamily are present in all plants, animals and fungi and that often more than one nitrilase-related protein is found when analysing the genome sequence of an organism (Pace and Brenner, 2001). In 2004, 137 unique nitrilases were discovered after screening more than 600 environmental cDNA libraries which shows their extreme and diverse presence in many areas of the biosphere, illustrating the ubiquity of these proteins (Robertson et al., 2004).

In the recent phylogenetic classification (fig 1.1.13) by Robertson *et al.* (Robertson et al., 2004), the nitrilases were grouped into six subfamilies based on enzymatic and biochemical properties and associated gene clusters. In further studies, they analysed the nitrilase neighbouring genes for insights into their metabolic roles (Podar et al., 2005). Although only members of sub-family one showed any particular associations with a conserved gene cluster, namely Nit1C, they were able to speculate about the biological functions implicated by the associations.

The four predicted operons downstream of the nitrilase gene which may have some significance in terms of functional understanding are: the radical S-adenosyl methionine superfamily member (SAM), which catalyses the reduction of S-adenosyl methionine, and is most closely related to bacterial genes annotated as biotin synthase-related enzymes. The fourth gene of the Nit1C cluster encodes a GCN5-related N-acetyltransferase (GNAT) superfamily member. Enzymes in this superfamily are involved in antibiotic detoxification as well as histone acetylation in eukaryotes.

ORF5 encodes a member of the 5'-phosphorybosyl-5-aminoimidazole synthase-related proteins (AIRs), superfamily members of which are involved in processes such as selenophosphate synthesis and de novo purine biosynthesis. The last ORF in the cluster is related to the pyridine nucleotide-disulphide oxidoreductases, a group which includes flavoproteins and flavin-containing mono-oxygenases thought to be involved in K^+ transport. These closely related sequences within the Nit1C gene cluster suggest a common function for this expression pathway such as the detoxification of plant or microbial defence compounds (Podar et al., 2005).

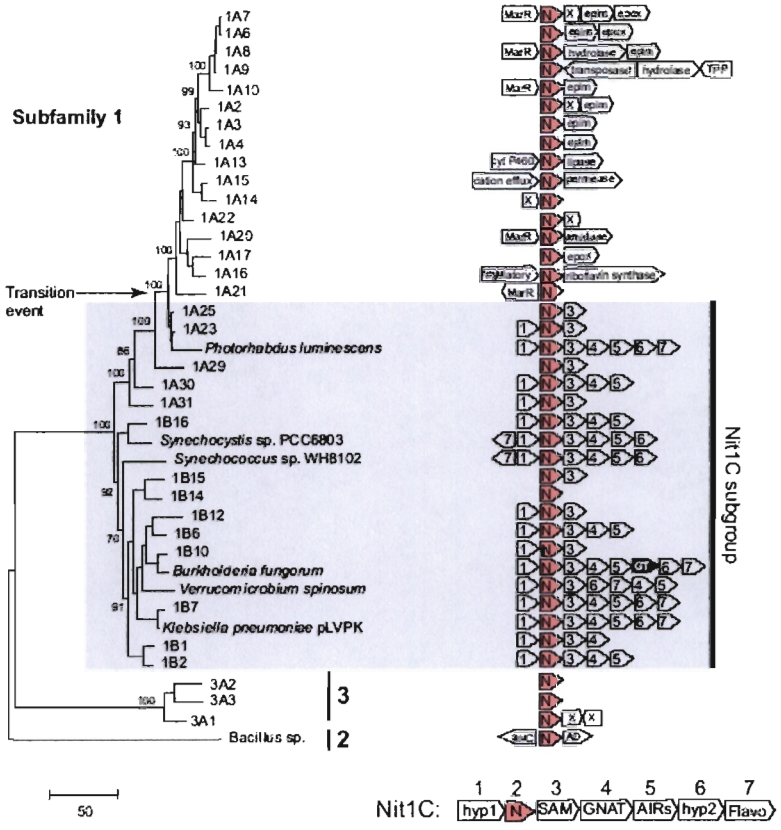


Figure 1.1.13: Maximum likelihood tree of subfamily 1 nitrilases, rooted with sequences from subfamilies 2 and 3, where numbered entries indicate genes from environmental samples. Adjacent to the tree are the diagrams of gene clusters that include the nitrilase ORF.

In another report, a multimeric flavoprotein exhibiting the properties of a peroxide-forming NADH oxidase (Nox) was identified as one of the components of a purported cyanide oxygenase (CNO) from *Pseudomonas fluorescens* NCIMB 11764. This complex catalyzes the oxygenolytic cleavage of cyanide (CN) to formic acid and ammonia. Four protein components including the Nox, along with a pterin cofactor were reportedly required for CNO activity. The second component was a homodimer that showed 100% amino acid identity at its N terminus to NADH peroxidase (Npx) from *Enterococcus faecalis*. A third component, namely a 29 kDa homodimer, was identified as carbonic anhydrase. Another component was reported to be a large oligomeric protein of ~300 kDa that exhibited cyanide dihydratase (CynD) activity. Interestingly, two separate polypeptides of 38 kDa and 43 kDa were detected in the isolated CynD enzyme (Fernandez and Kunz, 2005).

Due to it having similar size to other CynD enzymes, the protein of 38 kDa was assumed to be the domain required for catalytic activity but this was not confirmed by sequence analysis. The amino acid sequence of an internal peptide of the 43 kDa protein was identical to bacterial elongation factor *Tu*, which implicated it in a role as a chaperone in the assembly of CynD or a multienzyme CNO complex. Interestingly, this is the first report describing the collective involvement of four enzymes in the metabolic detoxification and utilization of CN as a bacterial nitrogenous growth substrate (Fernandez and Kunz, 2005).

Other than Nit1C, there is no conservation between the predicted ORFs of the different species or the metabolic functions encoded by the gene clusters. But, the nitrilase genes of *Bacillus sp.* and *Pseudomonas syringae*, members of sub-family two, are apparently co-transcribed with a phenylacetaldoxime dehydratase gene and an araC transcription factor (Podar et al., 2005) which regulates expression of the *ara* genes which are responsible for allowing the bacteria to grow on L-arabinose as its sole source of carbon and energy (Schleif, 1987).

Many nitrilases genes are associated with the MarR transcriptional repressors which are involved in regulating the expression of the operon encoding multiple antibiotic resistances. Epimerases and epoxide hydrolases are also often associated with nitrilase genes. Epoxide hydrolases hydrate chemically reactive epoxides to more stable dihydrodiols which is of most importance in detoxification of both endogenous and xenobiotic metabolites in all organisms. It is speculated that these gene associations indicate a requirement for nitrilase genes to be co-expressed and regulated with the MarR element and that these coupled reactions are responsible for the detoxification of cyanoglycosides in earth containing large quantities of decomposing plant material (Podar et al., 2005).

These recent observations are however merely confirmation of mechanisms that had been speculated as early as 1988. At that time, it was known that bacterial nitrilases are responsible for the detoxification of natural and xenobiotic nitriles produced as defence chemicals by other microorganisms and plants and that they could use the degradation products as a nitrogen source for growth (Knowles, 1988). It was also known that bacteria

and archaea which encode for nitrilases are often associated ecologically with plants and animals. Bacterial and archaeal enzymes are mostly utilized for nitrile or amide hydrolysis reactions or condensation reactions of acyl groups to polypeptide termini (Pace and Brenner, 2001).

Nitrilases are also fundamental to a number of secondary metabolite biosynthetic pathways (Robertson et al., 2004). Cyanide is formed as a secondary metabolite by a wide range of fungi and a few bacteria by decarboxylation of glycine (Knowles, 1988). Similarly, in animals, plants and fungi, biosynthetic pathways of natural products such as auxin, biotin and β-alanine are catalysed by members of the superfamily (Pollmann et al., 2002; Pace and Brenner, 2001).

Many fungi, particularly those which are pathogens of cyanogenic plants such as *Stemphylium loti*, *Gloeocercospora sorghi*, *Fusarium laterium*, are known to produce a cyanide hydratase enzyme. Enzyme activity is inducible in all phytopathogenic species by exposure to cyanide and in some species low constitutive levels of CHT activity is seen (Cluness et al., 1993). *Gloeocercospora sorghi* is a fungus which infects sorghum and when lesions are formed on the sorghum plant, cyanoglycocides degrade to form hydrogen cyanide (Wang and VanEtten, 1992).

Expression of CHT is induced in *G. Sorghi* by HCN and this enzyme can account for up to 18% of the protein in the organism (Wang et al., 1992). It was therefore postulated when first isolated that the production of CHT in fungi such as *G. sorghi* was a means of evading cyanide toxicity, and the presence of this enzyme was believed to be of importance in pathogenicity (Wang et al., 1992). However, in later studies it was discovered that the cyanide hydratase did not enhance the pathogenicity since a *G. sorghi* transformant with the CHT gene disrupted retained virulence on sorghum, causing lesions identical to those caused by the wild-type organism. This CHT mutant was nevertheless highly sensitive to HCN, confirming that CHT is the primary HCN detoxifying enzyme of the fungus (Wang et al., 1999).

Genome sequencing studies have now shown that both phytopathogenic as well as non-phytopathogenic fungi produce cyanide hydratase, illustrating that nitrilase production of these fungi is not linked to virulence. Previously the fungus *Aspergillus nidulans* was thought to lack CHT activity (Wang and VanEtten, 1992) but this has recently been disproven. The two non-pathogenic ascomycetes *A. nidulans* and *N. crassa* express CHT when grown on media containing 0.1mM KCN (Cope, 2005). The biological role of CHT in fungi is not yet known, and the above facts may provide further evidence that HGT or ancient gene duplication was responsible for the acquisition of nitrilase genes in these species.

The idea that nitrilases form an important means by which the bacteria can use environmental nitriles as a source of nitrogen is supported by evidence that these organisms (or transformants with nitrilase genes) are able to utilize nitriles as a sole source of nitrogen (Brady et al., 2004; Nolan et al., 2003; Robertson et al., 2004). Within the nitrilase superfamily, the fusion domains of the glutamine amidohydrolases (GAT) to some bacterial and all eukaryotic NAD synthetases accounts for the fact that only some NAD synthetases are able to use glutamine as a sole source of ammonia (Pace and Brenner, 2001).

Often the expression of a nitrilase is induced in cells grown in media containing a nitrile, but not when ammonia is also present in the medium. This is suggestive of N-catabolite repression, indicating that the *in vivo* role of the nitrilase is to use the nitrile as a source of nitrogen for growth. It is speculated that cyanide is converted to carbon dioxide and ammonia which is then assimilated by an NAD^+ linked oxygenase system (Fernandez and Kunz, 2005; Knowles, 1988). In some cases where more than one nitrilase ORF is present within the same organism, the enzymes may be differentially expressed by exposing the organism to either aromatic or aliphatic nitriles which supports the theory of N-catabolite repression (Banerjee et al., 2002).

One report mentioned that cyanide dihydratase is only expressed in the *B. pumilus* C1 cultures once they enter into late exponential and stationary phase which suggests that the enzyme may be linked to a sporulation pathway. Serendipitous mutagenesis experiments in which the transposon *Tn*917 was inserted into *spoOA* site produced *B. pumilus* mutants

which were no longer able to degrade cyanide. These non-sporulating mutants suggested that *CynD* could be regulated directly or indirectly by *spoOA*. The observation that WT non-sporulating *B. pumilus* strain 8A3 was unable to degraded cyanide under any growth conditions supported this hypothesis. The link between *CynD* expression and sporulation may be related to regulation of genes associated with the response to nutrient depletion (Jandhyala et al., 2003).

Not much is known about the function of nitrilase homologs in mammalian cells and enzymatic activity has yet to be demonstrated for any animal nitrilase homologs. In flies and worms (*Drosophila melanogaster* and *Caenorhabditis elegans*), Fhit is expressed as a fusion protein with the *FHIT* domain fused to a Nit domain which is homologous to plant and bacterial nitrilases (Pekarsky et al., 1998). The Fhit fusion protein has been shown to exhibit diadenosine triphosphate (ApppA) hydrolase activity, indicating its role as a tumor suppressor gene and has hence been shown to be of importance in the apoptotic pathway.

In human and mouse, the nitrilase homologs and Fhit are encoded by two different genes, namely *FHIT* and *NIT1*, which are localized on different chromosomes. Nit and Fhit are expressed as separate proteins in mammals but in murine tissues, the mRNAs have been shown to be coordinately expressed which suggests that suggest that Fhit and Nit1 participate in the same biochemical or biosynthetic cellular pathway in mammalian cells. However, there seems to be no connection between the ability to bind and hydrolyze diadenosine triphosphate ApppA and the hydrolysis of nitrile-containing compounds.

Pekarsky *et al.* therefore speculate the connection between the two enzymes to be that active form of Fhit stimulates or inhibits a nitrilase in production of growth-regulating compounds – this is an extrapolation of the fact that plant nitrilases are responsible for synthesis of the growth factor auxin. It is also thought that mammalian nitrilases may interact (directly or indirectly) with Fhit, causing recruitment of Fhit to the plasma membrane (Pekarsky et al., 1998).

The ability of many species of plant to withstand exposure to reasonable levels of cyanide (liquid and gaseous form) indicates that a cyanide detoxification mechanism may exist in all higher plants (Larsen et al., 2005; Blumenthal et al., 1968). However, the main biological function of nitrilases in plants has been implicated in the production of growth hormones. *Arabidopsis thaliana* expresses four nitrilases. Nit1, Nit2 and Nit3 are able to convert indole-3-acetonitrile to indole-3-acetic acid (IAA) in final step of auxin biosynthesis (Bartling et al., 1992; Hillebrand et al., 1996) while the isozyme Nit4 is a beta-cyano-l-alanine hydratase/nitrilase (Piotrowski et al., 2001). It has now been shown that the *Arabidopsis* enzymes Nit1, Nit2 and Nit3 convert 3-phenylproprionitrile into phenylacetic acid with the highest preferentiality and that specific activity and rate of substrate conversion of indole-3-acetonitrile is the lowest of those substrates tested (Wajant and Effenberger, 2002; Vorwerk et al., 2001).

It was proposed that the nitrilases of *Arabidopsis* play a role in breaking seed dormancy and also for the development of the plant embryo, however only the expression patterns of Nit2 were seen to fit with the putative role of auxin biosynthesis during seed development and/or germination. It has thus been suggested that the production of IAA by the plant nitrilases plays a role during certain physiological situations rather than for the general IAA supply of the plant (Kutz et al., 2002). The product phenylacetic acid - the most preferred substrate of Nit1, Nit2 and Nit3 - is found to have auxin activity in the plant *Tropaeolum majus* (nasturtium) (Ludwig-Muller and Cohen, 2002) which is another supporting line of evidence for the role of nitrilases in plant growth.

A role for Nit3 in the regulation of root morphology *Arabidopsis thaliana* was discovered upon analysis of the promoter under conditions of sulphur deprivation. Under conditions of sulphate starvation, plants developed an increased root system at the expense of shoot growth. The results showed that sulphur deprivation initiates a regulatory loop for glucobrassicin turnover and the production of nitrilase 3 in order to produce auxins to stimulate root development, thus allowing the plant to find new supplies of sulphur (Kutz et al., 2002).

Currently, members of the Nit1-3 subfamily are found only in the Brassicaceae, where their major function is to metabolize nitriles originating from glucosinolate breakdown in the plant to their respective carboxylic acids (Vorwerk et al., 2001). For example, the cyanide produced on the breakdown of cyanogenic glucosides in Lotus seedlings and shoots of Nandina domestica is readily converted into asparagine (Blumenthal et al., 1968). Though, the isoform NIT4 and homologs of thereof are also found in species not belonging to the Brassicaceae like *Nicotiana tabacum* (tobacco) and *Oryza sativa* (rice) (Piotrowski et al., 2001). The extensive metabolism of HCN in higher plants raises the question of the significance of nitrilase 4.

Purified nitrilase 4 enzyme shows a strong specificity for β-cyanoalanine (Ala(CN)) which is an intermediate product of cyanide detoxification in higher plants. Extracts from senescent leaves of *Arabidopsis* showed higher Ala(CN) nitrilase activities than extracts from nonsenescent tissue which may either indicate a role for Nit4 in cyanide detoxification during ethylene biosynthesis (Piotrowski et al., 2001), or it may indicate a link to an unknown apoptotic pathway as with the NitFhit association.

Even though the biological relevance of nitrilases seems particularly speculative at the present time, the rapidly forthcoming information obtained though biochemical studies and analysis of associated gene clusters will hopefully provide a greater amount of insight into the biological metabolic roles of these enzymes in the near future.

1.7 Catalysis and Specificity

It has previously been suggested that nitrilases be divided into three classes, depending on whether the preferred substrates are aromatic (or heterocyclic) nitriles, aliphatic nitriles or arylacetonitriles (Banerjee et al., 2002; Berman, 2003). However, it is now known that the nitrilases can be grouped into phylogenetic lineages which closely reflect the enantiomeric catalysis of the enzymes (Podar et al., 2005). This seems to be a far more intuitive way to categorize catalysis since there are many enzymes that cannot be categorized based on their substrate preferences since they are non-specific. Some cyanide dihydratases hydrolyse

hydrogen cyanide but do not generally function on larger nitriles and so cannot be accurately termed aliphatic nitrilases (Berman, 2003). Other nitrilases that have larger substrates are very broad in their specificity, and convert both aromatic and aliphatic nitriles (Banerjee et al., 2002; Brady et al., 2004).

Substrate specificity is often not completely known for an enzyme, as in the case of the nitrilase from *Rhodococcus rhodochrous* J1 which was originally thought to be an aromatic nitrilase but becomes activated for aliphatic substrates after exposure to aromatic nitriles (Banerjee et al., 2002). Certain nitrilase enzymes hydrolyze a variety of substrates, but their enantioselectivity varies depending on the substrate, thus large amounts of environmental screening is required to discover enzymes that convert a particular substrate with the correct enantioselectivity (Robertson et al., 2004). Through a comparative study of enzymatic reactions of nitriles, it may be seen that resting cell cultures and enzymes exhibit a distinctive substrate selectivity profile which is generally related to the length of the aliphatic chain and the position of substituents on the aromatic ring or aliphatic chain (Brady et al., 2004).

There is overwhelming evidence from sequence analysis (DeSantis et al., 2002), (Robertson et al., 2004; Pace and Brenner, 2001), comparative biochemical information (Brown et al., 1995; Novo et al., 2002; Watanabe et al., 1998) and crystal structure analysis (Chen et al., 2003; Wang et al., 2001) that the entire superfamily utilizes a novel catalytic triad consisting of glutamate/aspartate, lysine and cysteine. Site directed mutagenesis studies revealed that point mutations of the cysteine residue leads to inactivation of the enzyme (Novo et al., 1995; Brown et al., 1995). Other studies show that nitrilases are reliant on a catalytic cysteine residue since they are susceptible to inactivation by thiol-specific reagents (Novo et al., 1995; Harper, 1985) which indicates that they proceed through a covalent thioimidate intermediate. The aliphatic nitrilase of R. rhodochrous K22 was one of the few to exhibit a high resistance to thiol reagents but this is thought to be due to the active site cysteine being buried deep within the interior of the protein (Novo et al., 1995).

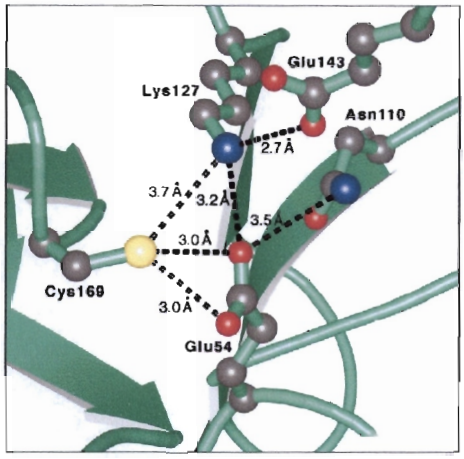


Figure 1.1.14: The catalytic centre of a Nit domain of the NitFhit protein showing the region around Cys169, a residue conserved in nitrilases. The catalytic triad is thought to be composed of residues homologous to Cys169, Glu54 and Lys127 (in Nit) in all members of the nitrilase superfamily (Pace et al., 2000).

The recent crystal structures of WT *N*-Carbamoyl-D-amino Acid Amidohydrolase (D-NCAase), as well as several mutant enzymes from *Agrobacterium radiobacter* with the substrate *N*-carbamoyl-D-*p*-hydroxyphenylglycine (HPG) have finally confirmed the common (Glu/Asp)-Lys-Cys triad. From an assessment of the stereochemical quality of the DCase crystal structure, it was noted that two residues Cys171 and His 284 were in the disallowed region of the Ramachandran plot. There was, however, good corroboration between the two crystallographically independent molecules in the asymmetric unit resulting in low B-factor values and clear electron densities for these residues after superposition - indicating that these residues indeed have prohibited dihedral angles (Nakai et al., 2000). The cysteine, which is the active-site nucleophile is positioned on a nucleophile elbow as with the serine hydrolases (Kumaran et al., 2003).

In an attempt to demonstrate the catalytic essentiality of glutamic acid and lysine, Glu59Gln and Lys134Arg mutants were shown to be inactive. However by CD spectra, these substitutions were shown to affect protein stability. Later, the inactive mutants containing Glu59Gln and Lys134Asn substitutions were shown to be stable thus proving the essentiality of these residues (Novo et al., 2002).

Upon analysis of a substrate-bound crystal structure of the mutant C172S the side chains of Glu47, Lys127, and Ser172 are seen to cluster toward the carbamoyl moiety of the substrate indicating that the Cys172-Glu47-Lys127 catalytic triad is involved in the hydrolysis of this group. Residues Asn173, Arg175, and Arg176 are seen to interact with the carboxyl group of HPG suggesting that they are responsible for conferring substrate specificity (Chen et al., 2003). Observation of the structures of N-carbamoyl-D-amino acid amidohydrolase, Nit-Fhit, and another carbamoylase suggests a conserved and robust mechanism that endured through evolution enabling functional diversification toward efficient catalysis of a specific nitrile or amide.

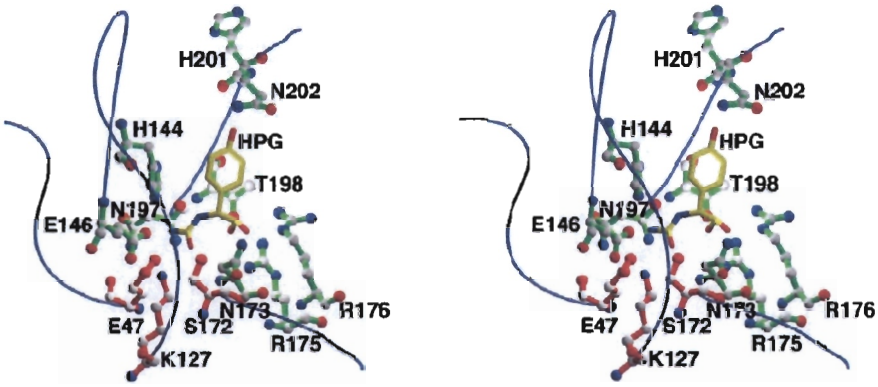


Figure 1.1.15: Stereoview of the binding pocket of DCase mutant C172S bound to the substrate HPG. HPG is in yellow, the four loops enclosing HPG are blue. "Catalytic triad residues", Glu47, Lys127, and Ser172 are in red and other residues which bind to HPG are shown in green.

The proposed mechanism for the nitrilase reaction indicates that the thiol group of the catalytic cysteine initiates a nucleophilic attack on the carbon atom of the nitrile with concomitant protonation of nitrogen to form a tetrahedral thiomidate intermediate. In the double hydrolysis reaction of the cyanide dihydratases, subsequent steps involve attack by two water molecules and protonation of the nitrogen atom which is then lost as ammonia. In the case of the cyanide hydratases, the tetrahedral intermediate is formed through a single hydrolysis reaction, which then breaks down to produce an amide as opposed to an acid product (Banerjee et al., 2002; Brenner, 2002). Other members of the nitrilase superfamily such as glutamine amidotransferases (Muchmore et al., 1998), signature amidases {Patricelli & Cravatt 2000 82 /id} and nitrile hydratases (Huang et al., 1997) have substrates in

common with nitrilases but are not structurally related to these enzymes and hence possess different catalytic mechanisms.

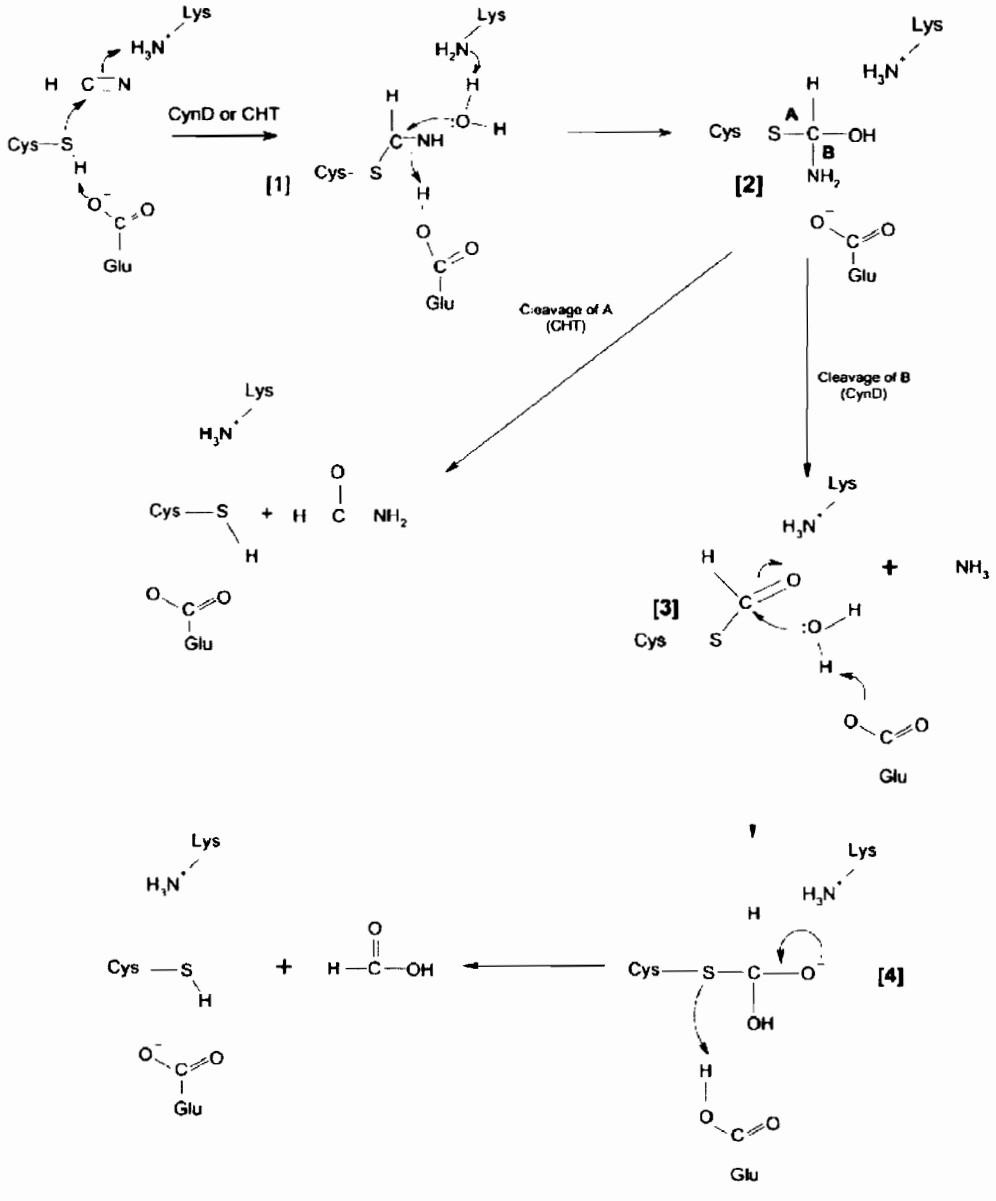


Figure 1.1.16: The covalent mechanism of cyanide degrading enzymes is such that the enzyme does not add water to a nitrile, but rather forms an enzyme linked thioimidate intermediate [1]. This is facilitated by the catalytic glutamate residue which acts as a general base and activates the nucleophilic attack by the cysteine on the nitrile. The catalytic lysine then catalyzes the first hydrolysis, leading to the tetrahedral intermediate [2]. Because ammonia is a better leaving group than the enzyme, the first enzyme-dependent water addition produces an acylenzyme as opposed to an acid amide [3]. In the cyanide dihydratases, the hydrolysis of the acylenzyme, activated by glutamate then produces the acid product [4]. In the cyanide hydratases the C–S bond breaks leading to the release of the amide product.

There are several lines of evidence indicating that nitrilase and cyanide hydratase (sometimes called nitrile hydratase) share a common tetrahedral intermediate in the hydrolysis reaction. Kobayashi *et al.* observed that *Rhodococcus rhodochrous* nitrilase J1 was also able to use an amide as a substrate (Kobayashi et al., 1998a; Kobayashi et al., 1998b) while Weiler's group showed that *Arabidopsis* NIT4 has both nitrilase and nitrile hydratase activity since it produces both asparagine and aspartic acid as products (Piotrowski et al., 2001). NIT4 did not utilize asparagine as a substrate, indicating that it's not an intermediate but rather the product of a single hydrolysis step. The first evidence of CHT having nitrilase activity was discovered when the recombinant CHT of *Fusarium lateritium* showed a low but significant nitrilase (CynD) activity with acetonitrile, propionitrile and benzonitrile. Site-directed mutagenesis of the *F. lateritium* CHT gene indicated that mutations leading to a loss of CHT activity also lead to a loss of CynD activity (Nolan et al., 2003). In the case of the commercially exploited Rhodococcus strains, the use of an amidase inhibitor stops the nitrile hydratase/amidase reaction at the amide intermediate (Brady et al., 2004). These facts suggest that the catalytic centre for both hydratase and dihydratase activity in the protein is the same (Nolan et al., 2003).

CHAPTER 1.2

METHODOLOGICAL REVIEW

1.2.1 The freeze-dry shadowing process

Freeze-dry electron microscopy, is a technique developed for the resolution of fine structure by preventing the phenomenon of particle collapse as occurs with conventional negative staining procedures (Harris, 1997). The freeze-dry heavy-metal shadowing approach is a well established extension of this approach used for structural studies of biological specimens (Fuchs et al., 1994). This method is particularly useful in the determination of the absolute hand of a structure since only the surface of the particle is imaged and hence the ambiguity that intrinsically arises from TEM (via projection of both near and far sides of the biological specimen) does not occur in a shadowing experiment. The method involves adsorption of isolated particles onto glow-discharged carbon grids which are plunged rapidly into liquid cryogen before they have a chance to air-dry. This procedure avoids the usual flattening of samples as experienced with regular shadowing experiments and thus reveals surface features which are in a well-preserved state (Harris, 1997).

At very low temperatures, the saturation vapour pressures of all residual gases besides H_2, are below 10^{-10} mbar. At this point, the specimen effectively acts as a cryopump and attracts residual gases (and also gas produced from the evaporation process) to its surface which then causes lateral movement of the condensing material. This was one of the reasons for developing the 'Midilab' apparatus which is capable of shadowing the samples under UHV (ultra high vacuum: 10^{-9} mbar) conditions, thereby minimizing specimen interactions with residual gases and consequently reducing lateral mobility of the shadowing material (Gross, 1987).

High resolution shadowing films such as Ta/W usually have to be protected from alterations during air exposure by a thick carbon backing layer when transferring samples through atmospheric conditions to the TEM. Unlike regular preparations that have to be transferred through air, the Midilab instrument does not require a carbon backing, which results in a

significant increase in resolution and a higher signal/noise ratio generated by a higher level of contrast (Gross et al., 1990). As with negative stain, the contrast created by the metal coat is so high that the contrast component resulting from unstained material below the surface is negligible.

The midilab instrument is also equipped with a high vacuum airlock with which the samples are loaded into the freeze-drying chamber, thus enabling efficient and contamination-free specimen transfer. It is known that removal of strongly bound water during freeze-drying takes place at the point of maximum sublimation which occurs at temperatures between 223K (-50°C) and 243K (-30°C) (Gross, 1987). To avoid this, freeze-drying is done at a temperature of 180 K (-93°) so that the hydration shells necessary for structural and mechanical stability of the proteins are preserved (Fuchs et al., 1994). It has been noted that structural collapse may occur in the process of freeze-drying, however, this is usually only seen upon removal of structural water shells when drying temperatures are far higher than 180K. Another admonition is that structural preservation can only be achieved during this process if the proteins are relatively robust since they are required to undergo brief rinsing with distilled water prior to plunge-freezing without exhibiting changes in morphology. The rinsing step is carried out in order to reveal uncontaminated surfaces after drying (Gross, 1987).

In a perfect shadowing experiment, if the sample is viewed from a different angle to that of the evaporation source, the resulting contrast should be directly related to the surface topography. Of course, in practice there are many factors which affect the resolution and representation of the sample. Such factors include thermal accommodation, surface diffusion, nucleation, crystal growth and coalescence as well as lateral mobility of the shadowing material (Gross, 1987). Generally though, the method provides high-contrast, low artefact images which reveals a great amount of information about the surface of the material being shadowed.

1.2.2 Iterative Helical Real Space Reconstruction

IHRSR is a method for the iterative real space analysis of helical filaments which determines helical symmetry by iterative refinement. This method applies the single-particle approach to boxed segments of the continuous extended fibres where each segment is treated as a single particle, assuming that each box contains a representative image of the same structure viewed from a different uniaxial angle. Therefore, if the symmetry of the helix is known, a comprehensive 3D reconstruction can be calculated from the untilted transform, since the asymmetric unit is imaged at different angles about the filament axis. A 3D reconstruction may be carried out according to standard single-particle techniques of reference based alignment, angle-assignment and filtered back-projection.

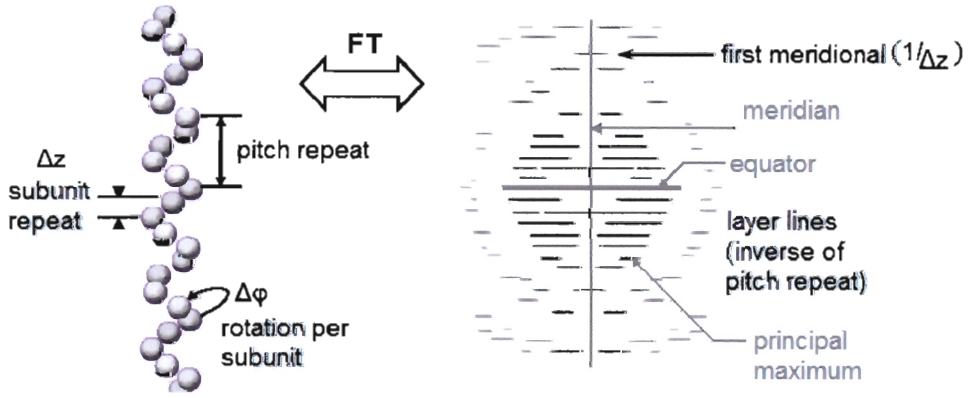


Figure 1.2.1: A helix can be considered as a 1 dimensional crystal since it has a repeating structure along the axis giving rise to a set of layer lines in the diffraction pattern (the Fourier transform is that of a continuous helix convoluted with transform of lines). Helical symmetry consists of a screw operation where the axial rise = Δz and rotation per subunit = Δφ. If the symmetry of the helix is known, a full 3D reconstruction can be calculated from the untilted transform, since the subunit is imaged at different angles about the filament axis.

Traditionally before IHRSR, Fourier-Bessel methods were used to solve helical structures (De Rosier and Klug, 1968), and x-ray diffraction patterns were used to obtain evidence about periodicities and symmetries from ordered fibres (Klug et al., 1958). The Fourier-Bessel methods involved formulating models compatible with the observed intensities of the layer-lines in the diffraction pattern (DeRosier and Moore, 1970). Aside from the fact that IHRSR is easier, both conceptually and in practice than Fourier-Bessel methods, there are

also several other advantages to bear in mind. Bessel overlap can occur for many helical structures at low resolution leading to problematic indexing of ambiguous layer lines. Since IHRSR operates in real space, it is unaffected by the problem of Bessel overlap. However, as will become evident at a later stage, it is still useful to extract information from the layer lines of the image diffraction patterns. Many helical structures have large variability so that *Fourier-Bessel* methods are rendered useless, whereas IHRSR can deal with disordered or heterogeneous filaments (Egelman, 2000). Of course, a potential problem that may arise with the single particle approach in this case is that components in the structure that are not related to the basic helical symmetry are averaged out. This is to some disadvantage since the asymmetric components are often functionally significant and their contribution to the overall structure is important (Paul et al., 2004).

The application of single particle techniques to the 3D structural analysis of helical protein complexes has been around since as early as 1996 (Boettcher et al., 1996), however the approach was only employed to a greater extent once it had been further simplified and extended by Egelman in 2000 (Egelman, 2000). The novelty of Egelman's approach is that the helical parameters Δz (namely the rise between adjacent subunits) and $\Delta\varphi$ (the rotation per subunit) are refined in each cycle. In addition he has prepared scripts to simplify the process. The helical search is a least-squares algorithm which uses an initial guess for Δz to determine $\Delta\varphi$. The mean-squared deviation in density ($<d^2>$) is calculated between voxels of density and the density in the symmetry-related positions as $\Delta\varphi$ is varied over the given search increments. The value of $\Delta\varphi$ is then used for new determination of Δz. Since the two helical parameters are related, the procedure is iterated only once using the updated values. The optimized parameters $\Delta\varphi$ and Δz are then imposed on the reconstructed 3D volume to generate a helically symmetrical reconstruction (Egelman, 2000).

1.2.3 Multi-reference alignment

Since the fibres lie in the plane of the EM grid, only projections around the helical axis are available without tilting which does not provide enough information to assign Euler angles ab initio, so reference free alignment is not a feasible option. This is because at least three

images related by rotations around three different common tilt axes are required to determine the three rotational degrees of freedom (van Heel, 1987). Thus, if no out-of-plane views are available, it is necessary to use a reference model to provide angle assignments. However, since different rotational views of the filament are contained within a single projection of a helical object, a full 3D reconstruction can be calculated from the untilted transform using multi-reference alignment (Saibil, 2000).

During single particle reconstruction, images within the same class (aligning to the same template) are arithmetically averaged after alignment to increase the signal-to-noise ratio. During IHRSR, the images themselves and not the class averages are back-projected but the same effect is achieved. Since each image is back projected with the same angle after alignment to the template an improvement in the signal-to-noise ratio is achieved because the random noise present in the images averages out. Another advantage to aligning images to reference projections is that the effect of damaging radiation averaged out when combining views from different particles.

Since averages are aligned against reference templates, their class averages will always resemble the templates. This problem of model bias, can however be overcome by the knowledge that different starting models converge to the same structure – indicating that convergence of the structure is independent of the starting structure. The distribution of images aligning to the references should be statistically random and should show an evenly distributed profile. If an uneven reference distribution is seen during the reconstruction it is often an indication of artefact. 3D maps produced from uneven reference distributions should be cautiously accepted since it will also influence the back-projection algorithm.

1.2.4 The back projection algorithm

The proper choice of weighting filter for the process of back-projection is essential for accurate 3D reconstruction of helices since the particles have such a large axial ratio. It has been proven that with an oversampled even distribution of reference projections, the simpler back-projection algorithm of the R weighted filter provides a more accurate reconstruction than the exact filter (Paul et al., 2004). The exact filter (Harauz and van Heel, 1986) takes

into account the overlap of central sections in Fourier space associated with the back projection of individual images but the thickness of a central section and amount of overlap is assumed to be from an object of equal dimensions. Consequently, the overlap arising from Fourier wedges of extended objects such as helices are considerably underestimated. Thus with the exact filter, low frequencies of the central sections are also insufficiently down weighted which results in overrepresentation of low frequency components in the 3D density map. Without oversampling or if the projection images are unevenly distributed, the ramp filter also tends to produce a map which is smeared in the direction of overrepresentation; exact filters tend to perform better in these conditions. It is evident that a simpler weighting algorithm such as the R weighted filter - which does not take into account section overlap - outperforms other filters when reconstructing volumes of a high axial ratio (Paul et al., 2004).

It is imperative that an appropriate box size be chosen because if the box size is too large, the above mentioned problems with back projection will be encountered. Also, problems with filament bending and internal disorder will pose a problem. On the other hand, box sizes should not be too short since the segment should be able to provide full information about the helical symmetry. If segments are too short they are also more likely to cause ambiguity during cross-correlation. Box size should be optimized for each sample and factors such as mass per unit length, signal to noise ratio of the images and filament flexibility should be taken into account (Egelman, 2000).

1.2.5 Fourier shell correlation threshold criteria

A quantitative means of determining the resolution of a structure is by means of the Fourier shell correlation (FSC) criterion. Generally, the image set is split in half and two separate reconstructions are generated and compared. The Fourier shell correlation, which has become the standard quality measure, is used to determine the correlation as a function of spatial frequency between the Fourier coefficients computed from the reconstructions (Harauz and van Heel, 1986). This resolution value should indicate the overall quality of the

reconstruction but considerable discussion exists in the literature as to the value of the FSC to be used as a resolution estimate.

It has been stated that the fixed-value FSC threshold such as the 'conservative' value of 0.5, should never be used as a reproducible criterion (Van Heel et al., 2005). Fixed-value thresholds were first proposed for use as a cut-off for DPR (differential phase residual) but once the DPR curve had been abandoned in favour of the FSC, fixed-value resolution thresholds were then used in conjunction with FSC curves. The 0.5 value as a resolution-defining threshold is extrapolated directly from the x-ray crystallographic cut-off criterion of the figure of merit used in x-ray crystallography. Recently a fixed-value threshold of 0.143 has been proposed as a more realistic value (Rosenthal and Henderson, 2003). This value was based on a calibration with the numerical values of the FOM threshold.

The σ-factor threshold curve is a proposed criterion which accounts for the varying number of voxels per shell since it compares the FSC curve of a structure to the behavior of the Fourier-space cross-correlation coefficient of a theoretical pure-noise curve. The most frequently used σ-factor threshold is a value of 3, which means that the significance threshold is set to three standard deviations above the expected random noise fluctuations (3σ-curve). The main criticism of the sigma curve is for information which has been collected with significantly high signal-to-noise ratios, the values for which the 3σ-curve crosses the FSC curve is generally unrealistic since the cross-correlation coefficient is related to the SNR within a particular shell.

When the 3D structure contains symmetry, as with helical structures, the symmetry is reflected in Fourier space and therefore the number of independent voxels in the two Fourier space shells will be lower than for an asymmetrical object. It should therefore be necessary to divide the number of voxels in a shell by the number of asymmetric units in the symmetric oligomer.

In his recent assessment of the different resolution criteria, van Heel (Van Heel et al., 2005) suggests use of the ½-bit threshold curve. This criterion defines a threshold curve for indicating a predefined Fourier-space SNR which has been achieved at a particular resolution level. These new information-based threshold curves are modified versions of the σ-factor threshold curves which include the effect of symmetry and also the degree of filling of the 3D volume. Van Heel states that although Rosenthal and Henderson and co-workers do not take the signal-noise cross-terms into account, the fixed value of the 0.143 criterion is almost correct in an asymptotical sense. Bearing in mind that fixed-value resolution thresholds cannot account for a number of factors which influence the behaviour of the FSC, such as the number of voxels in a given Fourier shell, the size of the structure within the reconstruction volume, as well as the symmetry of the structure, the '0.5' fixed-value still seems to be the most commonly quoted value in the literature to date. Since this seems to give more realistic or 'conservative' values, it will likely remain the most common criterion to be used in the future.

1.2.6 Fourier transforms of helical objects

A helix can be considered as a one-dimensional crystal since it has a repeating structure along the axis and thus its Fourier transform is that of a continuous helix convoluted with a transform of lines. Because of their intrinsic cylindrical symmetry, diffraction by helical objects is naturally expressed in terms of the Bessel functions. Furthermore, because helical structures are a regular array of discrete objects their diffraction pattern is non-zero only at discrete values called layer lines which are related to the pitch repeat.

The Bessel functions of the first kind $J_n(x)$ are defined as the solutions to the Bessel differential equation where n is a non-negative real number. For $J_n'(x)$, n is the startedness of the helix (Bessel order). The function $J_n (2\pi \text{ Rp } r_0)$ has a maximum when $J_n'(2\pi \text{ Rp } r_0) = 0$, where J_n' is the first derivative of J_n, Rp is reciprocal Å i.e. the distance in pixels from the meridian and r_0 is the radius. In the simplest case, each layer line contains a single Bessel function (Egelman, 2004).

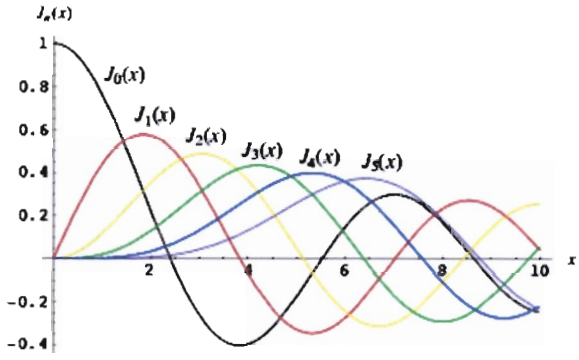


Figure 1.2.3: Graphs of the Bessel functions of order n. Values of the arguments producing the principal maxima are obtained from the zeros of the derivatives, i.e. $J_0'(3.831)$, $J_1'(1.841)$, $J_2'(3.054)$, $J_3'(4.201)$, $J_4'(5.317) = 0$. These values are used when calculating the effective radius r_e of a given principal maximum on a particular layer line (section 4.8).

A helical structure may be viewed as the superposition of a set of sinusoidal density distributions each giving rise to an identifiable part of the diffraction pattern. The superposition of coaxial thin cylindrical shells consisting of these wave-like densities follow a path of n parallel helices of pitch P. The transform is confined to layer lines in the z-plane, namely $l = \pm n/P$. Each line has two principal maxima (symmetry-related peaks across the meridian) , except when n is zero, then the helices become annular and a single maximum (a meridional reflection) appears along the meridian of the transform separated by $1/\Delta z$. Several subsidiary maxima also exist for any given Bessel order. Any helical structure can be defined as the sum of its various component density distributions with amplitudes and values of n and P and r (Moody, 1971).

The helical symmetry of the object may be best understood by the concept of the 'helical net' which is a representation of the helical structure viewed from the inside. It is as if the cylinder has been cut parallel to the axis and opened flat, thus revealing a flat sheet in which the component helices appear as parallel, equidistant, straight lines. Given a helical net, the subunits of a discontinuous helix can be illustrated as an infinite set of points which lie on a continuous helix and are separated by a constant axial translation z. The pattern of points will repeat itself after a distance 'c' along the axis, thus after $t = c/P$ turns of the helix there will be $u = c/z$ points.

The helix can therefore be described as having u equivalent points in t complete turns where u and t are obtained by expressing z/P as a rational function. Describing a discontinuous helix in terms of z, P and u are usually chosen in order to make the unit of twist a minimum (the lowest startedness with which the helix may be described) - this is known as the basic helix. There are however, an infinite number of helices that pass through the subunits of a helical net (Egelman, 2004). It is therefore best to represent the helices as a reciprocal lattice of the helical net in an (n,l) plot (Klug et al., 1958). Two adjacent lines of the (n,l) plot intersect the cut edge with a separation of $1/l$ and intersect any perpendicular line with a separation of $2\pi r/n$. Each point in the (n,l) plot corresponds to n-start helix with l related to the inverse of the pitch repeat. Since reflections from a regular helix obey its reflection rule, in order to determine the values needed for the (n,l) plot, the values for l are calculated according to $l=tm+un$ where n is the order of the Bessel function and m is any integer which satisfies the equation (Klug et al., 1958).

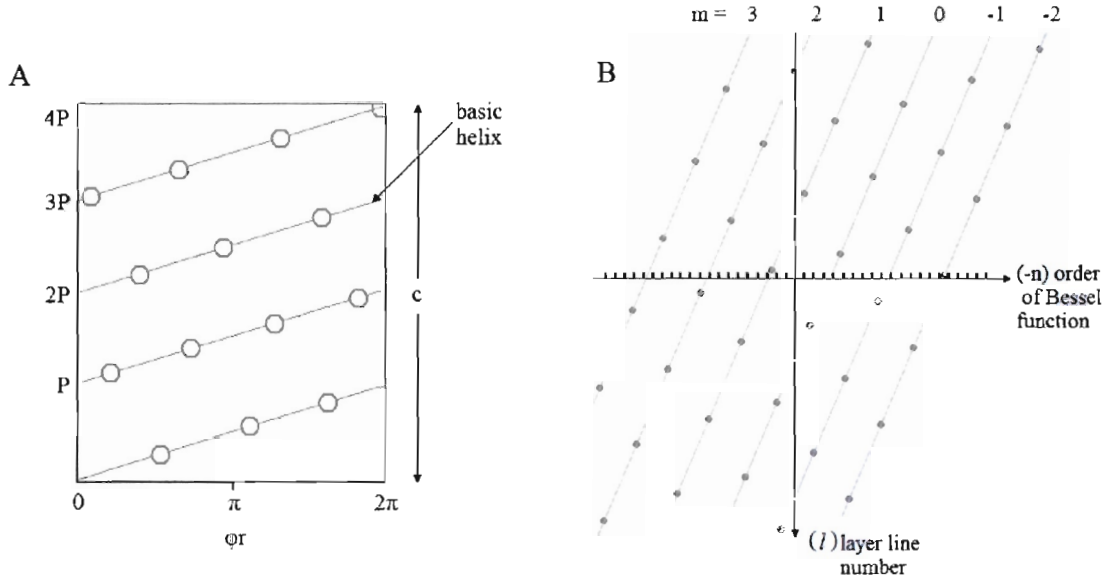


Figure 1.2.5: (A) A hypothetical helical net in which the axial repeat consists of 14 subunits in 4 turns. (B) A hypothetical reciprocal lattice of the helical net in an $(-n,l)$ plot which describes reflections from a helix in terms of the reflection rule l=tm+un.

As a general rule, indexing of layer lines assuming integral numbers of subunits per turn is highly sensitive to fluctuations in $\Delta\varphi$. In reality, the spacing along the vertical axis (ζ) as defined by Moody (Moody, 1967a), is given by n_i/P_i where n_i is the multiplicity of the set of helices and P_i is the pitch, and their handedness determines on which side of the vertical axis the points lie. The (n, ζ) plot is therefore the form assumed by the (n,l) plot (Klug et al., 1958), when considering helices in which the z-axis repeat is absent or unspecified as occurs when ζ is not restricted to multiples of 1/c.

Unless the helical particle has a very short repeat (if certain transform components have the same *l* coordinate they may interfere with each other), it should be possible to determine the *n* and *l* values for any pair of layer lines in the diffraction pattern. At low resolution Bessel functions start to overlap and a given layer-line might have Bessel function contributions of different orders (Moody, 1971) but this is overcome by computationally combining data from different views during reconstruction (Crowther et al., 1985).

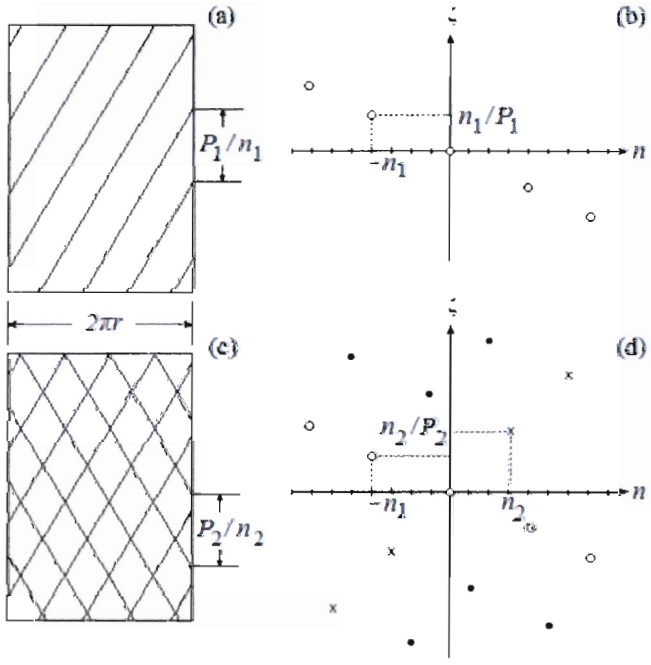


Figure 1.2.4: (a) Radial projection of a group of 4 parallel continuous helices related by a 4-fold axis. (b) The (n,ζ) plot of (a) where *n*-coordinates are multiples of the number of helices (n_1) which equals four in this case. (c) Radial projection of a helical surface lattice formed by two sets of intersecting lines represented as crosses on the corresponding (n,ζ) plot shown in (d). The resulting lattice generates additional points which are represented by dots.

CHAPTER 1.3

AIMS OF THIS RESEARCH

The quaternary structure of the terminating spirals of *P. stutzeri* and *B. pumilus* had already been solved through single particle reconstruction, however, the structure of an extended nitrilase fibre had yet to be elucidated before this study. The aim of this research dissertation was therefore to:

1) unequivocally determine the handedness of the extended nitrilase fibres by unidirectional metal shadowing,

2) determine the helical quaternary structures of the fibre-forming nitrilases from *G. sorghi* and *B. pumilus* using the single particle approach of Iterative Helical Real Space Reconstruction (IHRSR),

3) dock atomic models into the three-dimensional electron density maps produced from the reconstructions,

4) rationalize the layer line data from diffraction patterns of the images,

5) establish the possible interfacial contacts of the helical forming nitrilase enzymes in order to provide a better understanding of the interactions between the subunits leading to oligomerisation.

SECTION 2 – METHODS AND MATERIALS

CHAPTER 2.1

PURIFICATION OF RECOMBINANT CLONES

Growth and expression of the recombinant clones, as well as purification of the recombinant protein were performed according to the conditions as previously described by Mark Berman (Berman, 2003) and were optimized accordingly. Several protein preparations were made in order to collect the data used for the reconstructions. The *G. sorghi* protein purification was performed by Brandon Weber, Electron Microscope Unit, Physics Department, University of Cape Town. Data collected from the *G. sorghi* purifications shown here may be attributed to him.

2.1.1 Origin of bacterial clones and plasmids

Recombinant clones MB2890 and MB2313 were received from Professor. Michael Benedik, Department of Biology and Biochemistry, University of Houston, Texas. The host *E. coli* strain BL21pLysS, which carries a chloramphenicol resistant plasmid, was transformed with the expression vector pET26b which confers kanamycin resistance. The inserted sequence in MB2890 is the coding region of the cyanide dihydratase ($CynD_{pum}$) gene of *B. pumilus* C1, while that in MB2313 is the coding region of the cyanide hydratase (CHT) gene of *G. sorghi*. The stop codons of the insertions prevented expression of the hexahistidine tag of the pET26b vector.

2.1.2 Growth conditions for expression

For each purification, freshly plated colonies of either MB2890 or MB2313 were used to inoculate an initial pre-culture of 5ml Luria-Bertani (LB) broth (1% tryptone, 0.5% Yeast Extract, 0.5% NaCl in distilled water) (Davis et al., 1980). Solid medium for colony storage was as per LB with additional 1.5% (w/v) agar. The pre-cultures were grown overnight at 37°C under kanamycin (30µg/ml) and chloramphenicol (200µg/ml) selection. 1ml of the overnight pre-culture was then added to 1 litre LB cultures which were placed in

Erlenmeyer flasks of 5 litre capacity in order to achieve adequate aeration. Plasmid selection pressure was ensured by addition of 30μg/ml kanamycin to the batch cultures which were grown at 37°C to an OD600 = 0.9 at which point, the cells were induced with IPTG to a final concentration of 1mM. Cells were induced for four hours at 37°C.

After induction, the cells were harvested by centrifugation at 8000rpm (1000xg) for 10 minutes in a Beckman JA-17 rotor and resuspended in 1L fresh LB with chloramphenicol. The LB re-suspension was placed at 37°C for 1 hour. Chloramphenicol interferes competitively with the binding of the aminoacyl-tRNA 3'-terminus to ribosomal A-site (Moazed and Noller, 1989), thus inhibiting protein synthesis. It is believed that incubation with this antibiotic leads to retrograde transport of proteins from inclusion bodies if they are so located (Carrio and Villaverde, 2001; Carrio and Villaverde, 2002). No wash step was required subsequent to incubation with fresh LB.

2.1.3 Harvesting and lysis

Cells were once again harvested and resuspended in 50ml buffer containing 50mM Tris, 50mM NaCl pH 8. One tablet of a combination of protease inhibitors (Promega) was added to the resuspension. During some procedures, 1ml of 10mg/ml lysozyme was added to the cells which were incubated for 30min at 37°C prior to sonication. For the most part, the lysozyme treatment was omitted since sonication disrupted the cells adequately. The suspension was placed in a 50ml stainless steel vial and sonicated for a total of 4 minutes process time with alternation between pulse ON for 15 sec and pulse OFF for 15 seconds using the Misonix Sonicator® 3000 with a 12.5mm tip at output level 4. The temperature was measured with the temperature probe and the sample was kept at a constant temperature of 4°C in ethanol-ice slurry. The lysate was then centrifuged in a Beckman JA 20 rotor at 15000 rpm (4000xg) for 30 minutes.

2.1.4 Sample concentration

The supernatant of the lysate required a concentration step prior to loading onto the anion exchange column. For the *B. pumilus* MB2890 preparations, the samples were ultrafiltrated using a 50ml Amicon cell (Amicon Corporation, Lexington, Mass.) with a PM10 membrane which has a molecular weight throughput limit of 10kDa. Nitrogen gas was used to pressurize the cell in a 4°C cold room. The solution was mixed continuously to avoid precipitation of the protein onto the membrane. The magnetic stirrer was set on the lowest possible level to avoid denaturing the protein. When necessary, this method was used to replace buffers by concentrating and re-diluting the sample in the new buffer until the desired pH was achieved. For all samples at pH5.4, a buffer containing 10mM citrate titrated with HCl to pH 5.4 in 350mM NaCl was used.

For the *G. sorghi* MB2313 preparations, the samples were salt precipitated prior to loading onto the anion exchange column. Ammonium sulphate was added to 50% saturation [30g/100ml] and samples were then centrifuged at 18000 rpm (10000xg) in a Beckman JA 20 rotor for ten minutes at 4°C.

2.1.5 Anion Exchange Chromatography

For the *B .pumilus* MB2890 preparations, all subsequent steps were performed at 4°C however for the *G. sorghi* MB2313 preparations, the anion exchange chromatography was performed at room temperature. A Q Sepharose® anion exchange matrix packed into a XK 50/70 column (Amersham Biosciences) was used for the anion exchange chromatography. The column was equilibrated using a buffer with 50mM Tris, 50mM NaCl pH 8. The concentrated cell-free extract was loaded onto the column using a peristaltic pump, after which the column was flushed with at least one column volume of the equilibration buffer to rid the column of unbound material.

For *B. pumilus* MB2890, the column was washed with 10mM Tris, 250mM NaCl, pH 8 until the OD_{280} had decreased to less than 0.1 units in order to elute the protein interacting weakly with the matrix. A linear salt gradient from 300mM - 500mM NaCl in 10mM Tris pH 8 was then applied over a volume of 560ml to elute the enzyme. The fractions were collected at 7 minute intervals at a flow rate of 1 ml/min and were assayed for cyanide degrading activity using the picric acid assay described below. Active fractions were pooled and concentrated to 2ml using a 50ml Ultrafiltration cell (Amicon). For *G. sorghi* MB2313, the column was washed with 50mM Tris, 300mM NaCl, pH8. The enzyme was eluted using a linear salt gradient from 300mM - 1M NaCl. The enzyme eluted between 400 - 500mM. In order to perform gel filtration, fractions showing activity were ultrafiltrated and dialysed against 50mM tris, 300mM NaCl, pH8.

After use, the Q sepharose column was flushed with a column volume of 10mM Tris, 1M NaCl, pH 8. Further washing with 0.5M NaOH was passed through the column for 30 minutes if any discolouration was still present on the column. The column was stored in 20% ethanol at 4°C. All solutions used for columns were filtered and degassed.

2.1.6 Automated fraction collecting and absorbance monitoring

An automated Gilson fraction collector connected through the key port of Gilson 506C System Interface to a PC operating with Microsoft® Windows was controlled using the Unipoint™ system software (Gilson, Inc). This system software was also interfaced with a ThermoSpectronic Philips spectrophotometer with a flow-through cell so that the protein concentration of column eluates could be continuously monitored at $OD280_{nm}$. Data points were continuously collected by the Unipoint™ software which could later be plotted using a Microsoft® Office Excel spreadsheet. Additional to absorbance monitoring, protein concentrations were also measured using the Bradford assay (Bradford, 1976) with bovine serum albumin as a standard.

2.1.7 Gel filtration chromatography

Sephacryl 400HR in a 26/100 C column (AP Biotech) with a bed volume of 420ml was used for the *B. pumilus* MB2890 gel filtration chromatography. The concentrate was applied to the Sephacryl 400HR gel filtration column ($2X10^4$ - $8X10^6$ Mr fractionation range) and eluted with 10mM Tris, 350mM NaCl, pH 8 over a total volume of 387ml. Fractions were collected by a Gilson fraction collector at 7 minute intervals at a flow rate of 0.7 ml/min. Active fractions were pooled and Amicon concentrated to 5ml. The column was washed with buffer until baseline OD_{280} and thereafter stored in 20% ethanol. For *G. sorghi* MB2313, a pre-packed 40x1.2cm gel filtration column with Sephacryl S-300HR ($1X10^4$ – $1.5X10^6$ Mr fractionation range) was used with 50mM tris, 300mM NaCl, pH8 buffer.

2.1.8 Picric acid activity assay

Cyanide degradation was assayed using the picric acid activity assay according to the method of Fisher and Brown. Eighty microlitres of sample was added to 20µl of 100mM KCN in 500mM Tris pH 8 in a 1.5ml microtube and incubated at 37°C for 30 minutes. Afterwards, equal volumes of 1% picric acid and 0.5M Na_2C0_3 were premixed and 80µl of the mixture was added to each sample. Samples were then boiled for 5 minutes with lid locks on the microtubes to prevent them from opening. The sample was brought to 1ml with distilled water and assayed for absorbance at $OD520_{nm}$. Maximum $OD520_{nm}$ was determined for each experiment with a control which contained cyanide but no enzyme. The cyanide stock bottle was kept within a tightly sealed container at 4°C. Tips and microtubes as well as liquid picric acid/cyanide complex waste were disposed of in a glass Winchester bottle and disposed of accordingly by Waste-Tech.

To avoid unnecessary use of cyanide, every tenth sample of the chromatography fractions was assayed in order to narrow down the fractions containing active enzyme. Once a fraction was seen to be active, the other fractions within the range were also tested and activity curves were generated for those samples only.

2.1.9 Assessing sample homogeneity

A Bio-Rad Protean III 20cm cell was used to perform SDS polyacrylamide gel electrophoresis (PAGE) on the fractions to determine purity of the samples. SDS-PAGE gels and buffers were made up according to the method of Laemmli (Laemmli, 1970), however mass:mass ratio of acrylamide to bisacrylamide was 200:1. Separation gels were 15% acrylamide. Samples were boiled with sample application buffer for five minutes prior to loading on the SDS-PAGE gel. Sample application buffer consisting of 25% glycerol, 2% SDS, 0.1% bromophenol blue, 60mM Tris-HCl pH6.8, 14.4 mM and 2-mercaptoethanol was stored as a 5x stock at 4°C. Electrophoresis was carried out at a constant voltage of 150V.

As a general rule, gel filtration chromatography and SDS-PAGE do not completely reflect the status of sample purity. Often a protein may show up as a single band on an SDS PAGE gel, but the particles will look heterogeneous when viewed by negative stain EM (Ohi et al., 2004). For this reason, the active samples were also qualitatively analysed by negative stain electron microscopy. It was discovered that this method was quick and accurate for assessing the purity of samples.

CHAPTER 2.2

ELECTRON MICROSCOPY

2.2.1 Negative stain electron microscopy

Optimum protein concentrations for negative stain were determined to be between 0.2 and 0.3mg/ml. To achieve this, the purified enzyme preparation of the most active Sephacryl 400HR eluates were diluted $^{1}/_{10}$ in the required buffer. Copper-grids which were previously coated with thin carbon support film were glow discharged in air for 20 seconds at 20mV. Each grid was then held in reverse-grip tweezers over a petri dish and 5µl of the respective

protein solution was pipetted onto the carbon and left for 30 seconds. This method is preferred since any denatured protein aggregates on the air-water hydrophobic interface which may then be blotted off with filter paper. The grid was then washed with 2% unbuffered uranyl acetate (±100 µl) and blotted until a thin film remained. The grid was then air-dried.

Electron microscopy was performed on a LEO 912 AB transmission electron microscope operating at 120kV with zero-loss energy filtering. The microscope was aligned using the software EsiVision© Pro v3.2 (Soft Imaging System GmbH). Fine adjustments or corrections during use were done by hand. The grid was placed in the symmetric position (specimen holder θ value = 0). The microscope was stigmated at working magnification of 50 000X with the aid of the continuous FFT generated by the EsiVision© software. The minimum dose technique (Williams and Fisher, 1970) was implemented using the TVIPS software (Tietz Video and Imaging Processing Systems, 2003), in order to reduce exposure of the sample to damaging beam radiation. Images were taken at 50 000 times magnification and were recorded at 14µm pixels by a Proscan 2048×2048 slow-scan CCD camera. Images were stored as unformatted 16 bit arrays (RAW format) as well as MRC format.

For the *B. pumilus* data, electron microscopy was performed on a Jeol 1200CX transmission electron microscope. The microscope was aligned manually according to standard practice and the grid was placed in the eucentric position. Film micrographs were taken at 50 000 times magnification with spot size 4 using the minimum dose technique. Micrographs were then developed and scanned on an Ilford Leafscan™ 45. Resolution of the scans was 2540dpi which corresponds to a sampling of 2Å per pixel.

2.2.2 Shadowing experiments

Samples of *G. sorghi* CHT and *B. pumilus* CynD were taken to the ETH (Eidgenössische Technische Hochschule) Hoenggerberg, Zürich in order to determine the absolute handedness of the helical structures. Samples were applied to carbon-coated grids and absorbed for two to three minutes. The grids were washed twice in distilled water, followed

by blotting to remove excess liquid, and finally quick-frozen in liquid nitrogen. Frozen grids were then transferred to the Midilab instrument (Gross et al., 1990), a unique design in the laboratory of Heinz Gross at the Electron Microscope Unit of the ETH (Eidgenössische Technische Hochschule). The instrument consists of a freeze-drying/metal shadowing unit, mounted onto the column of a Philips-CM12 electron microscope at the height of the specimen stage which allows for a direct vacuum cryotransfer of specimens into the microscope onto a modified GATAN cryo-holder.

In the Midilab instrument, samples were freeze-dried for two hours at a temperature of 193K (-80°C) and a pressure of 1027 mbar. Completely dried specimens were unidirectionally shadowed with a 0.8 nm thick layer of tantalum/tungsten (Ta/W) at an elevation angle of 45°. Micrographs of shadowed specimens were recorded with a GATAN-794 Multiscan CCD camera (GATAN, Pleasanton, CA, USA), using an electron dose of 500–1000 electrons/nm^2. Nominal magnification of the microscope was 35 000 times, with an additional after-magnification ensued from the camera of approximately 1.5 times. The images were recorded on a 1K chip at 24μm pixels, providing a sampling of 5.3 Å/pixel.

2.2.4 Cryo electron microscopy

The cryo-EM data set was recorded by Prof. Trevor Sewell at the European Molecular Biology Laboratory in Heidelberg, Germany - courtesy of Dr. Andreas Hoenger. Samples of *G. sorghi* CHT were applied to holey-carbon grids and were plunge-frozen in liquid ethane. The frozen hydrated samples were placed on GATAN 626 cryostage cooled to ~90 K by liquid nitrogen. Transmission electron microscopy was performed on a Philips CM 200 FEG at 200 kV in low-dose bright-field mode. A Tietz 2k CCD camera was used to record the micrographs at a nominal magnification of 38 000 times. An attempt was made to set the underfocus values between -2.5μm and -3.1μm in order to achieve an optimal balance between contrast and resolution.

CHAPTER 2.3

IMAGE PROCESSING

2.3.1 Processing equipment

Image processing was performed on an Intel server with two 2.2 GHz Xeon processors and 2Gb RAM operating under the Fedora Linux operating system. The programme SPIDER (Frank et al., 1996) was used to perform all processing and reconstruction steps (http://www.wadsworth.org/spider_doc/spider/docs/spider.html).
From the EMAN suite (Ludtke et al., 1999) of programs, V2 was used to view the images and transforms and the program BOXER was used to select boxes along helices. The programme WEB (Frank et al., 1996) was used to generate densitometry profiles across the layer lines of the power spectra. The program CHIMERA (Pettersen et al., 2004) was used to interactively view the models in 3-space and was also used to generate the images of the electron density maps and pdb models.

2.3.2 Particle selection

Application of the single particle approach to continuous extended helical fibres is made possible by dividing images of individual filaments into segments of appropriate length. This is achieved by selecting boxes along the length of the helix.

Micrographs which had been recorded in MRC format could immediately be used by the program BOXER. This program greatly simplifies the picking process since it allows one to select a box on either end of a helix, and by defining the amount of overlap required, it automatically places boxes along the length of the helix. Helices were selected with 256x256 boxes using the option 'rotate helix' which then provides an initial particle alignment so that the length of the helix lies approximately at either 0° or 180° in plane. A 90% overlap was used.

Once the images had been picked, they were binned using the IP command in SPIDER, to decimate the image by a factor of two in both dimensions, resulting in a 128x128 pixel box. When boxes are reduced by an integer quantity in both dimensions, SPIDER performs a quadratic interpolation which is equivalent to two separate steps of bi-linear interpolation in each dimension. The interpolation step consequently reduces the amount of data and notably decreases processing time.

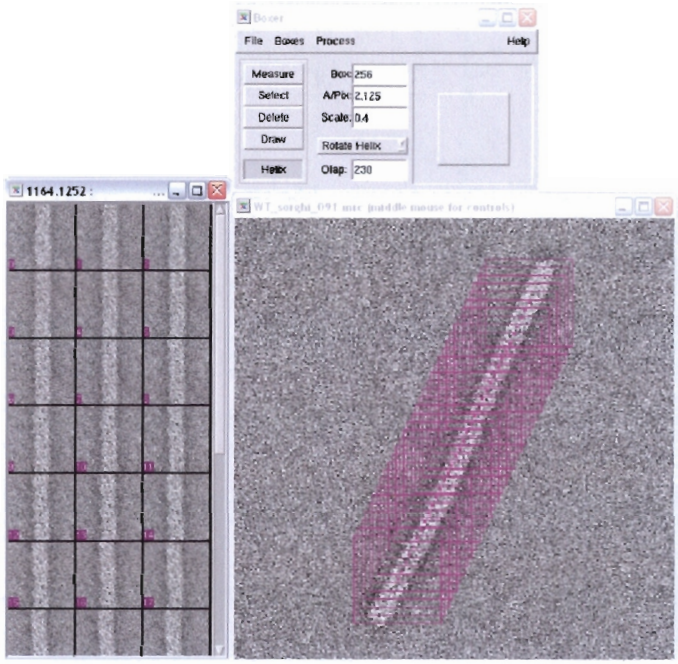


Figure 2.3.1: The interface of the EMAN program BOXER which simplifies the particle selection process of helices by automatically selecting along the length of the helix by a given overlap and also allows interactive selection and deletion of boxes.

The original sampling of the negative stain CCD micrographs was 2.125Å/pix so interpolation increased sampling to 4.25Å/pix, which was deemed to be more than adequate considering that expected maximum resolution attained from negative stain is 20Å. However, it is useful for the signal to be oversampled in order to minimize accumulating interpolation errors during the reconstruction. Sampling should at least be equal to the Nyquist-Shannon sampling frequency which is twice the value of the highest frequency present in the signal. If we accept 10Å as the required sampling rate according to the Nyquist-Shannon theorem, then 4.25Å/pix still corresponds to an oversampling of 2.35.

2.3.3 Filtering and normalization

Images were Fourier filtered to reduce the amount of noise and low frequency variation using the FF command in SPIDER. High-pass truncation was performed at a filter radius of 1/270Å which is twice the value of the helical diameter. A low-pass Gaussian filter function of 1/20Å was also applied. The images were then normalized to convert all pixel values to a common scale with an average of zero and standard deviation of one. This is calculated for each pixel in the image as the ratio of the difference between the actual pixel value and the average divided by the standard deviation. Normalization is required in order to prevent differences in image contrast from affecting the cross-correlation scores during alignment.

Images were visually inspected and those containing more than one helix were removed from the data-set using the program V2 (EMAN), since these images would not be normalized properly. Thereafter, images underwent a dust removal procedure using the command DU in SPIDER using an appropriate standard deviation factor in order to eliminate all pixels in the image that are more than a given multiple of the standard deviation away from the mode of the histogram. The eliminated pixels were then set to the value of the boundary of the range. This was to ensure that during the alignment procedure, no artificial cross-correlation scores would arise from alignment to areas of dust. For the cryo data-set, image grey-levels were inverted before continuing since the alignment procedure uses correlation of light-coloured pixels and in the cryo data set the biological material appears dark.

2.3.4 2D Averaging

For the shadowing data, the AP CA routine in SPIDER was used for reference-free alignment of rotationally aligned images via k-means clustering (Penczek et al., 1992). The alignment file produced by AP CA was then used by the AP SA routine to translate the images in order to produce a 2D average.

CHAPTER 2.4

ITERATIVE HELICAL REAL SPACE RECONSTRUCTION

The method of Iterative Helical Real Space Reconstruction (IHRSR) (Egelman, 2000) was used to reconstruct the filaments in three dimensions. The initial steps are the same as those implemented in the single particle approach but two additional procedures (described later) are used to inscribe the helical symmetry.

2.4.1 The generator program

A graphical user interface called generator, created by Egelman, was used to produce the IHRSR script to run the procedures in SPIDER. This script contains all the variables required for input at various stages of the reconstruction and also implements the SPIDER routines. Once it had been generated, the script was altered manually when the need arose.

2.4.2 Pre-centering images in the x-dimension

Initial centering of filament segments in the x-dimension is required in order to be able to eliminate helices with large shifts during the alignment procedure. This was done using an iterative procedure where images were aligned to an average produced by AS R using the AP MS routine in SPIDER to shift align the images (no rotation) in the x-dimension (with y search range set to zero). The INT function was then used to integerize the x-shift value so that no interpolation occurred when the shifts were applied. The SH function was used to create an x-shifted image stack which was then used to create a new average. It may also be possible to use a projection of an appropriately sized cylinder for this procedure.

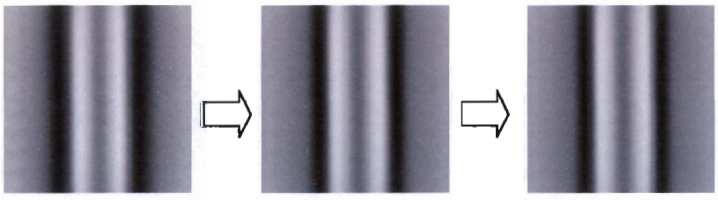


Figure 2.4.1: The three averages produced during the initial centring of the helices in the x-dimension. Noticeably the average becomes more defined and the x-shift values begin to remain the same after approximately three alignment steps.

2.4.3 Projection

The PJ 3 routine was used to compute projections of the volume according to three Eulerian angles, namely phi, theta and psi. During projection, phi is defined as the azimuthal rotation angle which is altered from 0° to 360° by a 4° increment, thereby producing ninety reference projections (θ and ψ are set to 90°). For the cryo 'out-of-plane' reconstructions, 630 reference projections were generated by combining two sets of Eulerian angles in a rotation about the azimuthal axis and a tilt of theta from 75° to 105° in five degree increments. A definition of these Euler angles as per the SPIDER context is shown in figure 2.4.2.

According to the Crowther criterion (Crowther et al., 1970), the number of evenly distributed reference projections required is equal to $\pi D/d$ where D is the largest dimension in the object reconstruction and d is the required resolution. Therefore, approximately twenty ((135*π)/20) evenly distributed reference projections rotated about θ are required to reconstruct the nitrilase helix to a resolution limit of 20Å. However, the *ramp filter* or *R* weighted filter, as of the BP 3F routine in SPIDER requires that there be an oversampled even distribution of projection images to avoid loss of detail (Paul et al., 2004). By using 90 projection images an oversampling factor of 4.5 is achieved.

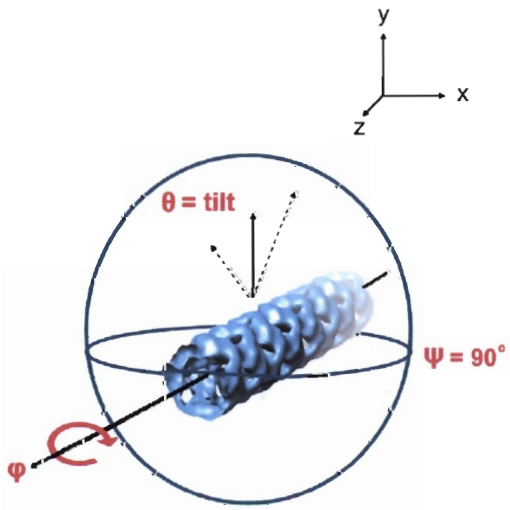


Figure 2.4.2: The coordinate system of Euler angles as defined by SPIDER during projection of a volume using the routine PJ 3. The volume lies along the φ axis, and the azimuthal rotation angle is defined by this angle. Tilted reference projections were produced by combining two sets of Eulerian angles in a rotation about the azimuthal axis and a tilt of theta from 75° to 105°.

2.4.4 Multi-reference alignment

In this procedure, all the experimental images are cross-correlated with all the references during alignment. Each image is then assigned to the reference that yields the highest correlation coefficient which thereby consigns its azimuthal orientation. The AP NQ routine implements the alignment procedure and also determines the required rotational and translation alignment (in-plane Eulerian rotation, x-shift and y-shift) by aligning the image to the most similar reference image. The minimum size of the search range in AP NQ is defined by the axial rise per subunit in pixels. In this case, due to their inherent two-fold symmetry, the dimers were considered as the asymmetric subunit during reconstruction. Since the axial rise of the dimers was initially estimated to be around 20Å (according to the terminating *P. stutzeri* structure), at a sampling of 4.25Å/pix the largest distance in the y-direction that any image could be from aligning correctly to a reference projection was 2.35 pixels. Initially, a search range of three pixels was used, but once the rise was discovered to be around 15Å, iterations were continued with a two pixel search range.

Once the alignment file is produced, images of poor quality may be eliminated on the basis of certain criteria. Considering that the picking procedure had constrained the helices to be oriented in-plane at either 0° or 180°, particles that deviated 10° either side of these values (providing a margin for imperfect picking) were eliminated. Images were also excluded on the basis of extraneous x-shift or y-shift values. Since the images had been pre-centred in x, there would be no reason - besides anomalous alignment to artefact - for images to have large x-shift values. The same applies for the y-shifts by virtue of the fact as mentioned above, that the largest distance in the y-direction any image could be from aligning correctly to a reference projection is 2.35 pixels. In some instances, a threshold was applied by which images with sufficiently poor cross-correlation were also excluded from back projection. This was subjectively applied after analysis of the correlation histogram. For the cryo data, two different strategies were applied - one in which the out-of-plane images were excluded from back projection, and another in which images aligning to templates obtained from models tilted by up to 15° were included in back projection.

2.4.5 Applying translation and rotation parameters

Using the parameters obtained from the AP NQ output file, images retained after the elimination process were then rotated and shifted using RT SQ. This routine averages adjacent pixels by quadratic interpolation. Images are first rotated counter-clockwise around the centre and thereafter the x-shift and y-shift is applied. The alignment was performed with only one interpolation of the input image utilizing the original image stack each time, thus avoiding the problem of accumulation of interpolation errors during alignment.

2.4.6 Back projection

This new stack of aligned images was then assigned a set of azimuthal angles according to the best matching reference templates and a 3D volume was generated by back projection using the BP 3F algorithm. This routine uses a Keiser-Bessel interpolation scheme where it is not required to specify a back projection radius (Frank, 1996). Commonly after the alignment step, images are grouped into class averages which are then back projected according to the projection angle conferred by the class reference projection in reconstruction of the 3D map. However, in IHRSR an equivalent but different process is implemented whereby images themselves as opposed to class averages are individually back projected. Once a volume has been produced, the symmetry search is executed.

2.4.7 Helical search and imposition

The helical search is performed by a program called hsearch which is implemented to identify and update the helical symmetry parameters which are written to a 'symdoc' file which then get applied to the refined structure in a helical imposition step (himpose). The hsearch algorithm is a least-squares search which explores the volume generated by back projection for two helical symmetry parameters, namely, the axial rise (Δz) and the rotation per subunit ($\Delta\varphi$). User input of the initial parameters, i.e. a reasonable estimate of the centre of the search and search increments, are required to start the procedure. The search is then made over a range of -10 to +10 times the search increment. In this case, search increments were both set to 0.1, search radius was set to 80Å and radius of imposition was set to 100Å.

The hsearch and himpose processes are performed outside of the SPIDER context and are implemented on the command line. A minor complication was that the intel byte order of the operating system required the model to be converted to opposite endedness using the routine CP TO OPEND in spider subsequent to back projection. The helical parameters are refined over successive cycles until they converge to a stable solution.

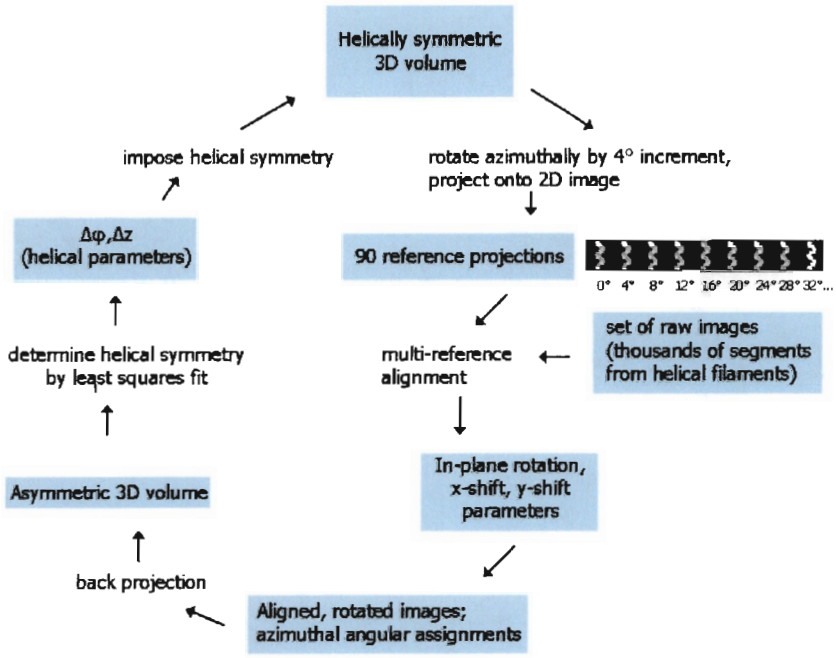


Figure 2.4.3: The cycle of iterative helical real space reconstruction (IHRSR). The process starts with a helically symmetric 3D volume from the previous round which is rotated azimuthally about the helical axis to generate a set of reference projections. Raw images are then aligned to the reference templates and five parameters are detemined, namely: the in-plane rotation angle, x-shift, y-shift, the cross correlation coefficient and the azimuthal angle. The in-plane rotation and shifts are then applied to each image. Aligned images are then used to generate a 3D volume by back projection. A least-squares procedure is then used to search for helical symmetry of the volume, which is then imposed once it has been detemined. The 3D reconstruction and inscription of helical parameters are refined over successive cycles until a stable solution is obtained i.e. the helical parameters Δφ and Δz converge. Adapted from (Egelman, 2000).

CHAPTER 2.5

POST RECONSTRUCTION ANALYSIS

2.5.1 Ancillary IHRSR data analysis programs

Several programs, written by Egelman, provided graphical display used to analyze the process involved in each cycle. The program HPLOT displays data from the alignment file produced by AP NQ. The program SPLOT was used to graph data from the symmetries file (symdoc). CPLOT was used to represent models as contour maps and is particularly useful for examining sections perpendicular to the helix axis. Initially, the HSEARCH algorithm allowed the searching of non-symmetrized volumes for helical symmetry. This program plots curves to shows mean-squared deviation in density as a function of changes in Δz and $\Delta\varphi$, and so a large search increment may be used to determine approximately where the mean-squared deviation is at its minimum.

2.5.2 Contour thresholds of volumes

It was necessary to calculate the appropriate threshold density value with which to contour the surface of the model in the graphics program CHIMERA. The volume of the protein can be estimated by dividing the molecular weight of the protein by the value of 0.73 daltons/$(Å)^3$ which is accepted for protein density, but this is difficult considering that an extended fibre is not comprised of a defined number of subunits. Therefore, to determine the number of subunits found within the box, the box size (128 pix) was divided by the previously determined axial rise of 3.53 pixels [Δz(Å)/scale(Å/pix)]. A value of 36 subunits derived from this calculation agrees closely to the number of subunits that can be counted by visual inspection of the model. A monomer of *G.sorghi* CHT is ~45 kDa, so a monomer volume occupies 45000/0.73=61644$(Å)^3$, which equals 2219184$(Å)^3$ for 36 subunits. To convert to a volume in voxels, one divides by the volume of a voxel $(4.25)^3$ so the final volume is 28908 voxels.

2.5.3 Generation of power spectra

Power spectra were produced using the PW 2 command in SPIDER. Images used were from the rotationally aligned and shifted stack produced by the RT SQ routine of the IHRSR process where poor quality images had been excluded. A power spectrum was generated for each image which was then padded to 512 pixels. The power spectra were then averaged using the AS R routine in SPIDER. For the volumes, the reference images generated from projection were used to generate power spectra. Since different rotational views of the volume are contained within a single projection, only one projection is needed to represent complete information about the diffraction pattern, but power spectra were generated for all projections and then averaged to increase the signal to noise ratio. These power spectra were also padded to a box size of 512 pixels before averaging. The average power spectra were windowed to a 200 box size using the WI function in SPIDER. Graphed amplitudes across layer lines were plotted using the line profile command in WEB.

2.5.4 Fourier shell correlation

Volumes were aligned rotationally and shifted in the z-direction after using the PK 3 in SPIDER to correlate the density peaks of the two volumes. The shifted volume was then windowed to remove the "wrap-around" effect that results from the shifting process in SPIDER. The windowed volume was then padded to its original size which has the effect of uncorrelating the axial mask function. A sharp-edged component is created in the volumes during the himpose procedure, so both reconstruction volumes were masked with a low-pass filtered cylinder at different radii to remove the sharp edge. The FSC was then calculated using RF 3 in SPIDER. A unity-width Fourier-space shell was chosen since this is accepted as the standard for reasons of consistency and comparability among independent data sets.

CHAPTER 2.6

DOCKING

Volumes were sliced using the VOLCROP command in SITUS to approximately three turns of the one-start helix (i.e. 14 subunits) and the cropped volume was used for docking. Initially, singular pdb models were used to dock the structure but once an approximate azimuthal angle had been found, helices were generated which were then used in the docking process.

The SITUS program COLORES (COrrelation based LOw RESolution) was used for the docking procedure which is a general purpose rigid-body search tool, suitable for situations where not all density is accounted for by the atomic model. The translational search is FFT-accelerated and supports the use of a Laplacian filter which increases the fitting contrast at low resolution. COLORES uses correlation based docking to perform an exhaustive 6D search (3 translational and 3 rotational degrees of freedom) required to find the best fit of an atomic structure/ high resolution model into a multi-molecule low resolution density map. For these docking procedures, an angular increment of 10° was used to search the EM map at a target resolution of 15Å. After the exhaustive search was performed, the best 20 on-lattice maxima were refined using Powell optimization. The automatically docked model was used to gain insight by annotating the positions of insertions and deletions and positions of deleterious mutations.

A FORTRAN script was written by Prof. Trevor Sewell in order to generate a helix given a pdb file, symmetry parameters (axial rise and rotation per subunit). The factors that could be altered in order to optimize the docking were the radius from the origin and azimuthal angle of rotation about the 2-fold axis of the dimer subunit.

The 3D volume was initially docked using a dimer model to fit to the structure. The approximate parameters of the 20 on-lattice maxima which had been refined were then used to generate a set of continuous helices by varying the azimuth around the 2-fold axis and the helical diameter at which the pdb was positioned. The helix models were then docked using SITUS to optimize these parameters.

SECTION 3 – RESULTS

CHAPTER 3.1

PROTEIN PURIFICATION

3.1 PURIFICATION OF RECOMBINANT CLONES

Recombinant clones of the cyanide degrading nitrilase genes of *B. pumilus* and *G. sorghi* were created in pET26b and were introduced into *E. coli* BL21(DE3) for expression in order to produce abundant active enzyme.

3.1.1 Fibre formation

It has been shown that several bacterial nitrilases form complexes which exist as terminating spirals with a defined number of subunits. As a case in point, cyanide dihydratase from *Bacillus pumilus C1* (CynD) forms a terminating spiral of 18 subunits at pH 7-8. At pH5.4, however, this enzyme forms large molecular weight complexes and has been shown to form extended fibres. In previous purification experiments (Berman, 2003), size exclusion chromatography was performed at pH5.4 in order to elute fibres in the void volume since this was found to rid the sample of contaminating GroEL particles. During these purifications, however, gel filtration chromatography was performed at pH8 and thereafter the buffer was exchanged using the Amicon cell to lower the pH to 5.4 to induce fibre formation. It was unnecessary to elute samples at pH5.4 since active fractions from the Sephacryl HR400 column were homogeneous. If any GroEL contamination were present, which was seldom the case, it was unproblematic to distinguish the nitrilase for picking once fibres had formed at pH5.4. Since the CHT from *G. sorghi* is consistently found as a large oligomer of MW greater than 300 kDa at any pH range, buffers were kept at pH8 throughout. *G. sorghi* CHT does not elute in the void volume and requires over-night incubation at 4°C at pH8 before fibres form.

3.1.2 The stabilizing effect of imidazole

It was serendipitously discovered during the purification of histidine tagged CHT that some of these enzymes require imidazole to induce fibre formation (Berman, 2003). Isolation of non-histidine tagged protein presented the opportunity to assess whether the fibres were able to form without the addition of imidazole. Although the addition thereof was seen to distinctly increase the length of the fibres, it was discovered that imidazole was not required and that fibre formation could be induced by leaving the samples overnight at 4°C in 300mM NaCl. The mechanism whereby imidazole assists helical extension is not yet known but is speculated to be comparable to increasing the ionic strength of the solution, thus forcing electrostatic interactions between the proteins. It may also possibly have to do with sequestering of repulsive positive charges, allowing the subunits to come into closer contact. For most of the *G.sorghi* samples, imidazole was added in order to stabilize the helices and lengthen them in order to simplify the picking process.

Imidazole thus proceeded to be a good indicator of fibre-forming ability which was utilized to test protein stability of recombinant mutants that had been expressed. Some only formed extended fibres when imidazole was added but others were inherently fibre-forming which was an indication as to the effect the mutation had on the enzyme.

3.1.3 Anion-exchange chromatography

For the *G. sorghi* samples, the active fractions eluted at 300mM NaCl. Fractions 13-23 (figure 3.1.2) of the *G. sorghi* samples from the Q-sepharose column were concentrated to 4ml to apply to the gel filtration column. For the *B. pumilus* samples, most of the protein from the cell free fraction eluted from the column in the unbound fraction (void volume) and a large percentage more in the washing step before the gradient was applied. As evidenced by the elution profile and the corresponding SDS PAGE gel, the fraction containing most of the protein during the gradient elution was fraction 45 but this was not active. The activity peak was shown to coincide with the small peak adjacently preceding the primary peak of the profile.

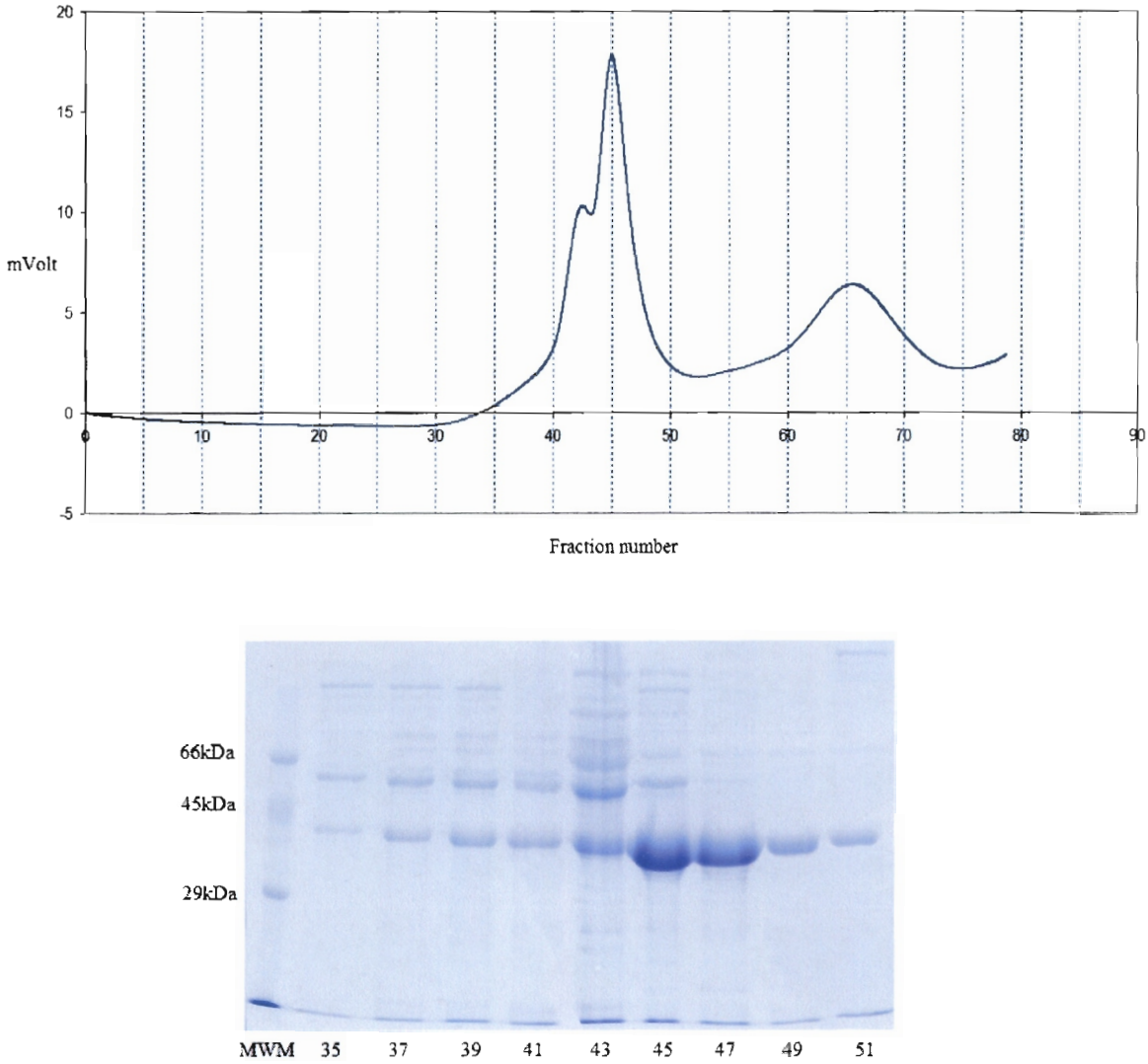


Figure 3.1.1: ABOVE: The elution profile from the Q-sepharose anion exchange column for *B.pumilus* CynD. The spectrophotometer output in mVolt is proportional to the absorbance at OD260nm. BELOW: SDS PAGE gel of some fractions from the Q-sepharose anion exchange column. The purified protein has a molecular weight of approximately 40kDa. Active fractions (33-37) were pooled and concentrated before applying to the gel filtration column.

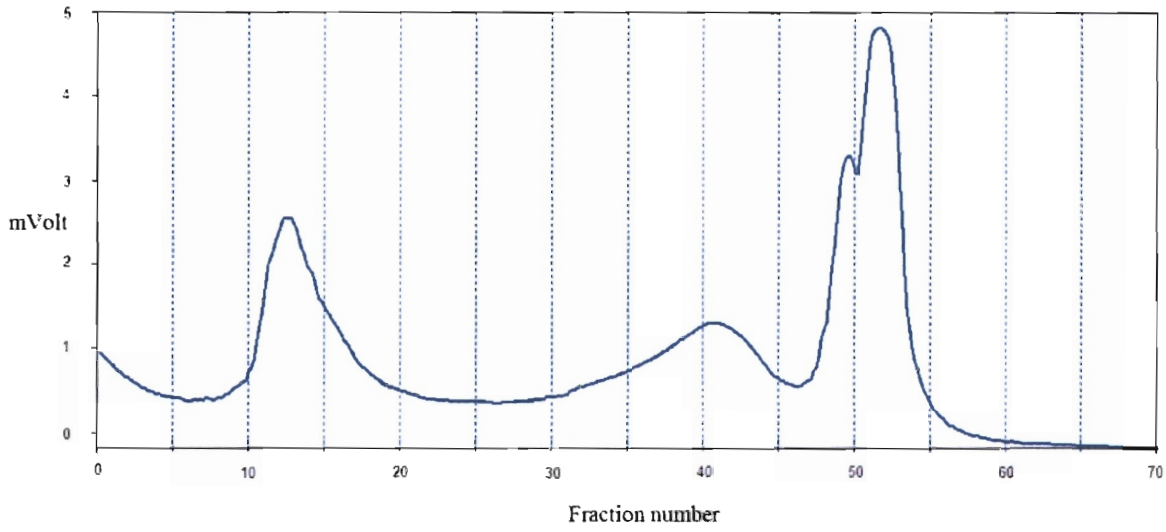


Figure 3.1.2: ABOVE: Elution profile from the Q-sepharose anion exchange column of the *G. sorghi* CHT purification. The spectrophotometer output in mVolt is proportional to the absorbance at OD260nm. Fractions 13-23 were active. All active fractions were pooled and concentrated before applying to the gel filtration column. CFE = cell free extract.

3.1.4 Gel filtration chromatography

Gel filtration profiles generally contained a solitary protein peak which corresponded directly to cyanide degrading activity. Both *G. sorghi* and *B. pumilus* profiles showed a distal trailing shoulder resulting in asymmetry of the peak. The purity of each fraction was assessed by SDS-PAGE using Coomasie blue staining and a 'three-component' band characteristic of the cyanide dihydratase was seen. After overnight incubation at pH5.4, preparations from the most active fractions (33-37) were negatively stained and viewed under the EM to investigate their homogeneity. Fibres of varying lengths were seen with no visible contamination of other proteins.

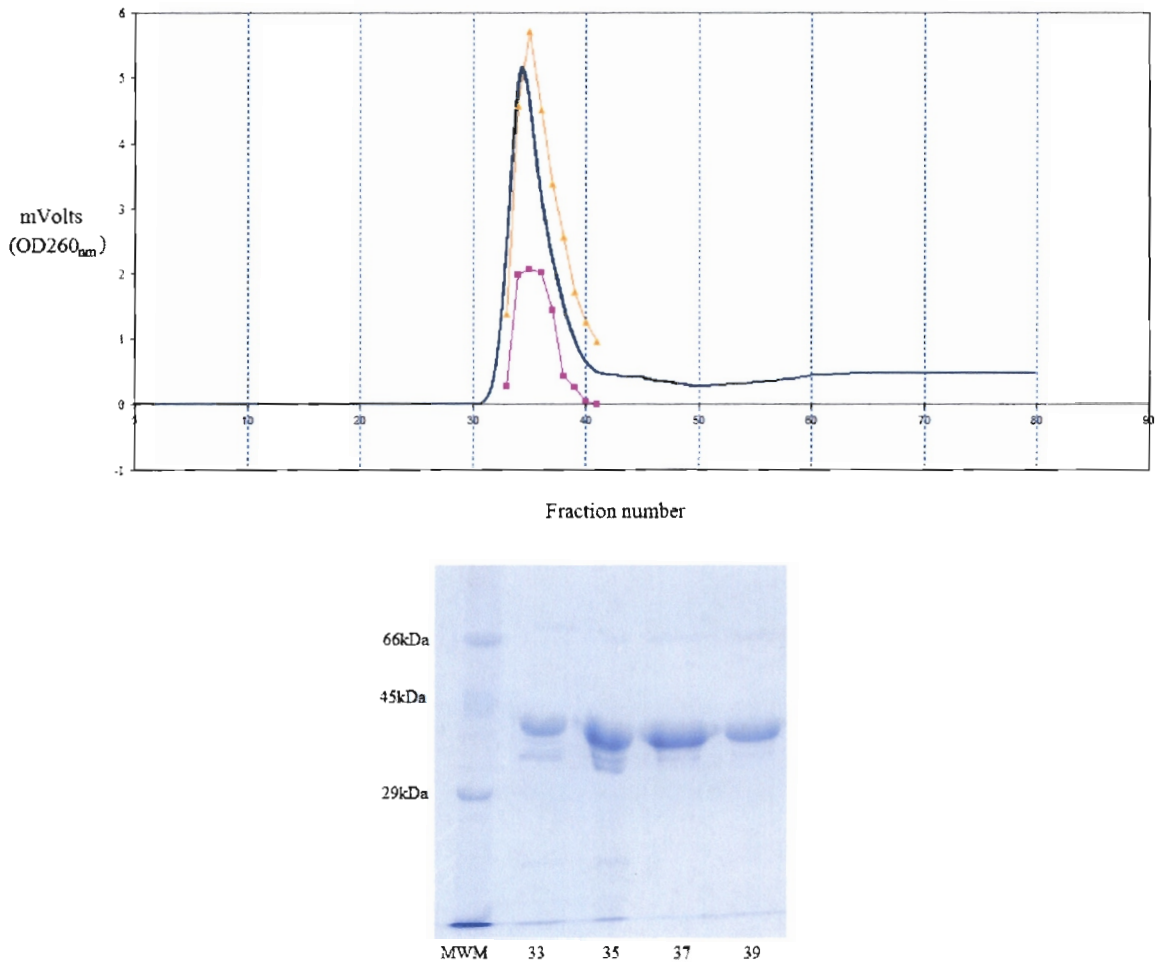


Figure 3.1.3: ABOVE: The elution profile from the S400 high resolution gel filtration column for *B. pumilus* CynD. The magenta curve shows relative activity from the picric acid assay and the yellow curve shows the normalized values from the Bradford assay. BELOW: SDS PAGE gel of some fractions from the S400 high resolution gel filtration column. Samples of fractions 33-37 were incubated at pH5.4 overnight and prepared for negative stain EM to assess the homogeneity.

3.1.5 SDS PAGE gel electrophoresis

The *B. pumilus* enzyme has been shown to consist of three separate bands on SDS-Page which are all part of the cyanide dihydratase. Previous experiments have shown that MALDI-TOF spectra of the different bands extracted from the SDS-PAGE are closely related in sequence. Also, during crosslinking experiments, all the bands disappeared from the monomer position upon increase of gluteraldehyde, indicating the polypeptides are from the same enzyme complex (Berman, 2003). These enzymes are possibly differentially post-translationally modified and thus run at slightly varying molecular weights. According to

the *G. sorghi* ORF, CHT should have a MW of 40.9 kDa, but on denaturing SDS PAGE, the polypeptide appears at ~45 kDa. Berman showed that these peptides appear to migrate anomalously on a gel – he showed that the CynD enzyme migrated at an apparent size of 41.4 kDa although its actual size is 37.4kDa. He attributed this phenomenon to an intrinsic property of the nitrilase polypeptide such as incomplete denaturation which causes their anomalous migration through a polyacrylamide gel. Wang *et al.* have previously resolved the 45-kDa polypeptide band into three isozymes of pI 6.1, 6.3, and 6.5(Wang et al., 1992).

CHAPTER 3.2
SHADOWING EXPERIMENTS

A factor which affected the resolution of the nitrilase samples to a large extent was sample drift which occurred upon exposure of the grids to the electron beam. This could be seen when inspecting the real-time FFTs of the images taken but could unfortunately not be avoided since all areas of the grid experienced drift effects. The method did however provide high-contrast, low artefact images which revealed a great amount of information about the surface of the nitrilase fibres. Reference-free averaging resulted in images with high signal-to noise ratio where striations were clearly seen. The power spectra of these averages revealed a great amount of information on the Bessel orders of the layer lines.

Handedness as revealed by shadowing

The shadowing experiments show striations running at a gentle slope from upper left to lower right indicating a left-handed one-start helix, as well as steep striations running from upper right to lower left indicating a right-handed helix of a different startedness (Figure 3.2.2). Although the samples from *B. pumilus* were in particularly poor condition, it was still possible to determine that these fibres had the same type of prominent left-handed one-start helix as the *G. sorghi* fibres – previous shadowing experiments had also suggested (but had not confirmed) the left-handedness of the *B. pumilus* fibres. The averages in particular

(figure 3.2.2 (2a & b)) demonstrate the similarity in the handedness of the fibres. A problem associated with the three-dimensional reconstruction process is that there are few ways to accurately determine the reliability and correctness of a reconstruction. The issue of enantiomorphic ambiguity is one factor to be considered, since both enantiomorphs of a structure will generate identical projections. One way of removing this ambiguity is by imaging metal-shadowed specimens to determine the correct handedness of the structure and subsequently applying the external information to the 3D reconstruction. The shadowing results were thus useful in the fact that the handedness could be constrained using the helical symmetry imposition during reconstruction.

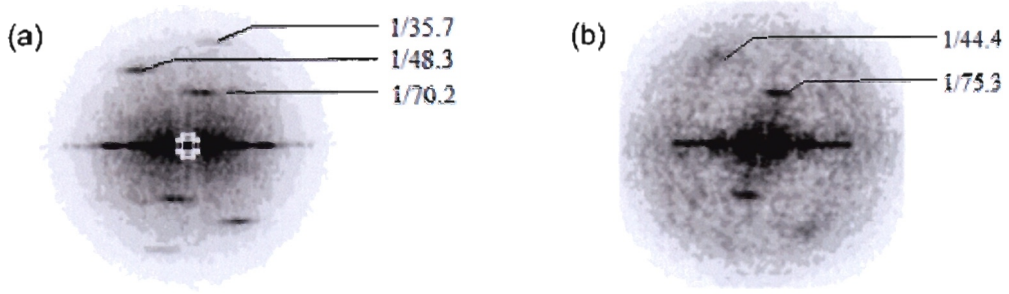


Figure 3.2.1: Power spectrum of the rotationally averaged image stack. Layer lines appear at 1/70.2, 1/48.3 and 1/35.7 for the *G. sorghi* CHT(a) and at 1/ 75.3 and 1/ 44.4 for the *B. pumilus CynD* (b).

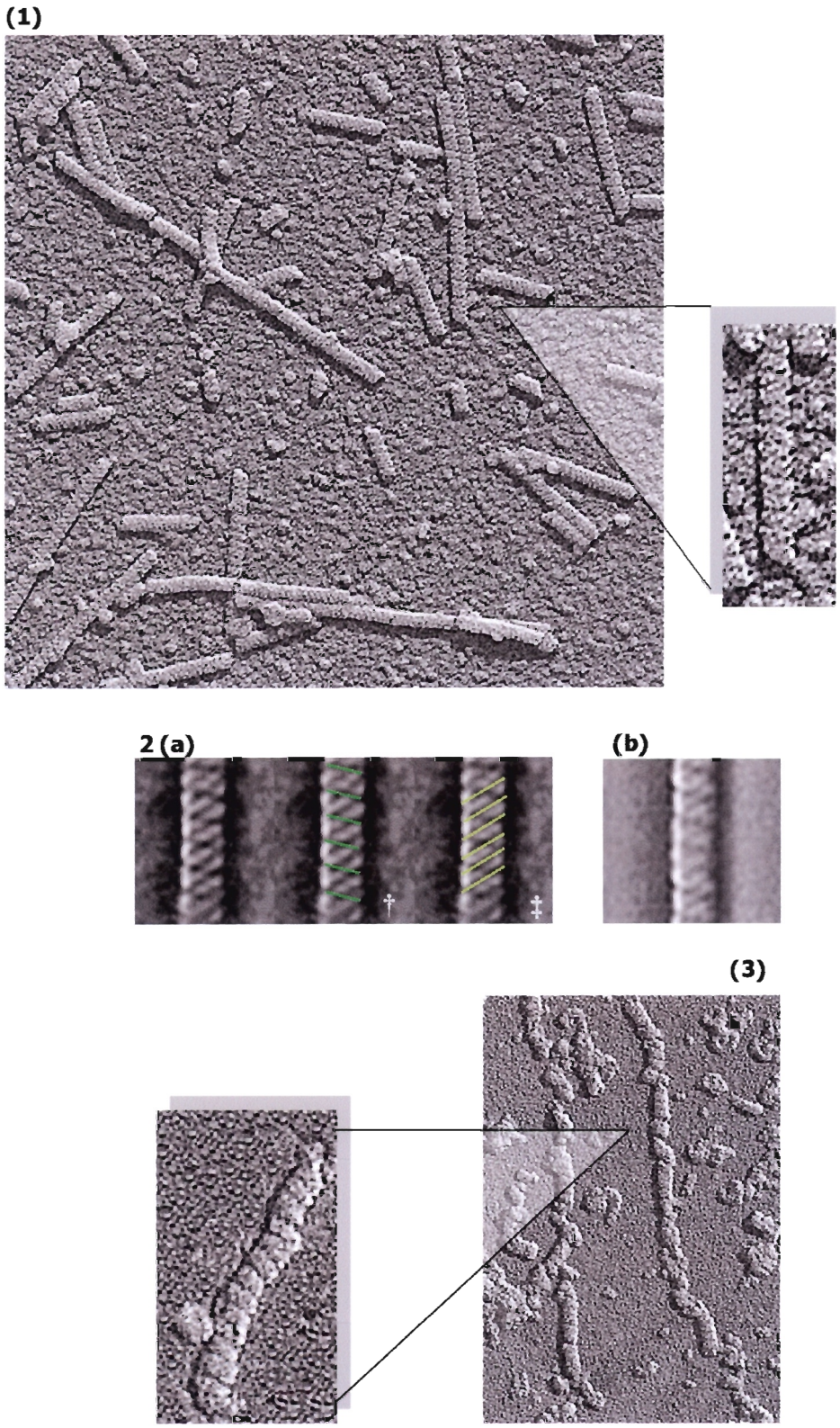


Figure 3.2.2: (1) Shadowed *G.sorghi* fibres shown with enlargement of a fibre. (2a) 2D average of the shadowed *G.sorghi* fibres showing the prominent leftward-sloping (†) and rightward-sloping (‡) striations. (2b) 2D average of the shadowed *B.pumilus* fibres. (3) Shadowed *B.pumilus* fibres with enlargement shown.

CHAPTER 3.3
ELECTRON MICROSCOPY

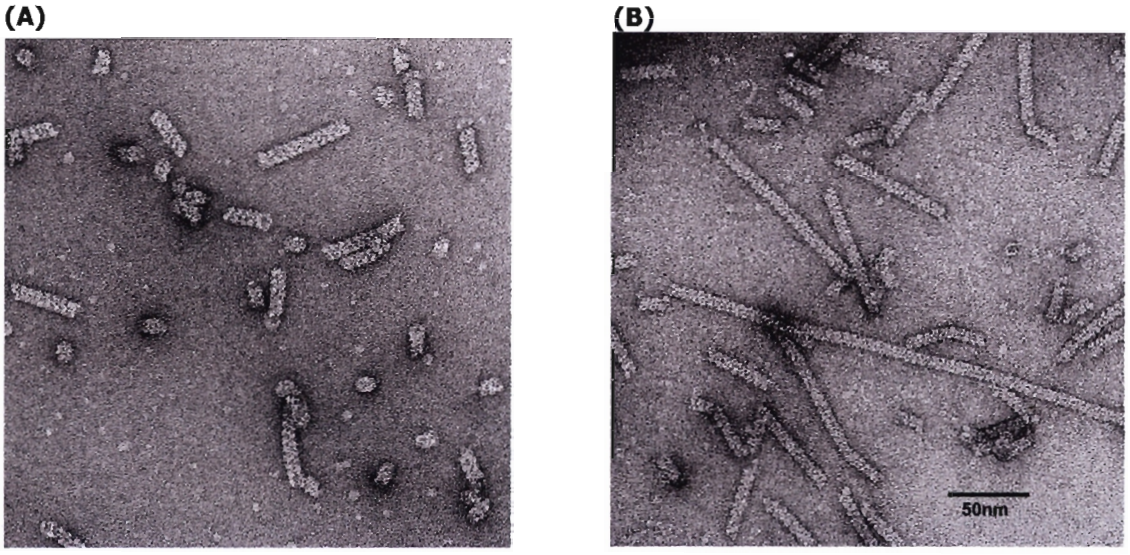


Figure 3.3.1: (A) *G. sorghi* CHT mutant R87Q without the addition of imidazole. (B) WT *G. sorghi* CHT produced extended long fibres of up to several hundred nanometres long.

Purified proteins were assessed by negative stain electron microscopy to be homogeneous long helical fibres. The fibres did not display any curvature and thus did not require processing with computational straightening algorithms. Fibres that aggregated or associated were not picked. The *G. sorghi* CHT mutant R87Q was active (Figure 3.3.1A), but as with the other active mutants (R84Q, R85Q, D92Q), the protein did not form extended fibres without the addition of imidazole. Wild Type *G. sorghi* CHT, on the other hand, inherently formed long extended fibres of up to several hundred nanometers. Some helices in the electron micrographs display clear one-start helices of the 'wrong handedness' (right-handed). Since images of negatively stained helices are formed by superimposition of the 'near' and 'far' walls of the molecule, it is the far surface nearer the carbon that is projected more clearly in these instances because of differential staining. The images shown here were not used for processing since they were taken at larger defocus values than usual in order to enhance the contrast for presentation purposes.

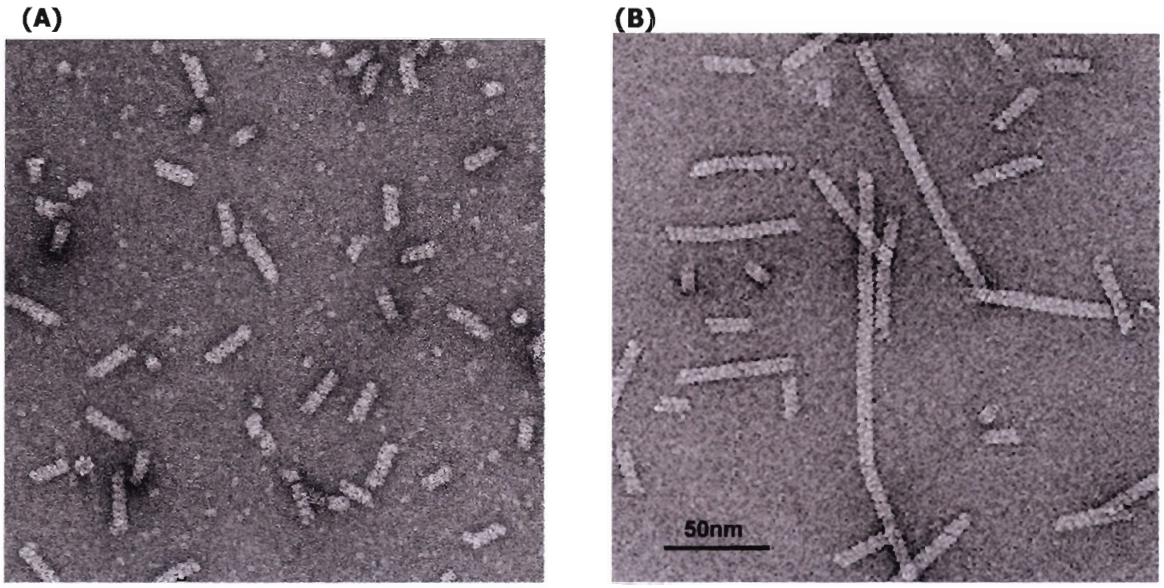


Figure 3.3.2: (A) Short *B. pumilus* fibres having eluted from the gel filtration column prior to incubation at 4°C overnight. (B) Long extended *B. pumilus* fibres subsequent to overnight incubation at 4°C .

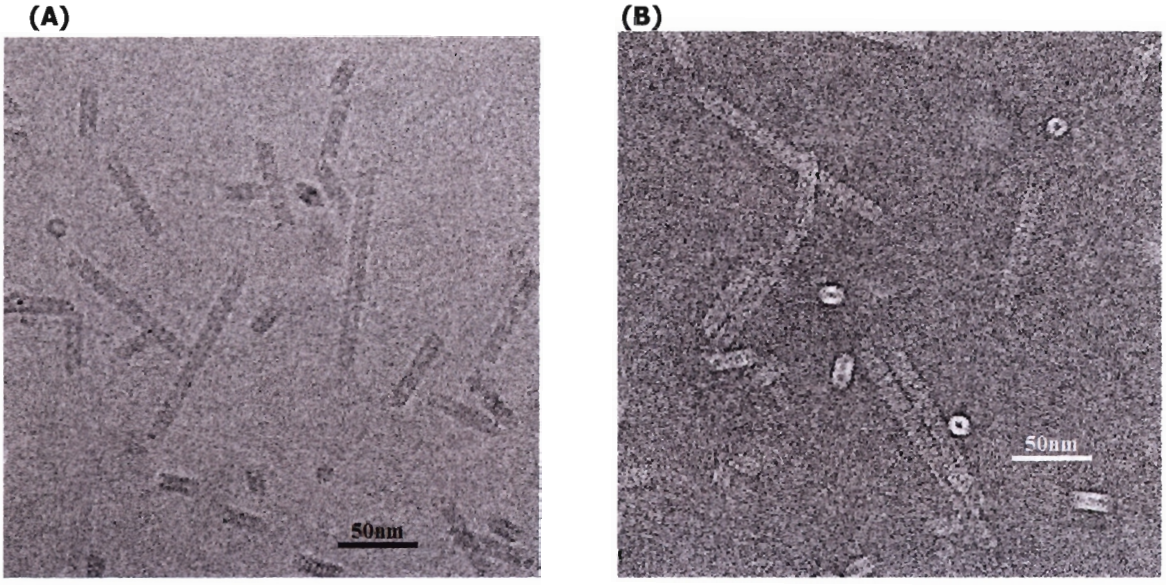


Figure 3.3.3: Cryo images of WT *G. sorghi* CHT. Many oblique and end-on views of the short fibres can be seen indicating the thickness of the ice. (A) Section of an original micrograph. For the cryo data-set, image grey-levels were inverted before continuing since the alignment procedure uses correlation of light-coloured pixels and in the cryo data set the biological material appears dark. (B) Section of a cryo image where the contrast has been inverted. Images were picked and processed with these grey-levels.

CHAPTER 3.4
ITERATIVE HELICAL REAL SPACE RECONSTRCUTION (IHRSR)

For all reconstructions, the helical imposition was constrained to be left-handed thus delta phi was set to a negative value. Initial starting symmetries for the search were estimated from the previous *P. stutzeri* structure which allowed the calculation of rise and rotation per dimer. This was a rough estimate since the *P. stutzeri* structure is a terminating spiral and it had already been suggested that the symmetries for the extended fibres would be different since the structure would have to accommodate the addition of more subunits. All $\Delta\varphi$ values quoted are in degrees and Δz values are in angstroms.

3.4.1 G. sorghi CHT negative stain reconstructions

Several independent *G. sorghi* CHT data sets were collected; one WT data set which was collected on film, as well as a WT sample and mutant (R87Q) which were collected using a CCD camera. For all data sets, two reconstructions were performed with different starting symmetries and different starting models. The first starting model was a cylinder filtered to 50Å and the second was a filtered structure produced by the helix generator program given the symmetry parameters $\Delta\varphi$ = -75.0, Δz = 15.0. The cylinder starting model was given initial search parameters of $\Delta\varphi$ = -78.0, Δz = 13.5.

After just a few iterations, the symmetry parameters of the first data set started to converge around the values $\Delta\varphi$ = -76.96 Δz = 14.99 which was almost identical to that of the second data set: $\Delta\varphi$ = -77.0, Δz = 14.87. The structures correlated to within 16Å resolution at a fixed-value threshold of 0.5. Even though individual images were back projected, 2D class averages were produced in order to visually assess whether any projection angles were not sufficiently covered. There was a distinct increase in signal to noise ratio when averaging images which aligned to the same template and all class averages looked like their corresponding reference projections (figure 3.4.1). The convergence of independent data sets to the same structure from different starting models indicated the presence of strong signal in the images and confirmed the reproducibility of the reconstruction (figure 3.4.2).

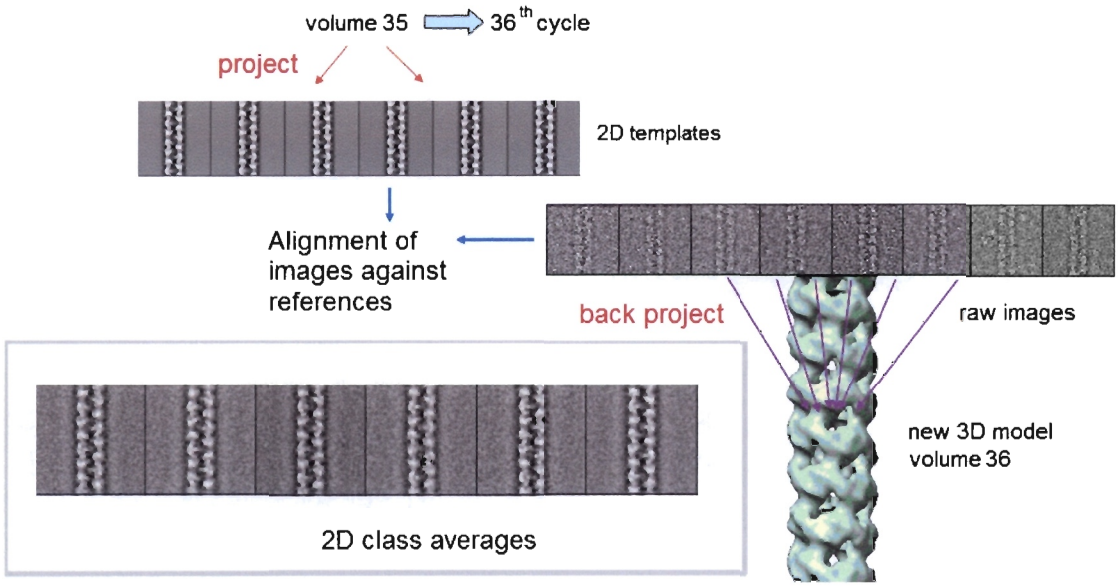


Figure 3.4.1: The reconstruction process shown on the 36th cycle where the volume generated from the previous cycle is projected to generate the reference images. The raw images are aligned to the templates; then rotated and shifted accordingly. Raw images are back projected to form a new 3D volume. The 2D class averages show a clear increase in signal to noise ratio as expected.

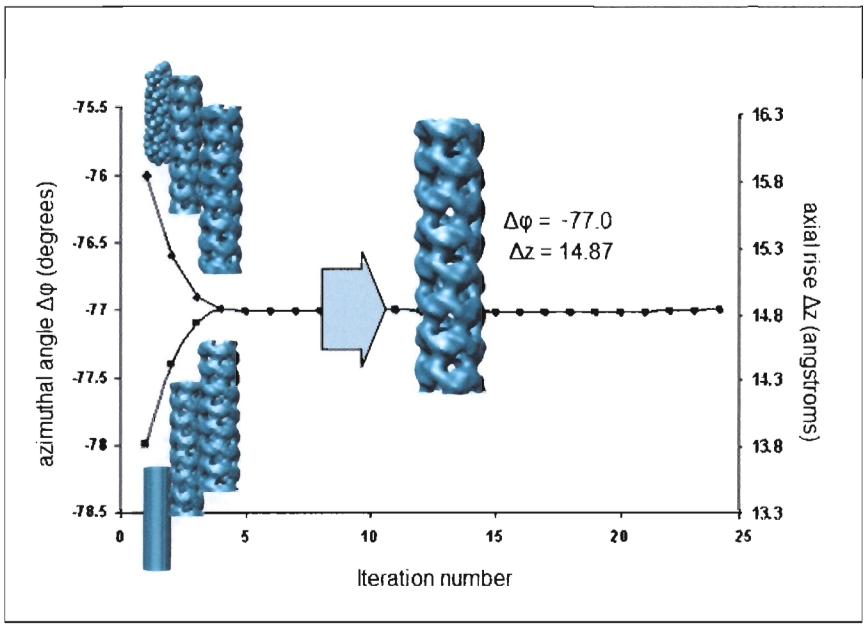


Figure 3.4.2: Convergence of the second WT negative stain reconstruction from two independent starting models, a cylinder and a structure produced by the helix generator program. The symmetry parameters converged around the values $\Delta\varphi$ = -77.0 and Δz = 14.87Å.

The validity of the reconstruction process was assessed by examining the distributions of in-plane rotations, reference frequency, x-shift and y-shift (see figure 3.4.3). In-plane rotations

showed characteristic modes around the values 0° and 180° with almost all images aligning within the (0°±10°,180°±10°) cutoff criterion. Importantly, the distribution pattern of images aligning to reference projections did not show preferentiality for a particular set of templates and thus could be considered statistically even. As is expected, correlation coefficients (CCs) showed a Gaussian distribution; no CC threshold was set as cutoff for back-projection during these reconstructions. In the first WT reconstruction, only 270 images were excluded from back projection and in the second WT reconstruction, 320 images were rejected on the basis of criteria mentioned in section *2.4.4*.

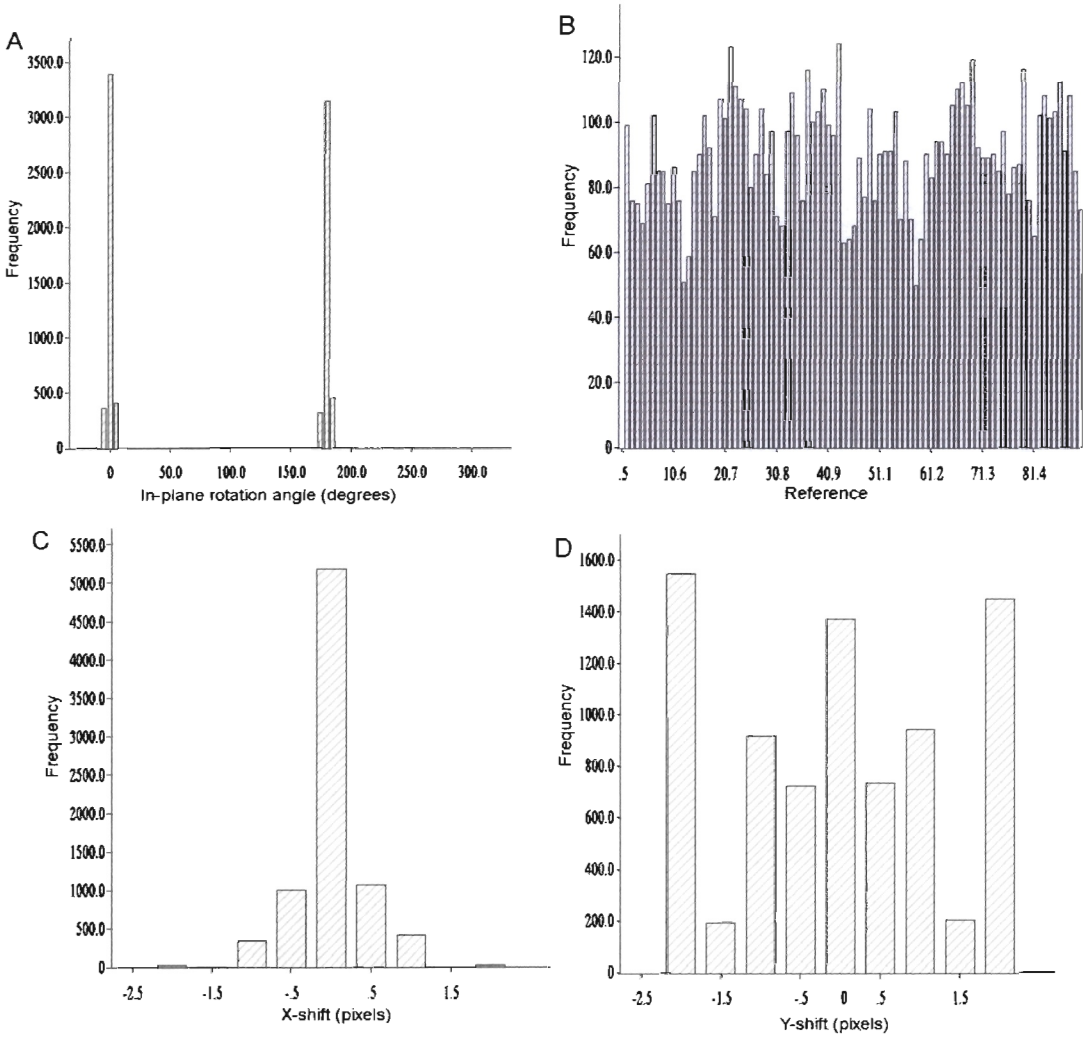


Figure 3.4.3: Figures produced from the hplot program of the last iteration for the second WT *G. sorghi* CHT reconstruction. The reconstructions from other negative stain data sets produced similar results (data not shown). (A) Distribution of in-plane rotation angles. (B) Uniformly distributed reference frequency with which images align to templates. (C) x-shift distribution centred around zero. (D) y-shift distribution showing a mode at zero and at ±2 pixels.

Y-shift distributions (figure 3.4.3D) show a mode at zero and at ±2 pixels, which is at the limit of the search range which possibly indicates that the images did not find optimal alignment within the search range given. The reconstructions were initially started with a larger search range of three pixels, but once the rise had been determined to be ~15Å a smaller search range was implemented for the remainder of the iterations. Reconstructions were restarted with larger search range of three and five pixels, but this did not make any difference to the structure or symmetry parameters at convergence. The image shifts in the y direction always seemed to cluster around the edges of the search range independent of the search range given. As previously mentioned, the minimum size of the search range in AP NQ is defined by the axial rise per subunit in pixels. Since the axial rise of the dimers was 15Å, at a sampling of 4.25Å/pix, the largest distance in the y-direction any image could be from aligning correctly to a reference projection was 1.76 pixels, so a search range of two should have been sufficient.

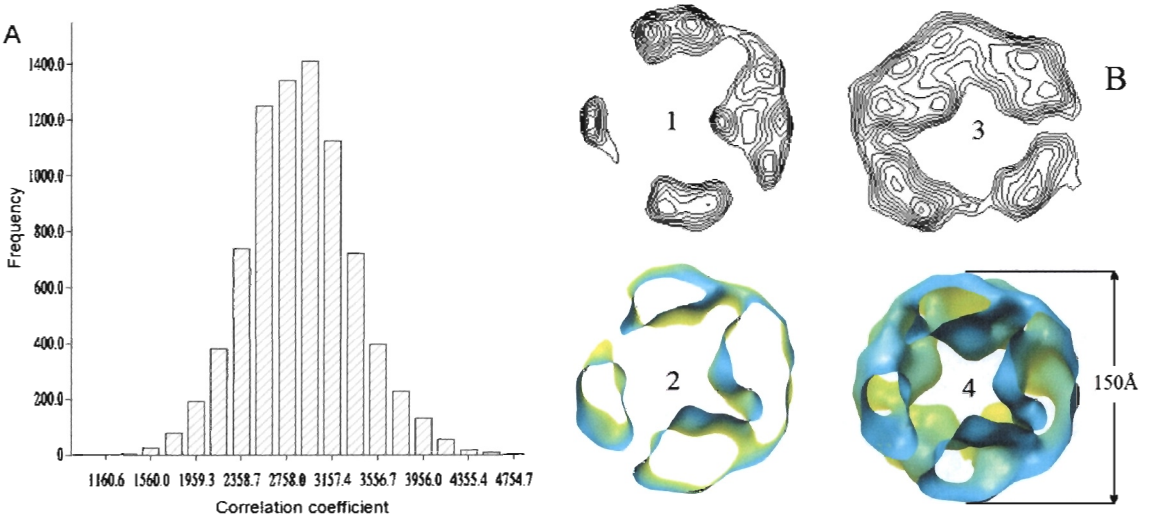


Figure 3.4.4 (A) Gaussian distributions of correlation of coefficients. (B) Lateral cross-sections of the WT *G. sorghi* CHT negative stain structure. (1) A single-pixel cross-section through z-section 3 at low contour threshold using cplot and (2) the corresponding cross-section shown in CHIMERA. (3) Z-section 64 at high contour thresholds and (4) the corresponding cross-section shown in CHIMERA, spanning ten pixels. Note the large protruding bulges as demonstrated in the lateral cross-section below. Diameter of the reconstructed helical density is approximately 150Å.

A prominent feature of the negative stain *G. sorghi* reconstruction is the left-handed one-start helix which has distinct leftward sloping ridges on its surface. A deep groove is formed by discontinuities in the structure above and below the one-start helix. At low contour levels the most outstanding feature is a right-handed steeply sloping four-start helix. The structure has large protruding areas of density (bulges) on the interior surface which are noticeable in both the lateral and the longitudinal cross-sections. The bulges protrude from corresponding areas on the inside surface of the structure where the centres of the ridges exist on the outside surface of the helix. Therefore, a dyad axis exists in the centre of the two rides and through the internal bulge on the inside surface. Narrow regions of density extend laterally in forming connections between the ridges and density arises vertically above and below the ridges to form a bridge across the groove.

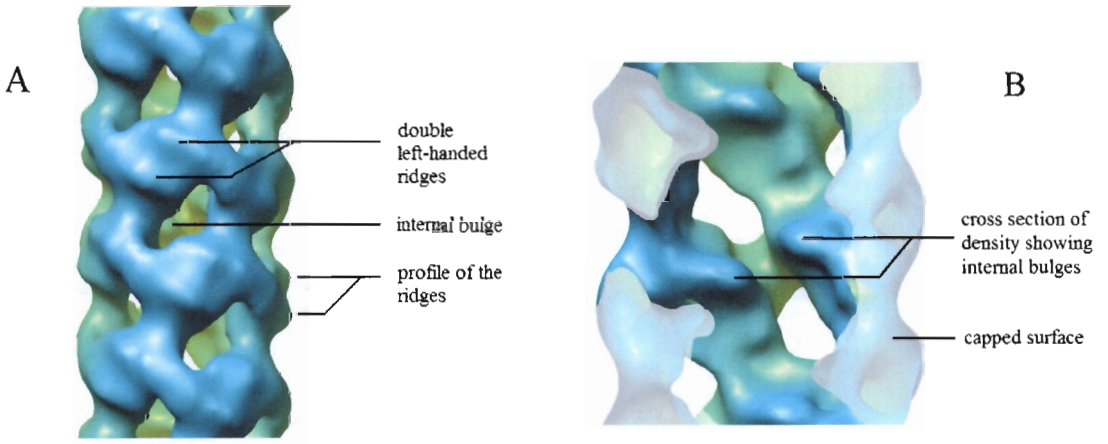


Figure 3.4.5: Features of the negative stain structure of *G. sorghi* CHT where depth cueing is shown in yellow (A) On the surface, the prominent left handed one-start helices can be seen with leftwards sloping 'ridges' (B) Longitudinal cross-section showing the prominent bulges of density on the inside of the structure. Surface capping is transparent white.

The convergence of independent data sets to the same structure from different starting models indicated the presence of strong signal in the images and confirmed the reproducibility of the reconstruction. Also, the power spectrum of the reconstructed volume looked like the rotationally aligned, averaged power spectrum of the image stack (figure 3.4.6). An FSC performed between the two independent data sets (film data set and CCD data set) which had converged from different starting models resulted in a value of 16Å resolution at a fixed-value threshold of 0.5 (figure 3.4.7).

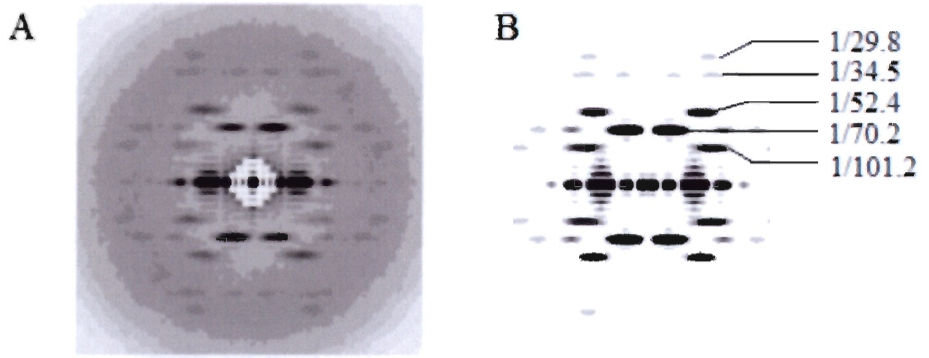


Figure 3.4.6: (A) Averaged power spectrum of rotated and aligned images (used for back projection). (B) 2D power spectrum generated from the reconstruction volume which mirrors the average power spectrum of the images.

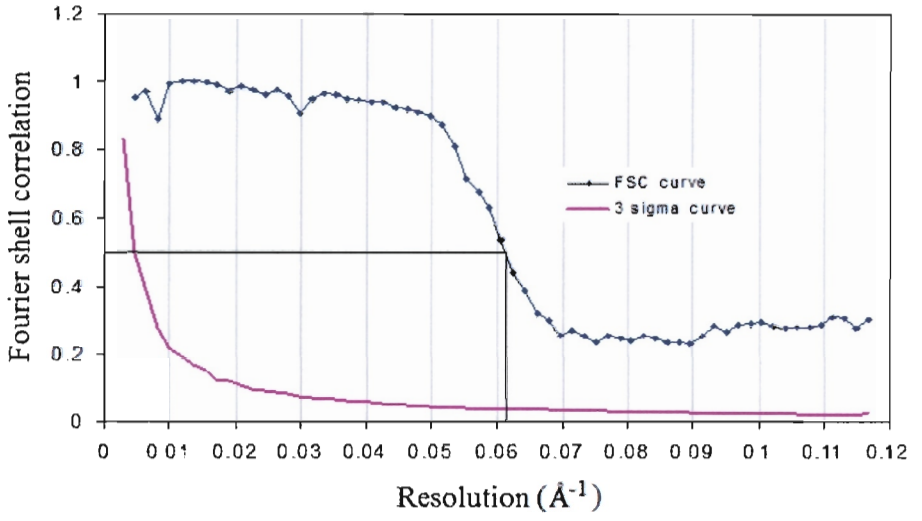


Figure 3.4.7: Fourier shell correlation for the first WT reconstruction (film data set) versus the second WT reconstruction (CCD data set) giving a value of (0.062Å^{-1}) at the fixed-value threshold of 0.5 which is equal to 16Å resolution.

3.4.2 G. sorghi mutant R87Q reconstruction

The negative stain structure of mutant R87Q converged to slightly different symmetry values than the WT, namely: $\Delta\varphi$ = -78.23, Δz =15.16. The two structures were also visibly different since at the same contour levels the narrow interfaces between subunits seen in the WT structure became much wider and the 'holes' became smaller. The prominent left handed ridges within the one-start helix were still visible and the protruding bulges on the inside of the structure were identical to the WT structure. An FSC performed between the WT and mutant correlated the two structures to 33Å resolution, indicating that at a finer level of detail the structures are inherently different.

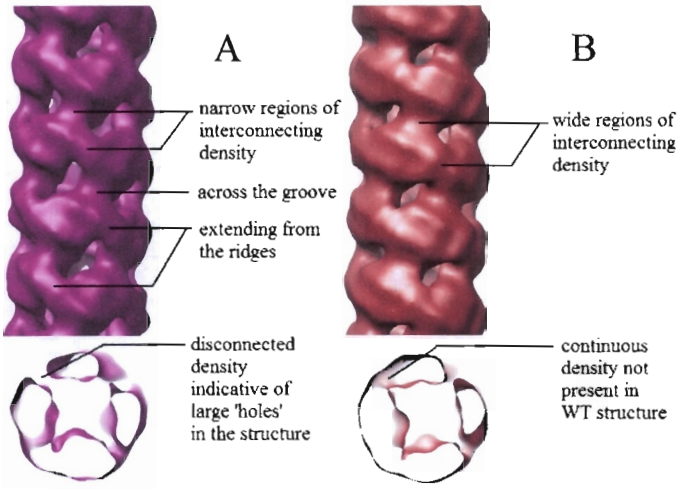


Figure 3.4.8: Highlighting the structural differences between the WT *G. sorghi* CHT reconstruction and that of the mutant R87Q. (A) The WT structure with narrow regions of interconnecting density extending laterally from the left-handed ridges and vertically across the groove. (B) Wide regions of density extending from the ridges, narrowing the holes across the groove.

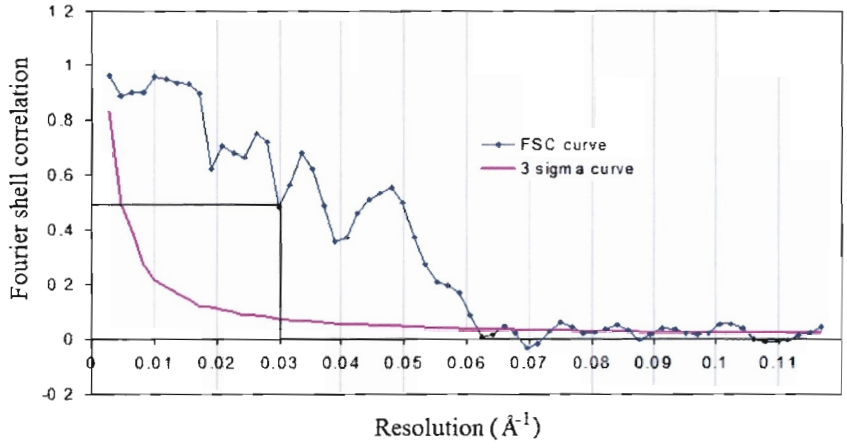


Figure 3.4.9: Fourier shell correlation for the second WT reconstruction versus the mutant R87Q reconstruction giving a value of (0.03Å^{-1}) at the fixed-value threshold of 0.5 which is equal to concurrence to 33Å resolution.

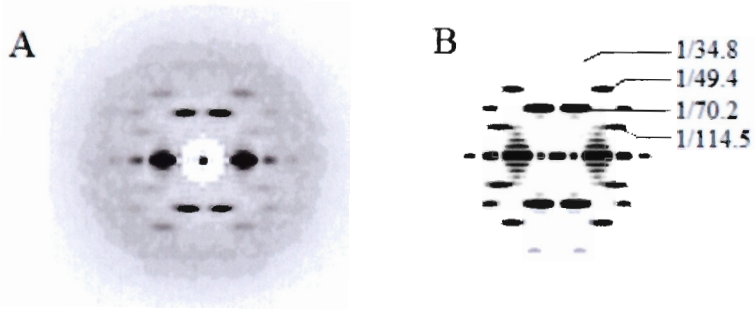


Figure 3.4.10: (A) Averaged power spectrum of rotated and aligned images of the *G. sorghi* mutant R87Q. (B) 2D power spectrum generated from the reconstruction volume which is identical to the average power spectrum of the images.

3.4.3 Cryo CHT reconstruction

Since most of the helical filaments were quite short, it was initially decided to use a small box size (128) to pick the particles. However, this proved to be an oversight since it was found that the short segments were mostly out-of-plane and the structure did not converge due to poor alignment. Images were re-picked with a larger box size (256) and only long helices were picked. Although this did not improve the ratio of out-of-plane to in-plane segments, it did improve the alignment and the structure converged within a few cycles.

It would be advisable in the future to work with cryo samples on carbon support since this would offer two strong advantages; the first being that it would eliminate most of the out-of-plane tilt of the helices which is of great advantage during the reconstruction process. Due to the nature of the negative staining process, a helix or elongated particle tends to lie in the plane of the EM grid with its long axis parallel to the carbon surface. However, in cryo-preparations, the particles are allowed to lie in orientations other than uniaxially depending on the thickness of the ice. Secondly, carbon-backing would allow for accurate CTF determination of the images.

For this set of cryo images, it was not possible to CTF correct because the contrast transfer function did not go past the first zero since ice generates poor Thon rings. As stated previously, it was intended to achieve an optimal balance between contrast and resolution at underfocus values of between -2.5μ and -3.1μ, but often the underfocus values of the images were far greater than desired. The process of minimal dose focusing of cryo samples with holey grids is challenging since the difference in height between the carbon film at the edge of a hole and the sample can be as much as 0.1 micron (pers. comm. Ed Egelman). Achieving the correct focus range for cryo samples is crucial when trying to recover high resolution. One advantage to working over holes however is that one achieves a better signal to noise ratio.

For the cryo data set, the sampling was originally given as 3.76Å/pixel. This however led to the layer line of Bessel order one to be found at 1/77Å, which did not agree with either the

negative stain or the shadowing for which the J1 layer line in both was found at 1/71Å. After inquiry it was discovered that the Philips CM120 had not been calibrated properly so the sampling at 33 000 was not accurately known. Since the structure did not converge when the sampling was set to 3.76Å/pixel, this was also evidence that the sampling was incorrect. It was thus decided that the J1 layer line should be constrained to lie at 1/71Å, since this would be in agreement with both the other data sets. When this was done, the sampling was calculated to be 3.42Å/pixel. The IHRSR iterations were restarted with this sampling value for the hsearch and himpose algorithms, which lead to a structure that converged within a few cycles. Unfortunately many images aligned to the tilted templates and were thus eliminated from the back projection because they were out of plane. Consequently only very few images of the entire stack were used for the reconstruction and thus the quality of the model as such was not of particularly good standard.

A reconstruction was performed where the images which were out-of-plane by up to 10° were incorporated in the back-projection. This allowed for more images to be used during reconstruction, and increased the amount of density visible on the interior surface of the structure. The non-symmetrized volume had extremely low signal to noise ratio when compared to the non-symmetrized volume of the negative stain (which was almost identical to the structure with symmetry imposed). A low signal to noise ratio is to be expected with cryo images and the fact that so few images were used in the reconstruction compounded this problem. The structure converged with the prominent left-handed 1-start helix and as with the negative stain structure the ridges within these helices were notably left-handed, whereas in the shadowing the 'tubes' within the one-start helices were shown to be right-handed three start helices.

Parameters of symmetry convergence were $\Delta\varphi$ = -76.2, Δz = 14.8. This was of course dependent on what value was given for the sampling – these parameters were those which resulted from constraining the sampling to be 3.4Å/pixel. The reconstruction where out-of-plane tilts (up to 10°) were included during the back-projection converged to the same parameters as the first reconstruction, however in comparison, more prominent bulges of density were seen on the inside surface of the structure. In a Fourier shell correlation of two

reconstructions from different starting structures which had excluded out-of-plane images, the resolution shells correlated to 27Å resolution. This value was only achieved by disregarding the inconsistent low correlation value of 0.372 at a resolution of 46.5Å. The x-shift histogram in 3.4.14 strongly suggests a problem with the alignment of the images such that the images are more than 3 pixels off centre. Unfortunately time did not allow for further investigation of this problem.

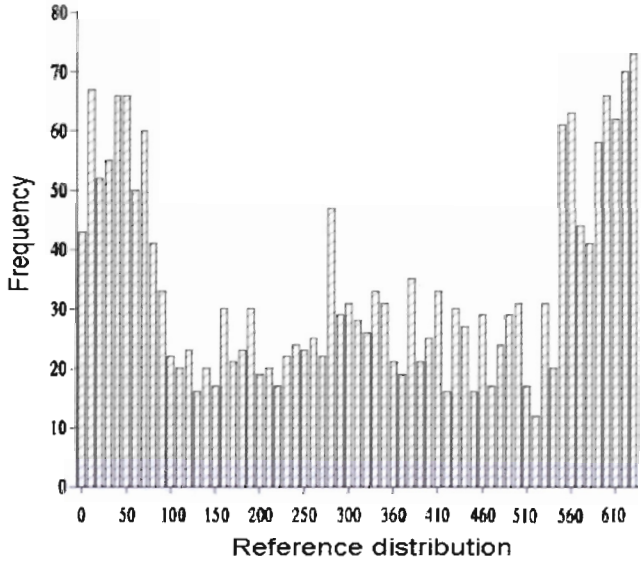


Figure 3.4.11: Reference distribution of images aligning to projection templates, showing preferential alignment to tilted reference images. Tilts were altered in 5° increments from 75° through to 105° over azimuthal rotations for 90 templates each. References between 270 and 360 have a tilt value of 90° which is un-tilted according to the SPIDER coordinate system.

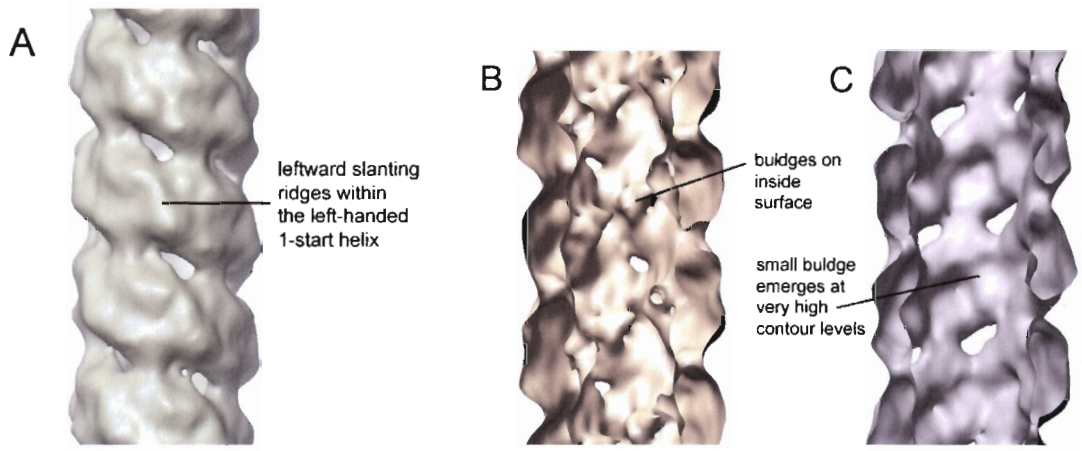


Figure 3.4.12: (A) *G. sorghi* cryo reconstruction where out-of-plane tilts had been excluded from back projection. Left-handed ridges within the one-start helices are seen in the symmetrized volume but are not clear in the reconstructed volume where symmetry has not been imposed. (B) Longitudinal cross section of the cryo reconstruction where out-of-plane tilts (up to 10°) were included during the back-projection. The exterior surface appeared the same, but more prominent bulges of density were seen on the inside surface of the structure when compared to the cross section of the first structure (shown in C).

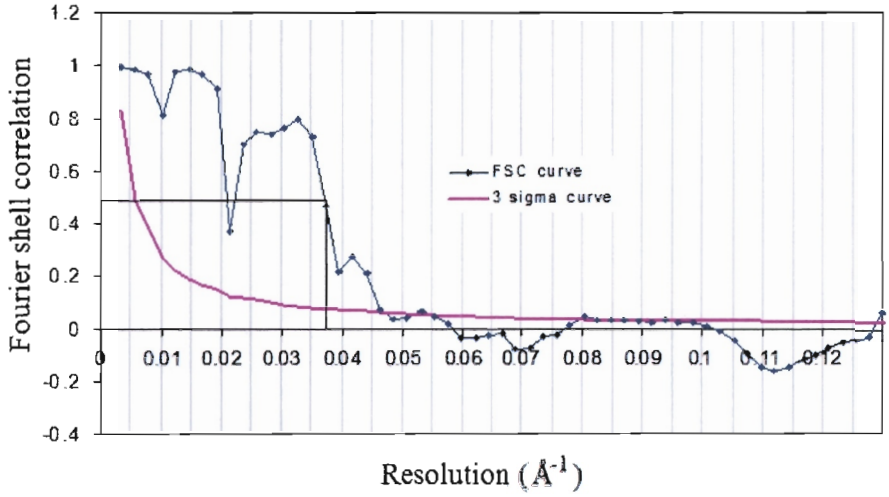


Figure 3.4.13: Fourier shell correlation for the cryo *G. sorghi* reconstruction of the cylinder start versus the *B. pumilus* starting structure giving a value of (0.037Å^{-1}) at the fixed-value threshold of 0.5 which corresponds to 27Å resolution.

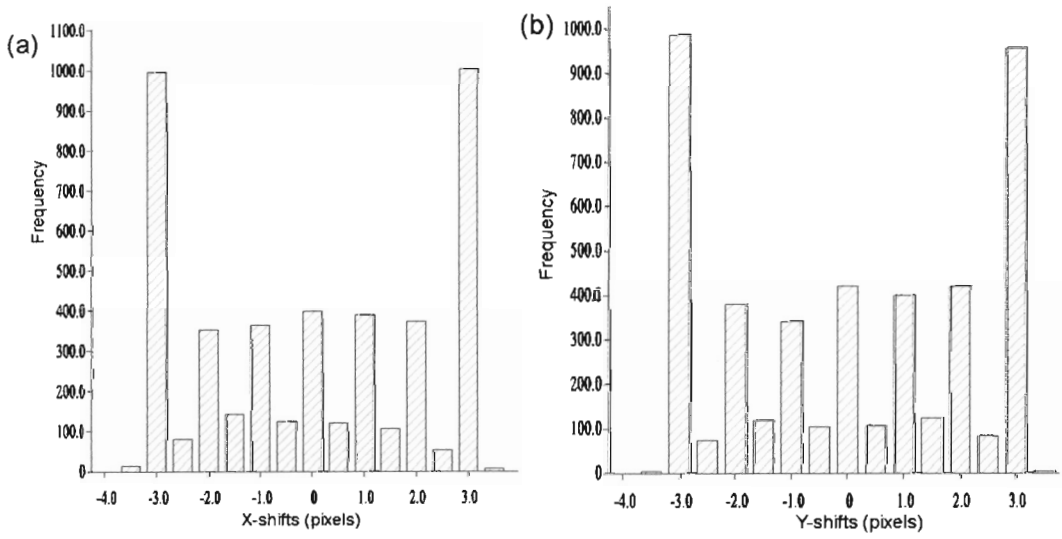


Figure 3.4.14: (a) X-shift and (b) y-shift distributions of the cryo reconstruction exhibiting multiple modes around integer values. As with other data sets, most of the images aligned with x-shifts and y-shifts at the limit of the search range which possibly indicates that the images did not find optimal alignment within the search range given.

3.4.4 Integer value multiple-modes for x-shift and y-shift distributions

Interestingly, during most of the reconstructions, the x-shift and y-shift distributions tended to exhibit an anomalous characteristic where both exhibited multiple modes around integer values. Other means of pre-aligning images in the x-direction and also not pre-aligning images were tried, but the results were always the same indicating that the result was not artefactual, but it could rather be attributed to a factor inherent in the alignment procedure. The main factor to be implicated in creating such a distribution would be the interpolation step during the 2D alignment of the images (pers comm. Ed Egelman). In most image processing applications, preference is given to rapid low-order interpolation schemes which have a fairly strong low-pass filtering effect on the images (Joyeux and Penczek, 2002). It is known that in the case of rotation, higher cross-correlation scores result when no interpolation is required i.e. at $\varphi = m90°$ where m = {0,1,2,3} since these rotations do not corrupt the image. Any intermediate angle will result in low-pass filtration of the image with the strongest effect for $\varphi = 45° + m90$.

Similarly, the characteristics of the associated Fourier-space filter depends on the amount of shift applied. During translation, a shift by k pixels does not corrupt the image but a shift by $k + x$, $x \in (0,1)$ does. The low-pass filtering effect is strongest for $x = 0.5$ (Joyeux and Penczek, 2002). Admittedly, since the AP NQ routine performs the alignment on the original images each time, interpolation errors are not cumulative and some may say that they are in fact negligible. This is because in the method of resampling to polar coordinates employed by the AP NQ procedure, the rotation angle is calculated first and the translation is found by applying the same procedure with respect to the origin of the system of coordinates to the translation considered. This reverse procedure may however be the reason for the increased corruption effect of interpolation during translation of images. The high axial ratio of images with helical filaments will augment the problem since the low-frequency overlap of the Fourier sections is considerable and thus a low-pass filtering effect will be more severe than for non-elongated objects.

3.4.5 B. pumilus CynD reconstruction

Two reconstructions were performed on the *B. pumilus* data set using different starting models. The first starting model was a cylinder filtered to 50Å and the second was a converged *G. sorghi* volume. The cylinder starting model was given initial search parameters of $\Delta\varphi$ = -75.0, Δz = 13.5. After about eleven iterations, the symmetry parameters started to converge around the values $\Delta\varphi$ = -76.64, Δz = 15.87 at which point they stabilized. The *G. sorghi* starting structure was given initial search parameters of $\Delta\varphi$ = -78, Δz = 15. After about eight iterations, the symmetry parameters started to converge around the values $\Delta\varphi$ = -76.63, Δz = 15.87.

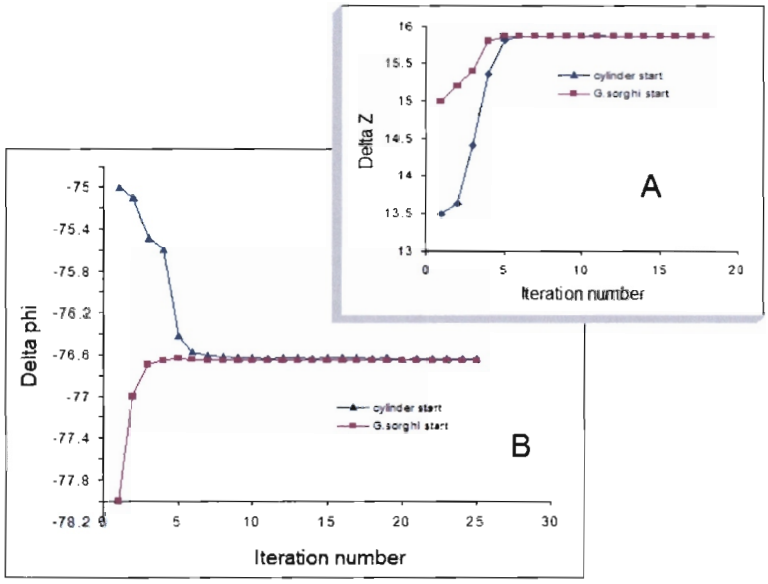


Figure 3.4.15: Convergence of symmetry parameters $\Delta\varphi$ and Δz from different starting points and with different models. The cylinder starting model initial search parameters $\Delta\varphi$ = -75.0, Δz = 13.5, and the *G. sorghi* starting structure was given initial search parameters of $\Delta\varphi$ = -78, Δz = 15. After only a few iterations, the symmetry parameters start to converge around the values $\Delta\varphi$ = -76.6, Δz = 15.87.

The average power spectrum of the rotationally aligned images is identical to the power spectrum generated from the reconstructed volume (figure 3.4.16), indicating that the structure is correct. The reference distribution was statistically even and the in-plane rotation angles were centred on 0° and 180°. X-shift and y-shift distributions were centred

on zero, but also showed preferences for integer shift values. The frequency of correlation coefficients showed a bimodal distribution which was further investigated.

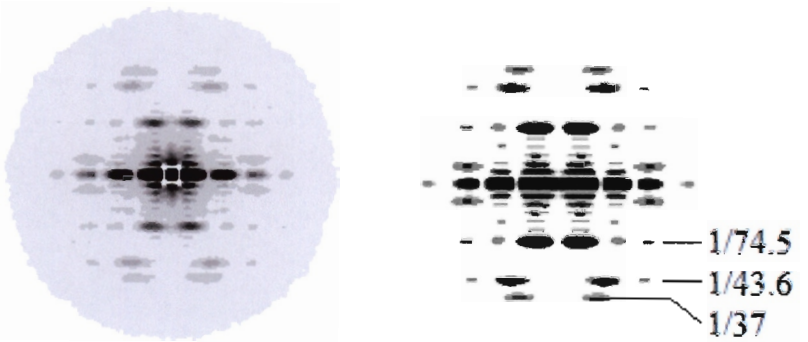


Figure 3.4.16: (A) Averaged power spectrum of rotated and aligned images of the *B. pumilus* CynD (B) 2D power spectrum generated from the reconstruction volume of *B. pumilus* which is identical to the average power spectrum of the images.

3.4.6 Bimodal distribution of correlation coefficients

During both *B. pumilus* reconstructions (different starting models using the same data set), a bimodal distribution of the correlation coefficients (CC) was noticed. By visual inspection of the two sets of images in WEB there was nothing notably different to be seen. After plotting CC versus image number it was evident that there was a distinct relationship between the two. Images with low correlation coefficients were generally from the first 1500 images of the image stack and those with high correlations were generally from the latter part of the stack.

This could arise due to a change of z-height between the first half and the latter half of the image set, leading to a change in magnification of the images being taken at an incorrect z-height. Visual inspection of the original images showed varying stain depths which could also be a cause. Since the relatedness between image number and correlation coefficient does not actually provide evidence as to what factor caused the bimodal distribution, the most effective test for this would be to perform multi-variate statistical analysis (MSA) on the image set to see if there is a principal component relating to differences in size or another unrelated component.

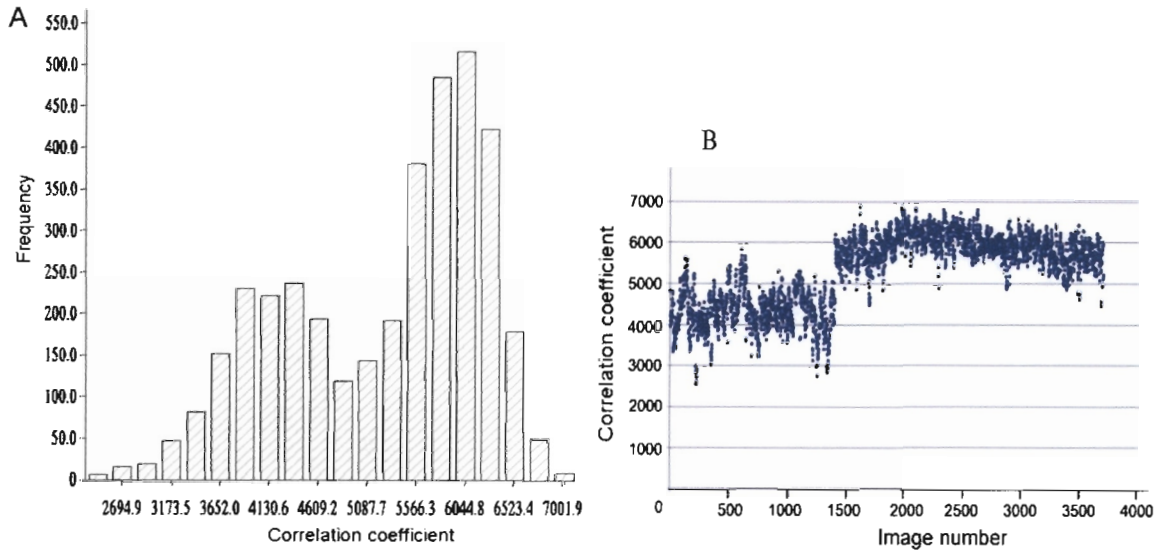


Figure 3.4.17: **(A)** Correlation coefficient vs number of images (frequency) showing a bimodal distribution. **(B)** Image number vs correlation coefficient presenting evidence to suggest that images 1-1500 from the stack are different in some way to the latter part of the stack since on average they have lower correlation scores.

The image stack was separated into two sets based on the correlation coefficient of the image with a cut-off at 5000 and separate reconstructions were performed on both sets of images. A visible difference between the 3D volumes was that the subunits appeared smoother in the reconstruction with high CC images. The reconstruction from the low CC images showed apparent higher resolution features with two small protruding bulges from each subunit. There was also a slight difference in the convergence symmetry parameters. High CC images yielded $\Delta\varphi$ = -76.67, Δz = 15.87 whereas low CC images yielded $\Delta\varphi$ = -76.2, Δz = 15.91. The higher value in the rise of the low CC images 3D volume suggests that these images may possibly have been taken at a slightly higher magnification.

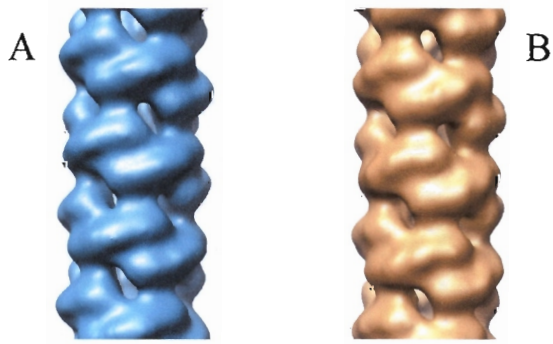


Figure 3.4.18: (A) Reconstruction of *B. pumilus* CynD using the set of images that had high correlation coefficients. (B) Reconstruction of *B. pumilus* CynD using the set of images that had low correlation coefficients.

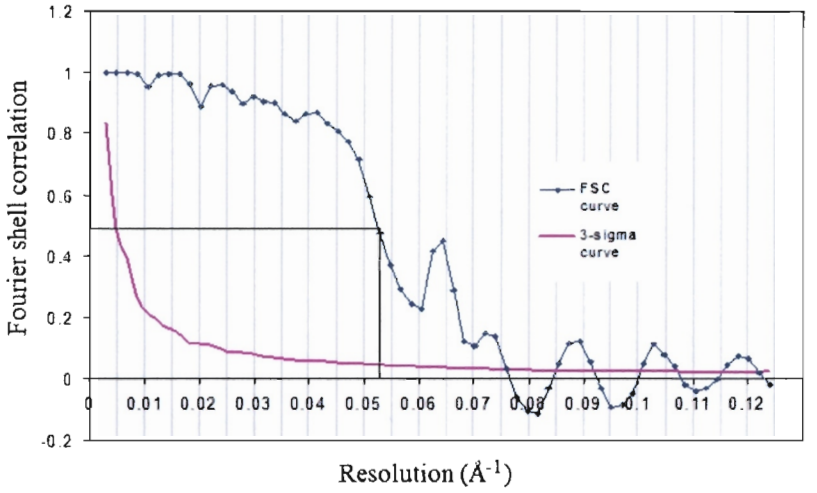


Figure 3.4.19: Fourier shell correlation between two *B. pumilus* volumes reconstructed from different starting structures. A fixed-value resolution criteria of '0.5' gives a spatial frequency of (0.053Å^{-1}) which is equal to a resolution value of 19Å. The 3σ curve crosses the FSC at (0.076Å^{-1}) which renders a resolution of 13Å, but this is not considered a realistic value since the grain size of the uranyl acetate negative stain generally limits resolution to around 20Å.

It is not known why there are oscillations in the FSC curve since all measures were taken to rid the structures of any sharp edges and their axial mask functions. Some would not consider this a verifiable measure of the resolution since the FSC was performed between structures arising from the same data set. However, others believe (Yang et al., 2003), (Egelman, 2000), that reconstruction of one data set with two independent starting models initiated with different starting symmetries result in two truly independent models. Since the *B. pumilus* starting models did not correlate to a greater resolution than 50Å and the structures converged from different starting symmetries, according to the above mentioned principle, the two volumes are independent and their FSC is valid and unbiased. The two structures generated from the *G. sorghi* data set are truly independent reconstructions since they arise from different data sets.

Yang *et al.* have shown that dividing an image stack in half and aligning them separately to the same starting model is not a valid way of producing two truly independent reconstructions since even aligned pure noise will start to resemble the starting structure after several iterations (Yang et al., 2003). They showed that an FSC produced from alignment of pure noise images to a low-pass filtered model of 12Å in fact gave a calculated resolution of ~10Å at the 'conservative' fixed value threshold of 0.5. This extension of

resolution was attributed to the fact that neighbouring shells in the Fourier transform are highly correlated when the thickness of the shell is less than $1/D$ where D is the diameter of the object. This of course is a problem with helices where the high axial ratio will affect the thickness of the shells in the Fourier transform.

3.4.7 Quality and correctness of the 3D volumes

Evaluating the correctness of a reconstruction, according to the data present in the images can be tested in several ways. Given the statistics of the data analysis, the reconstructed volumes in section 3.4, are correct in so far as the following criteria are concerned. Each structure converges rapidly on a stable solution of continuous, connected density given any starting model. The symmetry values converge regardless of the initial search parameters given. The power spectra of the reconstructed volumes are identical to the rotationally aligned, averaged power spectrum of their corresponding image stacks and the periodicities and symmetries of the 3D reconstructions are compatible with the observed diffraction patterns of the images. The reference distributions are statistically random and no particular templates are preferred. X-shift distributions are at a maximum around zero and have a Gaussian distribution.

Table 4.1: Summary of statistics and results from the various reconstructions

Reconstruction	no. of images in stack	no. of images used	search radius (Å)	imposition radius (Å)	Δφ	Δz
Pumilus	3701	3701	80	100	-76.64	15.87
First WT sorghi	3358	3088	80	100	-76.96	14.99
Second WT sorghi	8087	7767	80	100	-77.00	14.87
Mutant R87Q	4820	4820	80	100	-78.23	15.16
Cryo sorghi	4527	2221	85	85	-76.22	14.80

SECTION 4 – DISCUSSION

4.1 Docking of the 3D reconstructions

Through modelling of the polypeptide backbone of *B. pumilus* and *G. sorghi* using the structural templates of the existing atomic structures, it was possible to fit the predicted secondary structure elements into the molecular envelope of the low resolution model. It was established earlier that the A surface of the dimer will most likely be preserved in all structures, thus the dimeric models were used for docking the reconstructions. Since the insertions of the models could not be accurately modelled, they are represented as loops the general area of extension and are indicated as asterisks in some of the images. From the docking some useful parameters are determined, such as relative locations of the centroids of the dimers and the optimized relative tilt of the dimers. Docking also enables the interpretation of the biological interactions and allows the rationalization of various biochemical data.

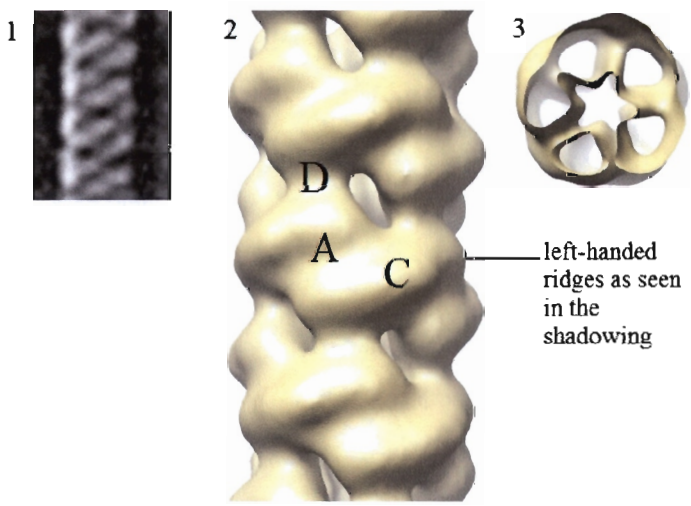


Figure 4.1.1: (1) The reference free average of the shadowed material placed adjacent to the *B. pumilus* reconstruction, highlighting the similarity between the surface features. (2) The *B. pumilus* reconstruction with labelled surfaces according to the *P. stutzeri* structure (shown figure 1.1.10). (3) End on view of the *B. pumilus* structure showing similar bulges on the inner surface as the *P. stutzeri* structure (figure 1.1.10).

When visually inspected, the *B. pumilus* structure looks very similar to that of the terminating spiral of *P. stutzeri*. Right-handed ridges within the one-start left-handed helix are clearly visible enabling the identification of surfaces A, C and D *a priori*. These ridges are of identical angle and dimension as those seen in the shadowed images (figure 4.1.1), providing a form of validation. A longitudinal cross-section does not clearly show the bulges on the inside of the helix because of its narrow diameter, but in a lateral section or from a top view of the reconstruction (figure 4.1.1 (3)) the bulges analogous to those in the *P. stutzeri* structure are visible.

A clear dyad axis could be seen in the reconstructed density which, in the *P. stutzeri* structure coincided with the dyad axis of the dimer. Automated docking with SITUS produced a series of plausible fits, some of which had the carboxy-terminus directed towards the inside of the structure; ambiguous docking occurred where some dimers were fitted with NS13 of the C-terminus facing outward, while others were docked with this motif facing inwards and the tilt angle of the model reversed (figure 4.1.2A). Upon closer inspection of the docking, it was clear that the A-surface should correspond to the global dyad axis at the interface between the ridges. Several lines of evidence supported this fact.

The NS13 outward-facing dimers displayed large regions of unoccupied density which could not be accounted for by the insertions (figure 4.1.2E), whereas a lateral view of the NS13 inward-facing dimers demonstrated that the 'wings' of the dimer fitted the shape of the density and that the C-terminus emerged from the density at the bulge on the inside of the structure (figure 4.1.2B). The NS13 inward-facing docking may also be more easily rationalized by the presence of insertions which would hypothetically extend into the region of unoccupied density. Other evidence suggests that the carboxy terminus fills this interior bulge since docking the dimers in this orientation provides interpretation of biochemical evidence for various inactivations resulting from mutations, as is discussed later in this section.

A similar problem existed for docking of the *G. sorghi* dimer model since some were docked with the carboxy terminus facing outwards and others were docked with NS13 facing inwards, albeit with the same tilt angle. Docking with the C-terminus facing inwards could be rationalized with the same lines of evidence as put forward for the *B. pumilus* structure. Docking with NS13 facing outwards would lead to insertions extending into regions where there is no density leaving unfilled density unaccounted for, whereas dimers with carboxy-termini facing inwards allowed regions of unoccupied density to be filled by their insertions (figure 4.1.2E).

Also, when the dimer is docked in this manner, the NS13 β-sheet of the C-terminal extension fits into the bulge on the inside surface of the helix as seen in the top view of the density (figure 4.1.3b) and corresponds to the shape of the bulge as seen by the longitudinal section (figure 4.1.3c), and the A-surface corresponds to the global dyad axis at the interface between the ridges (figure 4.1.3d). When dimers are docked with outward-facing C-termini, the residues known to lead to inactivation when mutated would be situated at an edge of the density indicating that they would not form part of an interface, making it difficult to rationalize the available biochemical evidence.

Locating the C-terminal extension (NS13) of the polypeptides with absolute certainty should be possible in the future through the visualization of active truncated enzymes to see if the bulge on the interior surface is still as prominent, which would provide conclusive proof that the internal density is that which accommodates the carboxy-terminal extensions of the helices.

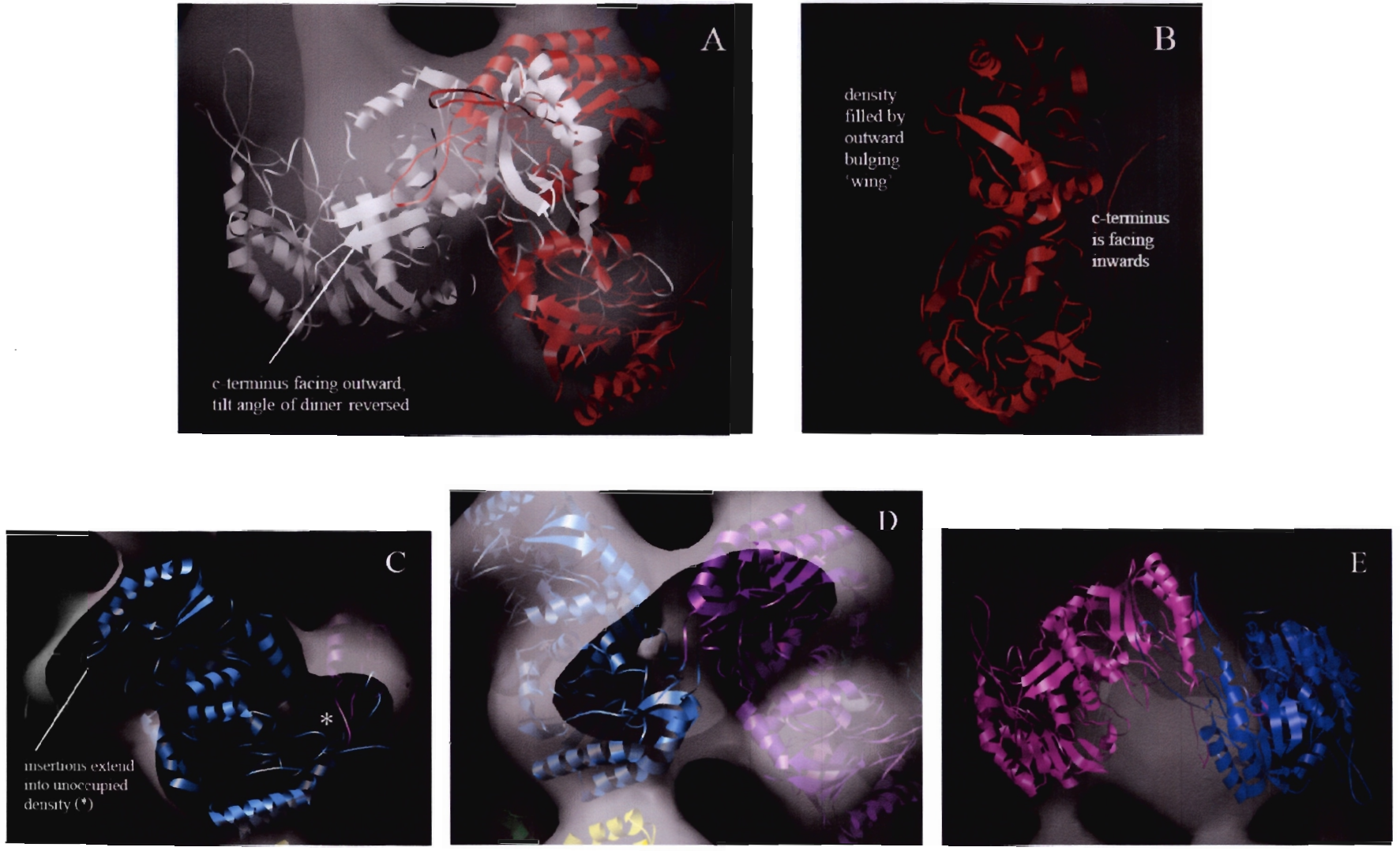


Figure 4.1.2 Docking of the *B. pumilus* density. (A) Ambiguously docked dimer models - NS13 of the white dimer is facing outwards whereas the red dimer has its C-terminal motif facing inwards. (B) Lateral view of the red model demonstrating that the 'wings' of the dimer fit the shape of the *B. pumilus* density and NS13 emerges at the point where the internal bulge exists. (C) Dimer model docked with NS13 facing inwards, illustrating that the regions of insertions as indicated would hypothetically extend into the region of unoccupied density. (D) A view of the C surface interface. (E) Dimer model docked with NS13 facing outwards displaying large regions of unoccupied density which cannot be accounted for by the extension of insertions.

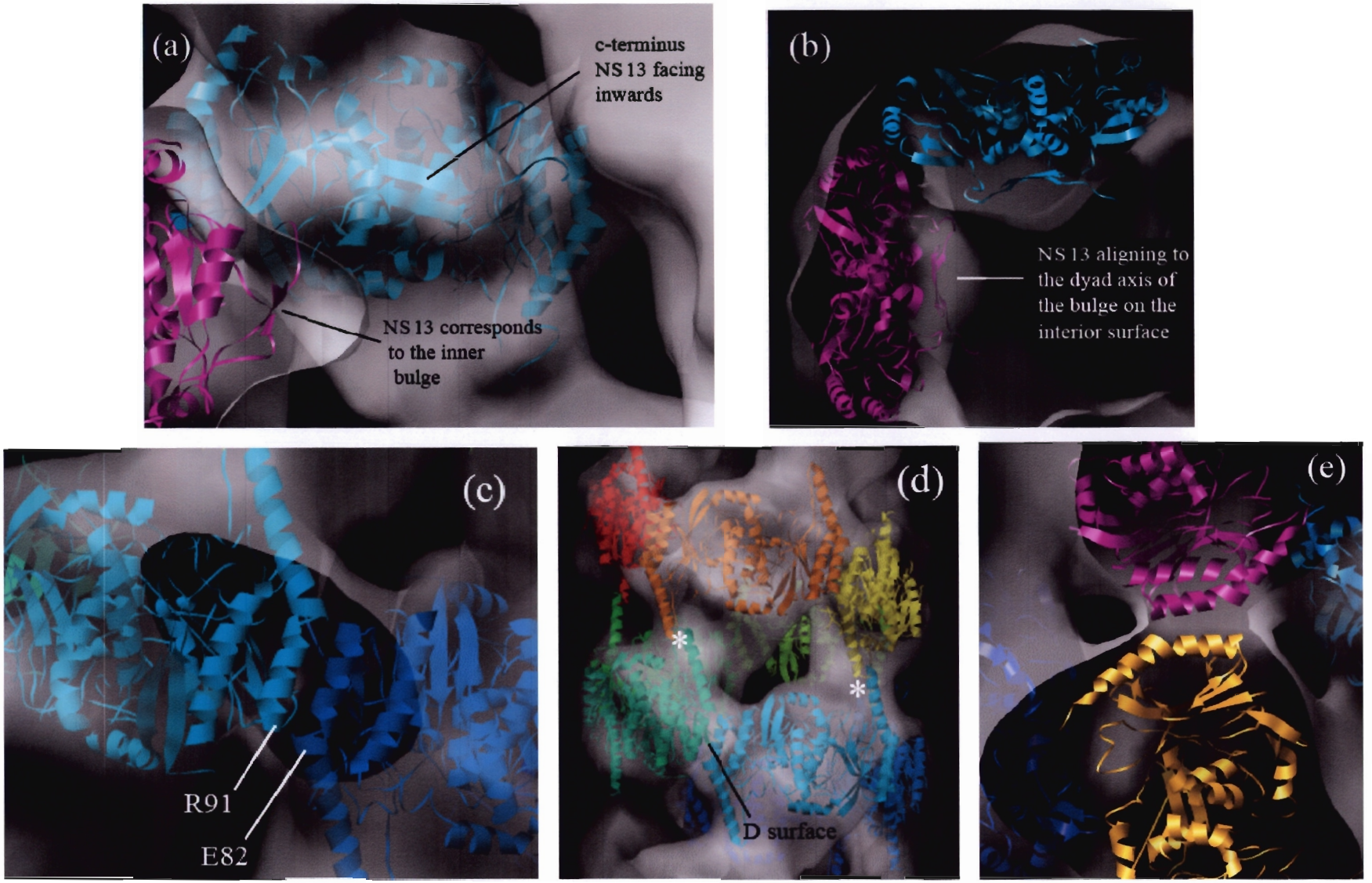


Figure 4.1.3 (a) *G. sorghi* CHT docked with the C-terminus facing inwards; the NH13 β-sheets of the C-termini correspond to the bulges on the inside of the helix. (b) Top view of the density. (c) A view of the D surface interface indicating the positions of the residues leading to inactivation. (d) Dimer model docked with NH13 facing inwards showing that regions of unoccupied density will be filled by the insertions (designated by *). The D surface as enlarged in (c) is also indicated. (e) Corresponding D surface of *B. pumilus* CHT, showing similar alignment of the NH2 helix. Residues forming the salt bridge across this interface have not yet been identified.

4.2 Nitrilase mutants and their corresponding activities

A series of mutants were constructed in order to destroy what was thought to be the C-surface on the basis of sequence alignment. From present experience in our lab, it seems that inactive mutants are difficult, if not impossible, to purify. The inactive *G. sorghi* mutants E82V and R91Q are proving to be elusive during the purification procedure - it is hypothesized that they form aggregates in the host cells and thus are localized in the inclusion bodies, thus making them difficult to purify. It may take some time before these structures are visualized.

From what has been seen so far, those mutants which are active, readily form extended helical fibres much like the WT. The reconstruction of mutant R87Q shows that the structure converges to helical symmetry parameters almost identical to the WT structure and when inspected visually, the structures cannot be considered as being greatly dissimilar. The FSC between the WT and mutant structures however only correlate to 33Å, demonstrating that there are differences in the structure beyond this resolution.

It is thus possible to speculate that all active mutants retain the same structure given that those that have been visualized by negative stain EM form extended fibre. It is feasible that they are less stable than the WT since the presence of imidazole was found to be obligatory for fibre formation. The structure of a mutant in which the alteration would affect stability or destroy interactions such as ion pairs at a surface required for oligomerisation would be expected not to form fibres since the exact orientations required for extension of the helix would no longer be present thus preventing fibre or even spiral formation.

Table 4.1: Mutant data indicating which surface the mutation is believed to affect and whether the resultant recombinant protein is active (refer to figure 1.1.11).

Mutant	Surface	Location	Activity
B. pumilus			
1. Delta 303	A	Vgtg → stop	Full activity
2. Delta 293	A	Matg → stop	Partial activity
3. Delta 279	A	Ytat → stop	Inactive
4. Delta 90	D	EAAKRNE → AAARKNK	Full activity
P. stutzeri			
5. Delta 310	A	Sagt → stop	Inactive
6. Delta 276	A	Kaaa → stop	Inactive
G. sorghi			
7. Delta 82	D	E82V	Inactive
8. Delta 84	D	R84Q	Full activity
9. Delta 85	D	R85Q	Full activity
10. Delta 87	D	R87Q	Full activity
11. Delta 91	D	R91Q	Inactive
12. Delta 92	D	D92Q	Full activity
13. Delta 217	A	Y217D	Inactive
14. Delta 217	A	Y217E	Inactive

4.3 Interpretation of structures from docking results

It has been shown that modification of some residues which are distant from the active site does not influence the activity of the enzyme, for example, R87Q and D92Q in *G. sorghi* and E90A *B. pumilus*. Mutations which do however cause inactivation may therefore be postulated to affect the interactions between subunits and oligomerisation, consequently affecting activity of the enzyme.

C surface

The interface leading to spiral formation in *P. stutzeri* was previously called the C surface (Sewell et al., 2003; Sewell et al., 2005). The residues speculated to be involved in this interaction are the insertions at the amino terminal end of NH2 (12-13 residues) and at the bend between NS9 and NS10 (13-15 residues). Both of these insertions are located in such a way that they could plausibly fit empty regions in the reconstructed density. According

to the sequence alignment (figure 1.1.11), in *B pumilus* these residues at the N-terminus of NH2 are residues 55-73 and those between NS9 and NS10 are residues 217-234. For the *G. sorghi* sequence, these insertions are found at residues 52-71 and 237-254 respectively. These insertions have thus been implicated in the formation of the spiral since the crystallographically determined proteins which form dimers/tetramers lack these particular residues.

Since there are almost five dimers per turn in the *B. pumilus* helix, dimer number six interacts across the groove with first dimer of the previous turn. This interaction is known as the D surface which lies on pseudo-dyad axis found directly on the opposite side of the spiral to the C-surface i.e. D surface between dimers 2 and 7 lies directly opposite the C surface of dimers 4 and 5.

D-surface

Residues E90 and R94 were thought to be involved in forming an ion pair across the D surface but the multiple mutation in *B. pumilus* of 90EAAKRNE→AAARKNK (residues 90-91 mutated to the corresponding *P. stutzeri* sequence) did not have any effect on the activity of the enzyme. It is not known yet if this mutant forms fibres, however, it is speculated that disruption of this surface will also disrupt formation of the short spirals since the D-surface interaction is found in both structures, thus damage to the salt bridge across this interface is speculated to lead to inactivation (as it does in *G. sorghi* CHT).

Therefore, residues forming this interface have yet to be identified but it is known that the interaction should be located in a position which would correspond to a dyad axis across the NH2 helix. Residues K94 and R95 were swapped, and thus there was no alteration of charge in that area so those residues could still be implicated in charge-charge interactions. It is plausible that residue E80 prior to residue S83 (position of the bend in NH2 according to homology with NitFhit) interacts with either K88, K94 or R95 but docking would suggest that the ion pair forms at the C-terminus of NH2 after the bend in the helix. Nevertheless, residues E80 and K88 are conserved between the *B. pumilus* and

P. stutzeri sequences, supporting this hypothesis. Mutagenesis experiments would need to be conducted to confirm this.

C-terminus

Residues involved in intersubunit contacts may be identified through site-directed mutagenesis. These mutants provide evidence as to whether full assembly of the complex is necessary for activity. Since the C-terminus is the most variable region, and for all tetrameric or dimeric nitrilases the extension is missing, it may indicate that this part of the polypeptide may influence the degree of oligomerisation of nitrilases. However, recent mutation evidence suggests that, if oligomerization is required for activity, that the C-terminus is not involved in oligomerization since *B. pumilus* CynD with up to 52 residues (mutations leading to removal of the C-terminus of the *B. pumilus* enzyme may extend up to Δ293 before the enzyme starts becoming less active, Δ279 renders an enzyme which is completely inactive) of the extension removed retains its activity, and *P. stutzeri* is slightly more susceptible to inactivation by removal of C-terminal residues, but the effect is similar. Since these nitrilases have not yet been visualized, it is not known whether or not they form spirals.

If they do not, it will indicate that the C-terminal extension is required for oligomerization but that oligomerization is not required for activity. Of course, the stability of the protein with the mutation must be taken into account since a mutation resulting in loss of activity may suggest an allosteric effect on the conformation of the active site and not necessarily on the quaternary structure of the enzyme. It was previously thought that this region is imperative for formation of the A-surface, however recent results have shown that truncation of these residues (to a certain extent) does not lead to inactivation of the enzyme. The fact that the crystallized nitrilases have 'shortened' C-termini but still form dimers suggests that the extension may not be important for dimer initiation and retention.

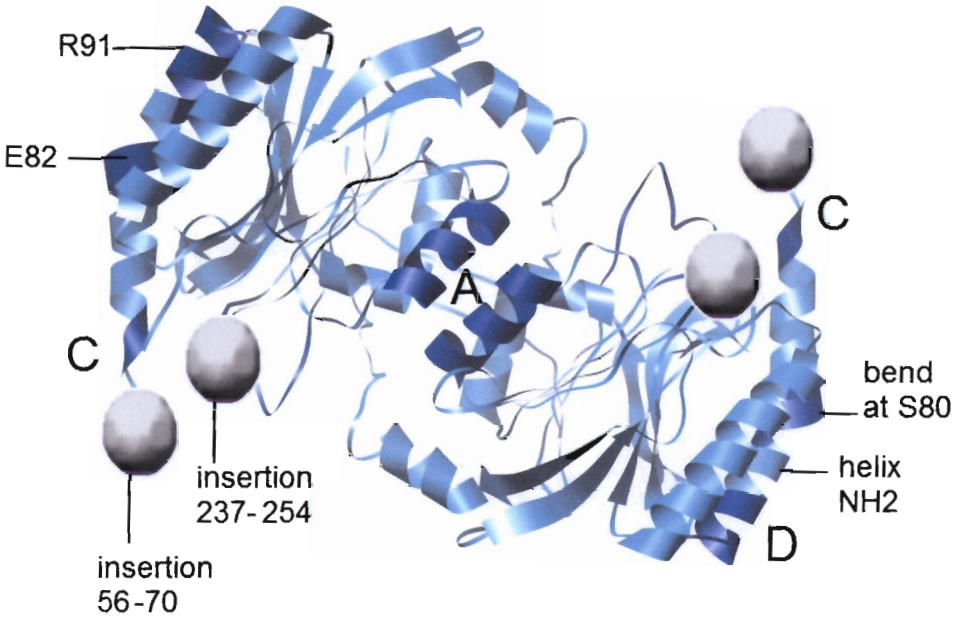


Figure 4.3.1: Model of a *G. sorghi* dimer based on secondary structure prediction. The main insertions, 56-70 and 237-254 are not modeled since there is no template provided from the known PDB structures; they are shown as spheres. The main surfaces as described in the text are indicated. The bend in α-helix NH2 at S80 is labeled, as well as the residues E82 and R91 implicated in formation of a salt-bridge in order to construct the D surface.

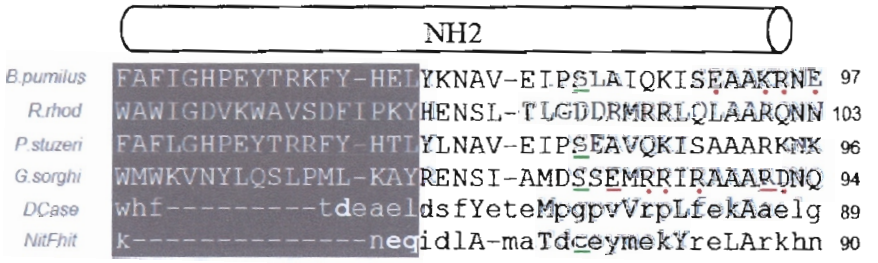


Figure 4.3.2: Sequence alignment of the NH2 helix. Underlined in red are the *G sorghi* residues in which a point mutation leads to inactivation thus indicating the presence of ion-pair formed by the two amino acids. Underlined in green is the conserved serine residue at which the bend in the helix occurs (modeled according to the cysteine residue in NitFhit), Residues with red dots beneath are those charged residues which did not alter enzymatic activity when altered. In *B. pumilus*, residues K94 and R95 were swapped, and thus there was no alteration of charge, so these residues could still be implicated in charge-charge interactions.

4.4 Mechanism of spiral termination

The quaternary structures of *G. sorghi* cyanide hydratase and *B. pumilus* cyanide dihydratase (at pH 5.4) form perfectly regular spirals of indefinite length with closely related helical symmetries. From the evidence put forward, it is possible to speculate that all multimeric nitrilases are left-handed helical structures in which the extent and strength of interactions breaking helical symmetry dictate the length of the spiral. In the extended fibre, a global dyad axis is found but in the terminal spiral the symmetry operator is reduced to a pseudo-dyad since termination leads to disruption of the dyadic symmetry of these interfaces.

In the *P. stutzeri* structure, docking of the individual dimers confirmed that there was a decrease of 10Å when comparing the diameter of the centre of the spiral to the ends of the spiral. The atomic model produced for the *P. stutzeri* sequence emerges from the density envelope at residue 291 (according to the NitFhit numbering). In this structure, there is no clearly defined density to accommodate the remaining 38 C-terminal amino acids which is theorized to be due to flexibility and heterogeneity in the region resulting in an average lower density.

Sewell *et al.* have already suggested two possible factors that could influence the length of the nitrilase spiral (Sewell et al., 2003). It was noted in the *P. stutzeri* structure that the diameter of the terminating spiral decreases from 120Å in the centre to 110Å at the end. The other noticeable factor was the increase in tilt of the terminal dimers inward towards the axis of the spiral by 12 degrees. The narrowing of the spiral at its ends and the inwards tilt of the terminal dimers is thought to prevent the addition of more dimers. This mechanism can be confirmed by the fact that for the extended fibre, there is no deviation from helical symmetry. The breach of helical symmetry is thought to be a result of accumulation of interactions across the groove of the spiral, consequently altering the geometry at the termini of the growing spiral which prohibits the addition of further subunits.

In the terminating spiral, an asymmetric interface is created where the terminal dimer tilts across in order to create the E surface (Sewell et al., 2005), which is suggested to involve residues 266EID268 in the structural element NS12 (in *G. sorghi* the corresponding sequence is DID). This interface is formed with the C terminus of NH2 of the centroid (possibly 92RKNK96) which would be involved in creating the D surface if another dimer was to be added to the end of the helix. Thus, if the E surface forms, it removes the availability of the D surface and favourable energetics are required to overcome the energy well created by this interaction, which could be facilitated by a process such as concomitant burial of hydrophobic residues upon extension of the helix. It is possible that for the *B. pumilus* structure, the energy well created by the E surface lies near an equilibrium which is easily perturbed by changes in the local environment of the protein.

The native *B. pumilus* structure was shown to consist of 18 subunits (at pH8), whereas the recombinant structure was thought to consist of 16 subunits (recombinant was reconstructed at pH6). At that time, it was not known that the A-surface is preserved, and consequently it was not identified that the structure is composed of dimers. A minimum number of five dimers are required to complete a turn and thereafter, dimers are successively added to the ends, resulting in structures composed of 7, 9, 11 dimers and so on. In terms of monomeric subunits, spiral structures would thus extend in increments of four, i.e. 14, 18, 22 subunits. A structure therefore, with 16 subunits thus seems unlikely.

Upon docking of the recombinant *B. pumilus* structure, two large regions of vacant density were observed adjacent to the terminal dimers. These regions of density were presumably mistaken for additional monomers. This does consequently raise the question as to the origin of these regions of density. One likely possibility is that a partial occupancy is observed of a structure that exists in dual states between a 14 and an 18 subunit structure. The other possibility is that the terminal dimer is flexible in its binding and tilts across the groove in order to close the structure by binding creation of the E surface.

Evidence suggests that the latter is true given that *B. pumilus* CynD does not always form the terminating spirals, but sometimes forms elongated spirals of twice the length of the

terminating spirals. Upon docking of the terminating spiral it can be seen that the optimal docking for the 'terminal dimer' (prior to the region of partial occupancy) is not tilted over in order to create the E surface as in the *P. stutzeri* structure which also indicates that there must be additional dimers at the terminal ends in order to close the spiral. This evidence suggests that the mechanism of spiral termination in *B. pumilus* is not as pronounced and precise as with the *P. stutzeri* structure. The inclination for the *B. pumilus* CynD enzyme towards extended fibre formation therefore already exists since the mechanism of termination is not absolute, possibly due to greater flexibility in the C surface or poorly defined interactions at the E surface, or a combination of both.

The mechanism for this remains to be elucidated, however preliminary studies have shown that the ATCC isolate of this enzyme does not form extended fibres below pH5.4. The ATCC sequence differs from the C1 isolate in a few residues but of special interest are the histidines in the C-terminal extension which are not present in the ATCC isoform. The solvent accessible histidines residues of the C-terminus are implicated in the extension mechanism since it is the only residue with pKa around pH6. It is speculated that these residues become charged below pH6 thus creating electrostatic repulsion on the inner surface of the helix. This would permit more space to accommodate the C-terminal extension thereby securing the interactions required for the preservation of helical symmetry and allowing additional subunits to continually bind to the ends of the helix in creation of an extended fibre.

Figure 4.4.1: Varying forms of the cyanide dihydratase (CynD) enzyme from *B. pumilus*. At optimal pH, the enzyme is seen as a short, terminating spiral of 14 subunits (left and centre) with a region of density of partial occupancy at the terminating ends that is indicative of dual states between a 14 and an 18 subunit structure. Below pH5.4, the structure becomes an extended helix as shown on the far right.

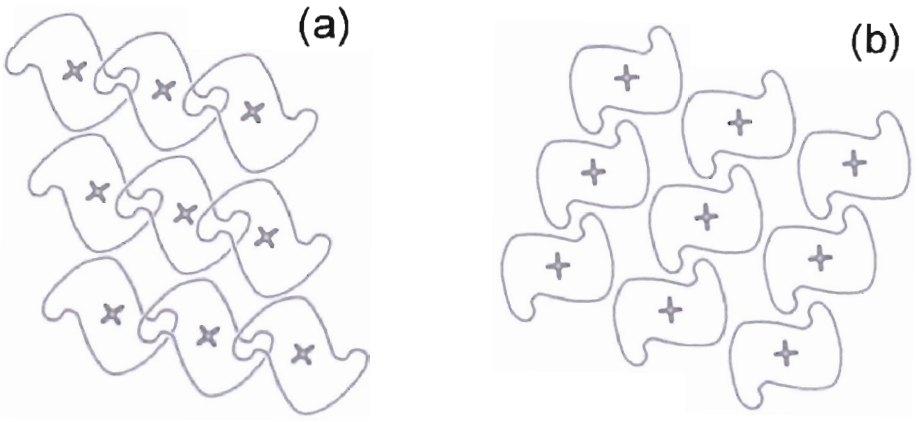


Figure 4.4.2: Schematics of the subunit binding of the dimers of (a) *B. pumilus* where it appears as if the insertions overlap. However, due to the narrow helical diameter of the fibre, the dimers are rotated by 77°, thereby allowing the extended insertions to interlock and (b) *G. sorghi.*

The optimum docking of the dimer model into the *G. sorghi* structure is almost at right angles to the optimum docking of the dimer into the *B. pumilus* and *P. stutzeri* structures. Given the inherent symmetry of these molecules and apparent preservation of the interacting interfaces, this seems highly unlikely and an alternative explanation is required.

4.5 The helical net and its reciprocal lattice - the (n, l) plot

Since the convergent symmetry parameters of both the *B. pumilus* and *G. sorghi* structures were very similar, the helical plots were almost identical and thus both of them could be described as 14:3 helices. The (*n*,*l*) plot was constructed using the formula $l=14m+3n$ (see section *1.2.6* for a more detailed explanation of the reciprocal lattice of the helical net). This allowed the indexing of layer lines in the diffraction pattern.

When the helical net is plotted with a pitch of 4.7, the closest integer number of subunits per complete turn (360* n) where {n=1,2,3...} is fourteen subunits (dimers) per three turns. Thus, the simplest way the structure can be described is as a 14:3 helix which is then used to construct an (*n*, *l*) plot.

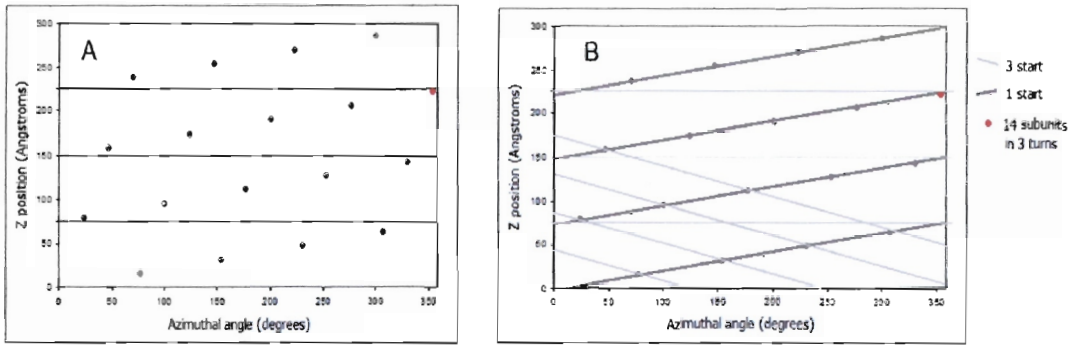


Figure 4.5.1: Helical net plotted with18 subunits. (A)The red dot indicates the subunit closest to an integer number of subunits per rotation (360* n) where {n=1,2,3...}. The simplest way the structure can be described is a 14:3 helix where there are fourteen subunits (dimers) per three turns. (B) The lines indicate the one-start left handed helix and the three-start right handed helix.

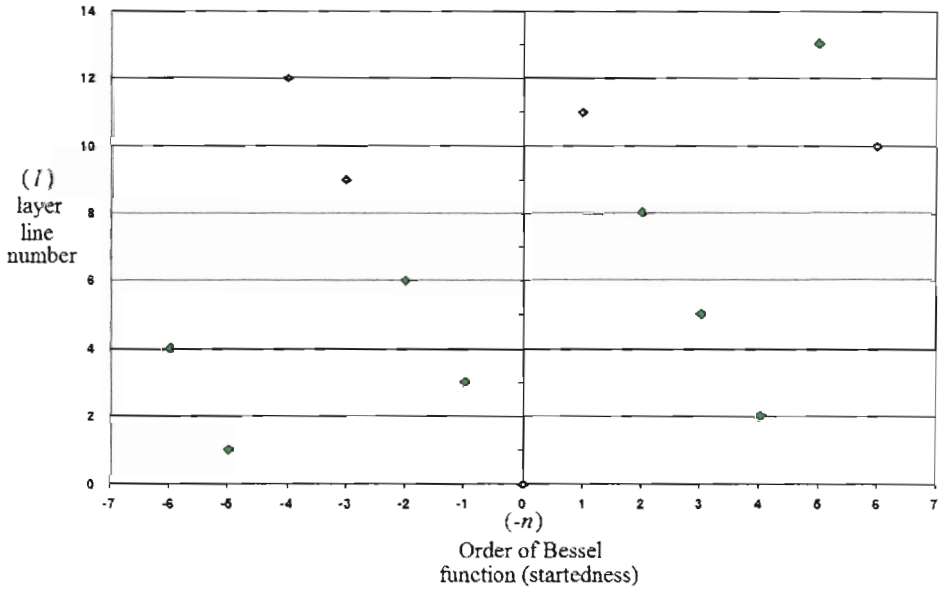


Figure 4.5.2: Reciprocal lattice (*n,l* plot) of the helical net for a 14:3 helix. Reflections found in the diffraction pattern of such a helix can be described by this selection rule. The order of a given layer line of the power spectrum can therefore be calculated by comparison with this reciprocal lattice of the helical net.

4.6 Analysis of diffraction patterns

By analysis of the (*n,l*) plot it was possible to index the layer lines of the diffraction patterns. In the shadowing, two left handed reflections and one-right handed reflection are present which can be indexed as Bessel orders J1, J2 and J3 respectively. Correspondingly, both the diffraction patterns of the negative stain images and cryo images showed reflections on layer lines 3, 5 and 6 accordingly indexed as Bessel orders

J1, J2 and J3. However, a difference was noticed in the negative stain reconstruction since reflections related to a four start helix were seen on the layer line two which were not present in the diffraction patterns from the other data sets. The appearance of this reflection in the negative stain was exclusive to the *G. sorghi* structure since the *B .pumilus* structure showed layer lines corresponding only to the J1, J2 and J3 helices. The power spectrum generated from the images of the *G. sorghi* CHT mutant also showed the presence of the anomalous reflections on layer line two, so the phenomenon seemed to be exclusive to the *G. sorghi* structures.

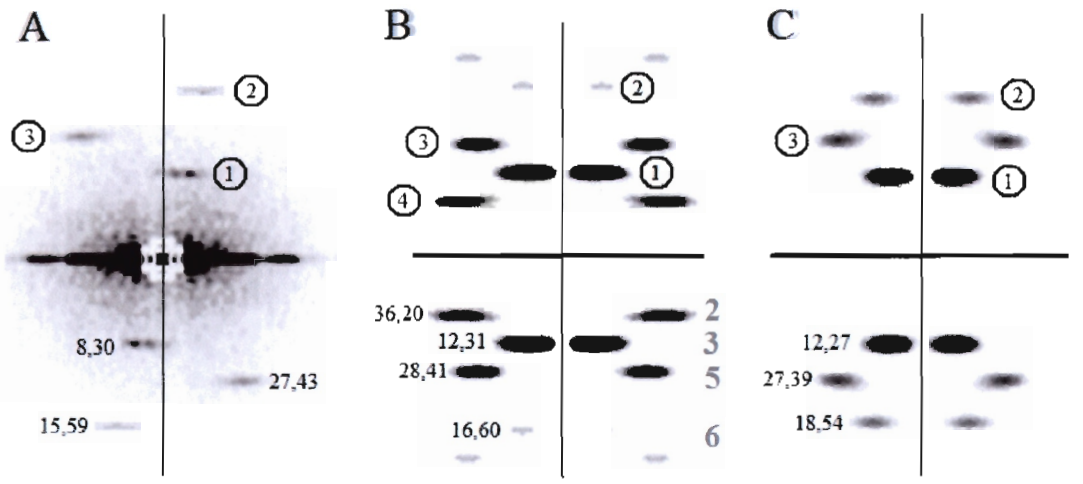


Figure 4.6.1: Principal maxima of the diffraction patterns from different data sets of the *G. sorghi* CHT. Bessel orders are shown as a ringed numbers aside their corresponding layer lines above the equator. Below the equator are the coordinates of the layer lines in pixels (distance from the meridian, distance from the equator). The layer line numbering is shown in grey according to the diffraction pattern in (B) based on a 14:3 helix. (A) Averaged power spectrum of rotationally aligned shadowed images. (B) Power spectrum of the negative stain reconstructed volume (C) Power spectrum of the cryo reconstructed volume.

4.7 Flattening in negative stain

A micrograph of a negatively stained biological particle gives an approximation to the projection of the density distribution of the surrounding stain but only if the variations in the distribution of stain are entirely caused by the structure of the particle. If the structure is perfectly helical, then the diffraction pattern can be regarded as a section through the three-dimensional transform the particle (Moody, 1971).

The notion of structural collapse and distortion of biological samples in negative stain has been reported in many instances in the literature (Trus and Steven, 1984; Moody, 1967b; Kiselev et al., 1990; Ohi et al., 2004; Harris, 1997; Baker and Caspar, 1984). The occurrence of this collapse and distortion is attributed to the surface tension forces that occur during dehydration of the specimen. In some literature, the extent of specimen flattening and shrinkage of helices has been as much as 30% (Moody, 1967a). In the single particle domain, ribosomes have been shown to be susceptible to flattening of up to 1/5 of the height of the particle without enlarging particle dimensions in the substrate plane (Kiselev et al., 1990). Flattening in negative stain depends on the preparation technique used; preservation has been shown to be superior in uranyl acetate as opposed to ammonium molybdate but a considerable amount of flattening of large particles can still occur in uranyl acetate. Distortion may be somewhat prevented by thickening the embedding layer via addition of 1% trehalose (Harris, 1997) – but this protectant was not added to the negative staining solution in these experiments.

Negatively stained grids are often seen with areas where there are numerous cracks which suggest that once the stain ceases to be fluid, it undergoes contraction. Hydrated protein embedded in the stain would therefore be exposed to contraction forces upon drying but the attachment to the grid of the far surface would restrict the lateral contraction of the base of the protein thus specimens will collapse to a greater extent on the side facing away from the grid. It is assumed that the effect of this contraction on a broad and uniformly thick layer of stain is to produce flattening of the embedded particles since contractile movement in the plane of the grid would be restricted by the carbon film. Nevertheless, if cracks are to develop in a thin layer of stain, contraction could lead to movements in the plane of the grid since it would no longer be restrained by the carbon film – but only in the upper parts of the stain layer (Moody, 1967a).

The structure of the biological specimen also affects the degree of flattening that it will experience in negative stain. It is possible that thin filaments or tubular structures may not undergo flattening if the diameter versus wall thickness ratio is small but in filaments with

thin walls or a large diameter, the supporting effect of the stain is insufficient to prevent flattening (Kiselev et al., 1990). Uneven distortion may occur if the specimen is supported anisotropically which presumably occurs if there is a thin layer of stain only supporting the side of the specimen in contact with the grid. Distortions such as flattening of helical particles are detected in the diffraction patterns of the images since such distortion leads to obvious loss of symmetry. Similar results may also result from undistorted specimens if they are non-uniformly penetrated by stain (Moody, 1967a).

Reports of flattening in the literature:

Moody notes that the most serious interference causing deviation from the requirements of the helical diffraction theory comes from flattening of the helix. In his studies (Moody, 1967a) of specimens of the polysheath of bacteriophage T4, he presented an express example of distortion effects on the diffraction patterns of helices. He utilizes the term 'surface-lattice' to define the intersection between the prominent helical grooves on the upper/near side (side facing away from the carbon) of the helical structure. In the polysheath helices, two types of 'grooves' were shown to exist, namely a 'long-pitch' 12 start helix and a 'short-pitch' 6 start helix. In shallowly-embedded specimens, curved sinusoidal lines derived from the 12 start helix on the far side (grooves on the side opposite the surface lattice) of the protein were visible as dark oblique lines on the electron micrographs. However, the width and separation of these lines were considerably greater than the straight lines derived from the 12 start helix of the surface lattice. The 12 start helix of the surface lattice gave rise to the diffraction spots L in fig5. The spot termed L' which derives from the sinusoidal lines on the far side of the helix is found on exactly the same layer line as L which means that it has the same axial repeat distance as the 12 start helix of the surface lattice. The 6-start helix of the surface lattice gave rise to diffraction spots S and the corresponding weaker spot on the same layer line from the opposite side of the helix which was termed S'.

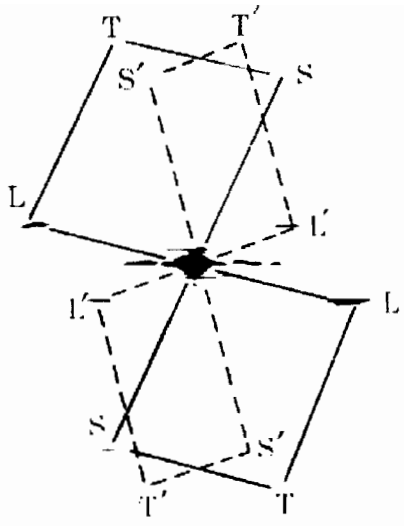


Figure 4.7.1: The diffraction spots derived from the prominent surface of the polysheath are joined with lines and marked with unprimed letters. Those relating to the other side are joined with dashed lines and are indicated by primed letters. S, S′ are diffraction spots resulting from the 6 start helix (short pitch) and L,L′ result from the 12 start helix (long pitch) (Moody, 1967a).

In the diffraction pattern, the diffraction spots L and S from the surface lattice were distinctly further from the meridian than the corresponding spots L' and S' from the other surface. This indicated either a lateral contraction of the surface lattice or a lateral expansion of the far surface. By analysis of deeply-embedded polysheath specimens, it was seen that the ratio of the axial repeat distance of the 12start:6start helix was approximately the same as for the shallowly-embedded specimen. However, it was found that both L and L' in the deeply-embedded specimen, as well as L' in the shallowly-embedded specimen were equally far from the meridian but the corresponding distance of L in the shallowly-embedded specimen was clearly larger than all the others. This indicated shrinkage of the lateral separation of the corresponding 12 start helix and implied that the surface lattice of the shallowly-embedded specimen had undergone lateral contraction when compared to the deeply-embedded specimen. The far surface however, nearer the carbon grid, did not show any distortion effects.

Trus *et al.* addressed the issue of how to integrate negative staining EM with X-ray fibre diffraction of helices by using computational simulations to investigate staining properties of helical models (Trus and Steven, 1984). Diffraction patterns of stained and unstained

models of F-actin, myosin S1-decorated actin, and microtubules were compared to test the reliability of selecting well preserved specimens as those whose diffraction patterns most closely matched the X-ray patterns of hydrated fibres. They concluded that the criterion is likely to be reliable if the stain layer is considerably thicker than the outer diameter of the specimen but noted that dynamic scattering effects such as beam-broadening would limit the resolution of such specimens. The more relevant scenario in a practical situation is that there would be relatively thin staining thereby producing "edge effects" in the stain distribution, resulting in differences between diffraction patterns of stained and unstained specimens. They also modelled shrinkage and flattening of helices and noted the changes in diffraction patterns due to molecular distortions. It was reported that it is not possible to distinguish between the edge effects or molecular distortions when analysing the differences between diffraction patterns of stained and unstained helices.

Using Moody's terminology as described in section 1.2.6 (methodological review), the diffraction pattern of a flattened continuous helix of pitch P wound around a cylinder of radius r is considered. Since the z-axis coincides with the long axis of the cylinder, flattening will compress the cylinder into the xz-plane and the Fourier transform of this projection in the reciprocal $\xi\zeta$-plane should be affected. Provided they are uniform along the length of the particle, the distortions cannot change the ζ co-ordinates of the layer lines. Uniformity can be checked since any serious departure from uniformity should cause splitting or elongation in the ξ dimension of the layer line. The ξ co-ordinates of the principal maxima (ξ_M) are dependant on the degree of flattening of each surface.

4.8 Evidence for flattening of the G. sorghi structure in negative stain:

It has been shown that lateral flattening of helical structures in stain causes the diameter of the particle to appear larger. In comparison to cryo experiments, where the biological material itself is reconstructed, in negative stain the radius and resultant reconstruction is dependant on the artificial stain-excluded volume which depends on the type of stain used. By measurement of the helical diameters of the various reconstructions, it was clear that the *G. sorghi* negative stain reconstruction had a far greater radius than both the *B.*

pumilus and the cryo reconstructions which was the first indication that flattening might have taken place.

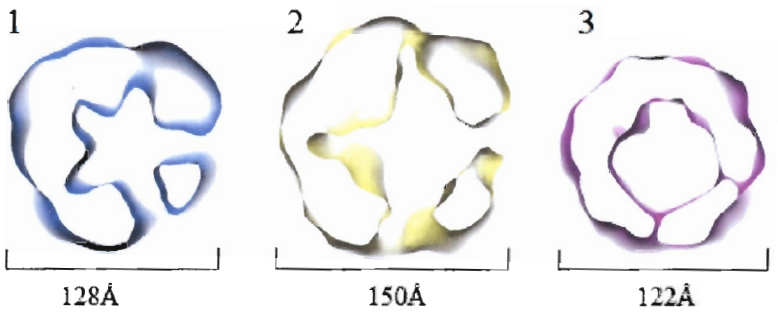


Figure 4.8.1: Effective diameters of the reconstructed volumes (A) *B. pumilus* negative stain (B) *G. sorghi* negative stain (C) *G. sorghi* cryo reconstruction.

There was slight departure from mirror symmetry in the diffraction patterns of the negative stain structures noted in the difference in intensity of the two principal maxima of layer line 3, with slight movements in their relative positions. Loss of mirror symmetry across the meridian may be attributed to the effects of differential flattening (Baker and Caspar, 1984) but is usually due to differential staining of the two surfaces of the helix rather than distortion effects (Moody, 1967a). The anomalous diffraction spot that appeared in the *G. sorghi* negative stain diffraction pattern which was not found in either the cryo or the shadowing data has a large repeat length of 109Å. The layer line is believed to be of the 4th Bessel order when compared to a 14/3 (n,l) plot.

An advantage of the freeze-drying shadowing approach is that the structural features are better preserved, and since surfaces which are not in contact with the specimen support are imaged, an unaltered representation of the surface of the helix will be provided. If the anomalous layer line originated from an internal structure, it would not be displayed in the diffraction pattern of the shadowed material since the interior of the helical particle is not imaged during the shadowing process. To investigate this hypothesis, the radial distance of the anomalous layer line was calculated so as to determine the helical diameter from which the structural information resulted.

If one takes the Bessel order of this layer line to be 4 according to the (n,l) plot of the 14:3 helix, the corresponding effective radius of the helical density which gives rise to this diffraction spot can be computed according to the position of the first zero of $J'_n(2\pi\xi r_e)$, where ξ is the distance from meridian (reciprocal Å) and r_e is the effective radius - the real space coordinate which corresponds, through reciprocity, to the position (ξ) of the principal Fourier-Bessel maximum on a given layer-line. Since $J'_4(5.317) = 0$ (see section 1.2.6), the effective radius r_e is calculated to be ~51Å.

For a simple helix, consisting of only points at the same radius from the axis, r_e is equivalent to the radial distance. Relating r_e to its radial reference point is not as easy for a helix which has a continuous density distribution but this should at least provide an approximation as to where in the structure the layer line originates. Arising from information at a diameter of 102Å within the structure, this layer line would not be visualized during shadowing; however since the diffraction of the cryo data did not exhibit these diffraction spots either, an alternative explanation was sought.

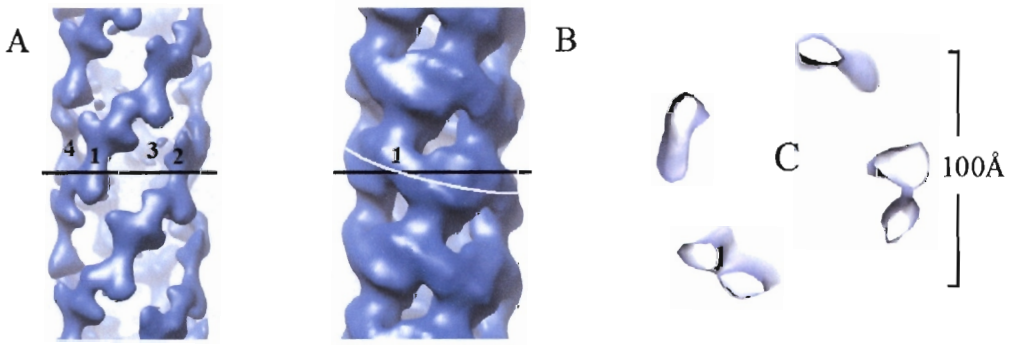


Figure 4.8.2: (A) Negative stain structure of *G.sorghi* CHT showing the four-start helix which is prominent at high contour levels. (B) Illustration of the basic helix which may be described as a left-handed one-start helix. (C) Lateral cross section of the reconstruction in (A) showing the four-start helix from above. The effective radius between the centres of density of the helices is approximately 50Å.

The pitch (P) of the 4 start helix above is measured to be approximately 440Å, therefore P/n for this feature is equal to 110Å. This corresponds to the repeat length of 1/109Å as calculated for layer line 2 from the (n,l) plot which would indicate that this diffraction spot indeed emerges from the four start helix as demonstrated in figure 4.8.2 A.

It was shown by Moody that flattening with spreading of a simple point helix results in displacement of layer line maxima towards the meridian (Moody, 1967a). This was confirmed by Trus *et al.*, who demonstrated that the ξ_M values of modelled flattened microtubules were lower in negative stain than in the unstained structure. Nevertheless, once partial flattening is extended beyond 70-50% of original thickness, the ξ_M values show a marked increase. Flattening can thus most obviously be recognized through an increase in the ξ_M values of the layer lines. The observed ξ_M values were calculated for layer lines 3, 6 and 5 (of Bessel orders n = 1,2 and 3 respectively) and compared to their expected values which should be roughly proportional to the order of the layer line. For layer line 5 (figure 4.6.1c), the observed ξ_M values for the different data sets agree. The ξ values of principal maxima of layer line 3 and layer line 6 also agree, although the second and third maxima of the layer line 3 from the negative stain data have far greater ξ_M values. Also, the amplitude of the secondary maximum of layer line 6 is equal to that of its principal maximum.

Differences that are noticeable on some layer lines may not been seen in others since certain layer lines are more sensitive to adjustments of particular parameters than others (Baker and Caspar, 1984). Baker *et al.* noticed in their studies for example, that on one particular layer line the outer peaks were sensitive to differences in contrast between the near and far sides of the helix whereas the two principal maxima were more affected by other parameters. In figure 4.8.3, the principal maxima do not vary substantially between the different samples but most of the subsidiary maxima of the negative stain data are displaced relative to the cryo data and change in amplitude is also detected. Flattening may be an explanation for the oserved phenomena, although, aaccording to Klug and Berger, a greater equatorial spacing may also be indicative of deep penetration of stain (Klug and Berger, 1964).

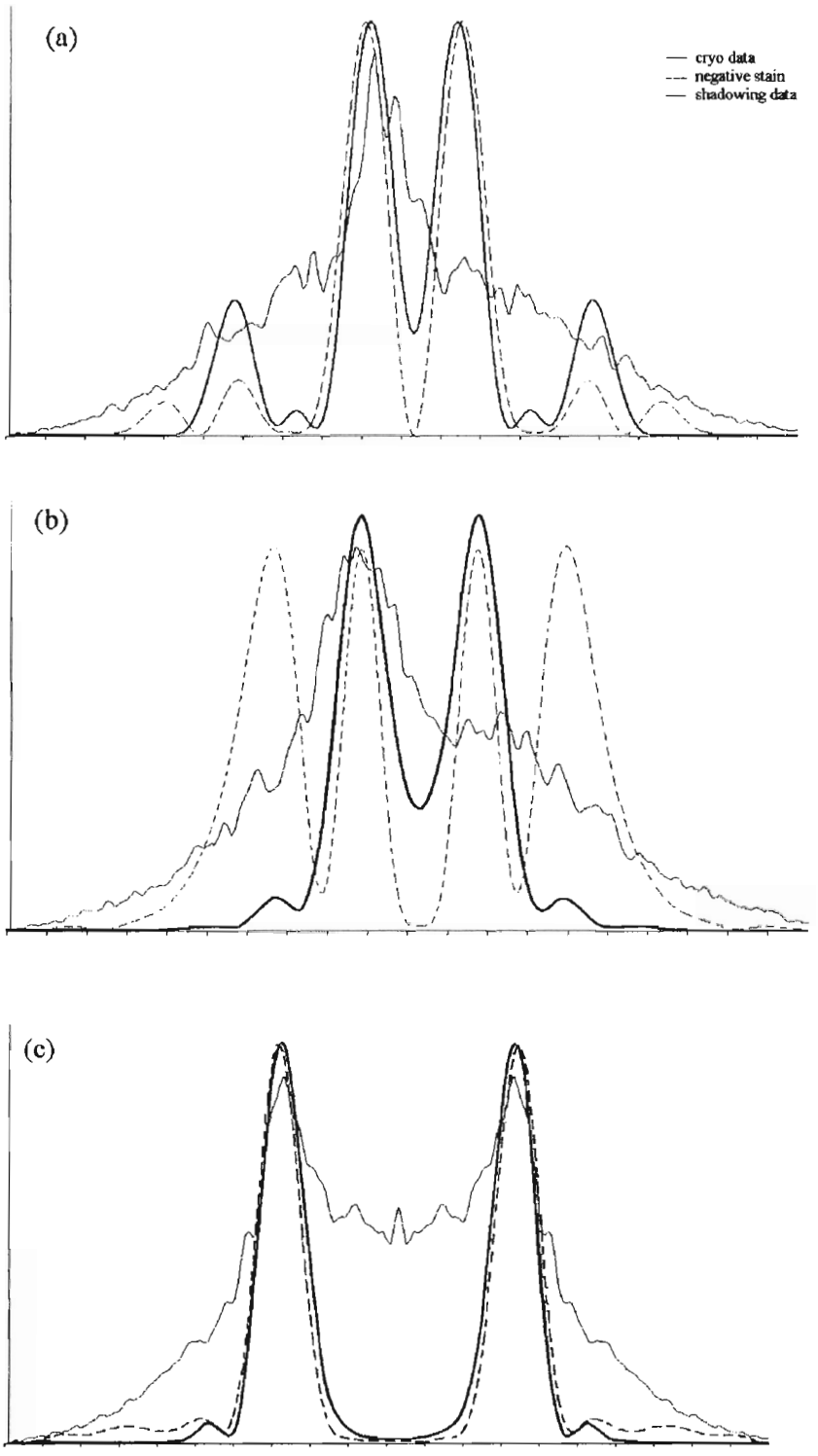


Figure 4.8.3: Comparison of the amplitudes along non-equatorial layer lines 3, 6 and 5 (n = 1,2 and 3 respectively) for the cryo (solid) and negative stain (dashed) and shadowing (dotted) data. The data sets have not been scaled. (A) Layer line 3, (B) layer line 6 and (C) layer line 5. The data sets differ most on layer line 6 (n = 1) where ξ_M values of the second and third maxima of the negative stain structure are greater than in the cryo and the shadowing data. The shadowing data were taken from the power spectrum of the rotationally aligned average and the other data were taken from the power spectra of the reconstructions. In (c) the shadowing data is shown as reflected across the meridian.

Table 4.1: Distances measured of diffraction spots on layer lines 3,6 and 5 corresponding to Bessel orders 1,2 and 3 of the different samples. ξ denotes distance from the meridian and ζ denotes distance from the equator. Distances are in pixels – all samples were at 4.25Å/pixel padded in a 512 box.

Distance measured	1start1	1start2	1start3	2start1	2start2	3start1	4start
shadowing ξ	8				15	27	
shadowing ζ	29.5				59	43	
ratio shadowing	0.271				0.254	0.627	
negative stain ξ	12	44	63	9	36	28	36
negative stain ζ	31	31	31	60	60	41	20
ratio negative stain	0.387	1.419	2.032	0.15	0.6	0.683	1.8
cryo ξ	12	45			18	27	
cryo ζ	27	26.5			54	39	
ratio cryo	0.444	1.698			0.333	0.692	

In table 4.1, it can be seen that the measured ratios of the ξ vs ζ of the negative stain data compared to the cryo and shadowing data are different to what the expected value would be according to the selection rule. It may be speculated that this is mainly due to a change in ζ spacing, as seen in the diffraction pattern shown in figure 4.6.1B. Uniform contraction or dilation in the z direction or a variation in the pitch of the helix would result in a change in the spacing of the layer lines in ζ, which would provide a rationalization for the change of the ξ vs ζ ratios seen here (Baker and Caspar, 1984). The fact that there may be two or more conformations present may further be exemplified by the fact that the layer spacing between layer line 5 and 6 is altered in the negative stain data when compared to the cryo data (figure 4.6.1).

The fact that the $n = 4$ layer line can be described by the selection rule suggests that the otherwise dormant feature, from which the diffraction spot arises, is enhanced by the negative stain. During the reconstruction process, the stain exclusion volume is reconstructed, therefore the only mechanism whereby a pseudo feature may arise is if the stain does not sufficiently penetrate the sample. If there is a crevice in the molecule which is not filled by the stain, the crevice will appear as density in the reconstruction. This may provide a rational explanation for the unexpected appearance of the 4 start structure in the reconstruction of the negatively stained *G. sorghi* fibres. An experiment was conducted in order to investigate the effect of excising the layer plane 2 from the negative stain structure. A masking function was created to remove the anomalous layer line from the

images and the new image stack was used to reconstruct the filaments in three dimensions. The only visible effect of this experiment was that the 4 start helix had become discontinuous when the model was viewed at high contour levels. However the experiment revealed that the structure was not greatly altered in the absence of this information. It is therefore evident that there are more factors involved than insufficient stain penetration.

Differential flattening has been shown to lead to overlap of reciprocal lattice points from different layer lines especially if the two lattices necessarily overlap at the meridional layer line which is usually the case for a helical structure with (in this case 14/3) screw symmetry. Although, differential flattening can only be implicated upon presentation of a particular line of evidence which has not been assessed at this stage. As shown in figure 4.7.1, there is a good indication that differential flattening has occurred if layer lines display asymmetry across the meridian. This has not been assessed at this stage since one would require knowledge of the phases of each layer line in order to sort the images into those which display information from the near side, and those that display information from the far side of the helix. Without this information, these images will be averaged and the asymmetry across the meridian will be displayed as symmetrical extended diffraction spots on a layer line.

It is clear that several explanations may be presented in an attempt to clarify the discrepancy in the diffraction pattern of the *G. sorghi* negative stain structure, but further investigations are required in order to unambiguously identify a particular causative factor. In conclusion, the negative stain structure derived from the *G. sorghi* is correct as far as the available data is concerned, however an anomalous layer line in the diffraction pattern does not concur with the available shadowing and cryo data. This leads to the suspicion that the negative stain structure is incorrect and interpretation of the structure should be approached with caution. It is possible that the larger diameter of the *G. sorghi* fibre influenced the degree of flattening in negative stain allowing the structure to collapse in contrast to the *B. pumilus* fibre with its small diameter and potentially more rigid structure which would be able to withstand the dehydration forces.

4.9 Conclusions and future prospects

In light of the aims of this research, it may be said that the following list of accomplishments were achieved. The handedness of the extended of fibres of CHT and CynD were unequivocally determined to be left handed helices by high-resolution unidirectional metal shadowing. The helical quaternary structures of CynD from *Bacillus pumilus* and the CHT from *Gloeocercospora sorghi* were solved using the method of Iterative Helical Real Space Reconstruction (IHRSR), showing that they converged on similar symmetry parameters. The structure of an active mutant of *G. sorghi* CHT, namely R87Q was also solved. Both the *G. sorghi* CHT and *B. pumilus* CynD structures were docked with homology models of dimers with the A surface preserved. Models were sometimes docked ambiguously but it was premised that the carboxy-terminus should face inwards since unoccupied densities filled by insertions and available knowledge of ion-pair interactions could be more easily rationalized in this manner. Docking thereby allowed the interpretation of biological interactions and provided justifications for the inactivities of mutants E82V and R91Q. The interacting residues leading to spiral formation were speculated upon and the C and D surfaces involved in oligomerisation were identified.

It was sought to rationalize the layer line data from diffraction patterns of the images. Even though the negative stain *G. sorghi* CHT structures converged to the same solution, it was impossible to reconcile the diffraction pattern of the negative stain structure with the information obtained in the shadowing experiment and the cryo reconstruction due to an anomalous layer line present in the power spectrum as well as displacement and change in amplitude of subsidiary maxima on other layer lines. The average power spectrum of the rotationally aligned images was however identical to the power spectrum generated from the reconstructed volume, indicating that the structure of the *G. sorghi* CHT was self consistent. Through closer inspection and analysis of the layer line data it was possible to put forward a 'flattening in negative stain' hypothesis to explain the origin of the layer line.

The possibility of the *G. sorghi* IHRSR model being incorrect due to the collapse of the structure in negative stain was discussed. An obvious area for future work is to solve this structure to high resolution using CTF correction on cryo-EM data.

It is hoped that insight into the oligomerization process will allow for the design of mutants that are susceptible to crystallization. Currently no crystal structures have been solved, largely due to the fact that helical structures are not likely to crystallize. Also the helical symmetry parameters of the nitrilase structure (4.6667 subunits per turn) creates a point group close to that of a 5_1 screw which cannot crystallize since 5-fold rotation axes are forbidden in ordinary periodic crystals (Dunitz, 1996; Steinhardt, 1996). Through better understanding of the quaternary structure of these nitrilases, mutants in which the dimer A-surface is preserved may be crystallized. The atomic structure may then be solved by x-ray crystallography, ultimately leading to greater knowledge for use in the industrial domain.

The increasing amount of sequence-space exploration as well as high-throughput screening of environmental samples is rapidly leading to discovery of novel nitrilases and their gene associations. Continuing biochemical studies will hopefully assist in the characterization of these enzymes and the elucidation of the biological functions and insights gained through the structural investigation will provide a detailed understanding of the structure – function relationship. It will ultimately lead to the possibility of engineering the enzymes according to specific substrate requirements as well as increasing factors such as temperature and pH stability.

Baker, T.S. and Caspar, D.L.D. (1984). Computer image modeling of pentamer packing in polyoma virus "hexamer" tubes. Ultramicroscopy *13*, 137-152.

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

Bartling, D., Seedorf, M., Mithofer, A. and Weiler, E.W. (1992). Cloning and expression of an Arabidopsis nitrilase which can convert indole-3-acetonitrile to the plant hormone, indole-3-acetic acid. Eur. J. Biochem. *205*, 417-424.

Basheer, S., Kut, Ö.M., Prenosil, J.E. and Bourne, J.R. (1992). Kinetics of enzymatic degradation of cyanide. Biotechnol. Bioeng. *39*, 629-634.

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

Berman, M. N. Quaternary structures of the cyanide dihydratases of *Bacillus pumilus* and *Pseudomonas stutzeri* AK61. 2003. UCT.
Ref Type: Thesis/Dissertation

Blumenthal, S.G., Hendrickson, H.R., Abrol,Y.P., and Conn, E.E. (1968). Cyanide metabolism in higher plants. 3. The biosynthesis of beta-cyanolanine. J. Biol. Chem. *243*, 5302-5307.

Boettcher, C., Stark, H. and Van Heel, M (1996). Stacked bilayer helices: a new structural organisation of amphiphilic molecules. Ultramicroscopy *62*, 133-139.

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

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

Brady, D., Beeton, A., Zeevaart, J., Kgaje, C., van Rantwijk, F. and Sheldon, R.A. (2004). Characterisation of nitrilase and nitrile hydratase biocatalytic systems. Appl. Microbiol. Biotechnol. *64*, 76-85.

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

Brown, D.T., Turner, P.D. and O'Reilly, C. (1995). Expression of the cyanide hydratase enzyme from Fusarium lateritium in Escherichia coli and identification of an essential cysteine residue. FEMS Microbiol. Lett. *134*, 143-146.

Carrio, M.M. and Villaverde, A. (2001). Protein aggregation as bacterial inclusion bodies is reversible. FEBS Lett. *489*, 29-33.

Carrio, M.M. and Villaverde, A. (2002). Construction and deconstruction of bacterial inclusion bodies. J. Biotechnol. *96*, 3-12.

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

Cluness, M.J., Turner, P.D., Clements, E., Brown, D.T. and O'Reilly, C. (1993). Purification and properties of cyanide hydratase from Fusarium lateritium and analysis of the corresponding chy1 gene. J. Gen. Microbiol. *139*, 1807-1815.

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

Cope, J. Immunolocalization of CHT in *Aspergillus nidulans* and *Neurospora crassa*. 2005. University of Cape Town.
Ref Type: Thesis/Dissertation

Crowther, R.A., DeRosier, D.J. and Klug, A. (1970). The reconstruction of a three-dimensional structure from its projections and its application to electron microscopy. Proc. R. Soc. Lond. A Biol. *Sci. 317*, 319-340.

Crowther, R.A., Padron, R. and Craig, R. (1985). Arrangement of the heads of myosin in relaxed thick filaments from tarantula muscle. J. Mol. Biol. *184*, 429-439.

Davis, R.W., Botstein, D. and Roth, J.R. (1980). Advanced Bacterial Genetics. (New York: Cold Spring Harbour Laboratory).

DeRosier, D.J. and Klug, A. (1968). Reconstruction of three-dimensional structures from electron micrographs. Nature *217*, 130-134.

DeRosier, D.J. and Moore, P.B. (1970). Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. J. Mol. Biol. *52*, 355-369.

DeSantis, G., Zhu, Z., Greenberg, W.A., Wong, K., Chaplin, J., Hanson, S.R., Farwell, B., Nicholson, L.W., Rand, C.L., Weiner, D.P., Robertson, D.E. and Burk, M.J. (2002). An enzyme library approach to biocatalysis: development of nitrilases for enantioselective production of carboxylic acid derivatives. J. Am. Chem. Soc. *124*, 9024-9025.

Duke, S.O. (2005). Taking stock of herbicide-resistant crops ten years after introduction. Pest. Manag. Sci. *61*, 211-218.

Dunitz, J.D. (1996). Symmetry arguments in chemistry. Proc. Natl. Acad. Sci. U. S. A *93*, 14260-14266.

Egelman, E.H. (2000). A robust algorithm for the reconstruction of helical filaments using single-particle methods. Ultramicroscopy *85*, 225-234.

Egelman, E.H. Lecture series: Image analysis and 3D reconstruction of helical polymers. 2004. Presented as part of the Electron Microscopy section of the coursework component of the Masters course for the South African Structural Biology Initiative.

Fernandez, R.F. and Kunz, D.A. (2005). Bacterial cyanide oxygenase is a suite of enzymes catalyzing the scavenging and adventitious utilization of cyanide as a nitrogenous growth substrate. J. Bacteriol. *187*, 6396-6402.

Frank, J. (1996). Three-Dimensional Electron Microscopy of Macromolecular Assemblies. (San Diego, California: Academic Press, inc.).

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

Fuchs, K.H., Tittman, P. and Gross, H. Cryo-microscopy and Surface Imaging of Biological Macromolecules. IECM 13-Paris , 23-24. 1994.
Ref Type: Conference Proceeding

Gross, H. (1987). High Resolution Metal Replication of Freeze-Dried Specimens. In Cryotechniques in Biological Electron Microscopy, R.A.Steinbrecht and K.Zierold, eds. (Berlin Heidelberg: Springer-Verlag), pp. 205-215.

Gross, H., Krusche, K. and Tittmann, P. Recent progress in high resolution shadowing for biological samples. Proceedings of the XIIth International Congress for Electron Microscopy , 510-511. 1990. San Francisco, CA, San Francisco Press, Inc.

Harauz, G. and Van Heel, M. (1986). Exact filters for general geometry three dimensional reconstruction. Optik *73*, 146-156.

Harper, D.B. (1985). Characterization of a nitrilase from Nocardia sp. (Rhodochrous group) N.C.I.B. 11215, using p-hydroxybenzonitrile as sole carbon source. Int. J. Biochem. *17*, 677-683.

Harris, J.R. (1997). Freeze-dry negative staining. In Negative staining and cryoelectron microscopy, (Oxford, UK: BIOS Scientific Publishers Ltd), pp. 57-58.

Hillebrand, H., Tiemann, B., Hell, R., Bartling, D. and Weiler, E.W. (1996). Structure of the gene encoding nitrilase 1 from Arabidopsis thaliana. Gene *170*, 197-200.

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

Hymes, J. and Wolf, B. (1999). Human biotinidase isn't just for recycling biotin. J. Nutr. *129*, 485S-489S.

Jandhyala, D., Berman, M., Meyers, P.R., Sewell, B.T., Willson, R.C. and Benedik, M.J. (2003). CynD, the cyanide dihydratase from Bacillus pumilus: gene cloning and structural studies. Appl. Environ. Microbiol. *69*, 4794-4805.

Jandhyala, D.M., Willson, R.C., Sewell, B.T. and Benedik, M.J. (2005). Comparison of cyanide-degrading nitrilases. Appl. Microbiol. Biotechnol. *68*, 327-335.

Jones, D.T. (1999). GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J. Mol. Biol. *287*, 797-815.

Joyeux, L. and Penczek, P.A. (2002). Efficiency of 2D alignment methods. Ultramicroscopy *92*, 33-46.

Kiselev, N.A., Sherman, M.B. and Tsuprun, V.L. (1990). Negative staining of proteins. Electron Microsc. Rev. *3*, 43-72.

Klug, A. and Berger, J.E. (1964). An optical method for the analysis of periodicities in electron micrographs, and some observations on the mechanism of negative staining. J. Mol. Biol. *10*, 565-569.

Klug, A., Crick, F.H.C. and Wyckoff, H.W. (1958). Diffraction by helical structures. Acta Cryst. *11*, 199-213.

Knowles, C.J. (1988). Cyanide utilization and degradation by microorganisms. Ciba Found. Symp. *140*, 3-15.

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

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

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

Kutz, A., Muller, A., Hennig, P., Kaiser, W.M., Piotrowski, M. and Weiler, E.W. (2002). A role for nitrilase 3 in the regulation of root morphology in sulphur-starving Arabidopsis thaliana. Plant J. *30*, 95-106.

Larsen, M., Ucisik, A.S. and Trapp, S. (2005). Uptake, metabolism, accumulation and toxicity of cyanide in Willow trees. Environ. Sci. Technol. *39*, 2135-2142.

Ludtke, S.J., Baldwin, P.R. and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82-97.

Ludwig-Muller, J. and Cohen, J.D. (2002). Identification and quantification of three active auxins in different tissues of Tropaeolum majus. Physiol Plant *115*, 320-329.

McGuffin, L.J., Bryson, K. and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics. *16*, 404-405.

Moazed, D. and Noller, H.F. (1989). Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. Cell *57*, 585-597.

Moody, M.F. (1967a). Structure of the sheath of bacteriophage T4. I. Structure of the contracted sheath and polysheath. J. Mol. Biol. *25*, 167-200.

Moody, M.F. (1967b). Structure of the sheath of bacteriophage T4. II. Rearrangement of the sheath subunits during contraction. J. Mol. Biol. *25*, 201-208.

Moody, M.F. (1971). Application of optical diffraction to helical structures in the bacteriophage tail. Philos. Trans. R. Soc. Lond B Biol. Sci. *261*, 181-195.

Muchmore, C.R., Krahn, J.M., Kim, J.H., Zalkin, H. and Smith, J.L. (1998). Crystal structure of glutamine phosphoribosylpyrophosphate amidotransferase from Escherichia coli. Protein Sci. *7*, 39-51.

Nagasawa, T., Wieser, M., Nakamura, T., Iwahara, H., Yoshida, T. and Gekko, K. (2000). Nitrilase of Rhodococcus rhodochrous J1. Conversion into the active form by subunit association. Eur. J. Biochem. *267*, 138-144.

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

Nolan, L.M., Harnedy, P.A., Turner, P., Hearne, A.B. and O'Reilly, C. (2003). The cyanide hydratase enzyme of Fusarium lateritium also has nitrilase activity. FEMS Microbiol. Lett. *221*, 161-165.

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

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

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

Ohi, M., Li, Y., Cheng, Y. and Walz, T. (2004). Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. Biol. Proced. Online. *6*, 23-34.

Pace, H.C. and Brenner, C. (2001). The nitrilase superfamily: classification, structure and function. Genome Biol. *2*, REVIEWS0001.

Pace, H.C., Hodawadekar, S.C., Draganescu, A., Huang, J., Bieganowski, P., Pekarsky, Y., Croce, C.M. and Brenner, C. (2000). Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers. Curr. Biol. *10*, 907-917.

Paul, D., Patwardhan, A., Squire, J.M. and Morris, E.P. (2004). Single particle analysis of filamentous and highly elongated macromolecular assemblies. J. Struct. Biol. *148*, 236-250.

Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y., Tillib, S., Draganescu, A., Wermuth, P., Rothman, J.H., Huebner, K., Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M. (1998). Nitrilase and Fhit homologs are encoded as fusion proteins in Drosophila melanogaster and Caenorhabditis elegans. Proc. Natl. Acad. Sci. U. S. A *95*, 8744-8749.

Penczek, P., Radermacher, M. and Frank, J. (1992). Three-dimensional reconstruction of single particles embedded in ice. Ultramicroscopy *40*, 33-53.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605-1612.

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

Podar, M., Eads, J.R. and Richardson, T.H. (2005). Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study. BMC. Evol. Biol. *5*, 42.

Pollmann, S., Muller, A., Piotrowski, M. and Weiler, E.W. (2002). Occurrence and formation of indole-3-acetamide in Arabidopsis thaliana. Planta *216*, 155-161.

Robertson, D.E., Chaplin, J.A., DeSantis, G., Podar, M., Madden, M., Chi, E., Richardson, T., Milan,A., Miller, M., Weiner, D.P., Wong, K., McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Snead, M.A., Keller, M., Mathur, E., Kretz, P.L., Burk, M.J. and Short, J.M. (2004). Exploring nitrilase sequence space for enantioselective catalysis. Appl. Environ. Microbiol. *70*, 2429-2436.

Rosenthal, P.B. and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. *333*, 721-745.

Saibil, H.R. (2000). Macromolecular structure determination by cryo-electron microscopy. Acta Crystallogr. D. Biol. Crystallogr. *56 (Pt 10)*, 1215-1222.

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

Schleif, R. (1987). The L-Arabinose Operon. In Escherischia coli and Salmonella typhimurium: cellular and molecular biology, F.Neidhardt, ed. (Washington DC: American Society for Microbiology), pp. 1473-1481.

Sewell, B.T., Berman, M.N., Meyers, P.R., Jandhyala, D. and Benedik, M.J. (2003). The cyanide degrading nitrilase from Pseudomonas stutzeri AK61 is a two-fold symmetric, 14-subunit spiral. Structure. (Camb.) *11*, 1413-1422.

Sewell, B.T., Thuku, R.N., Zhang, X., and Benedik, M.J. (2005). Oligomeric structure of nitrilases: effect of mutating interfacial residues on activity. Ann. N. Y. Acad. Sci. *1056*, 153-159.

Stanley, C.M., Hymes, J. and Wolf, B. (2004). Identification of alternatively spliced human biotinidase mRNAs and putative localization of endogenous biotinidase. Mol. Genet. Metab *81*, 300-312.

Steinhardt, P.J. (1996). New perspectives on forbidden symmetries, quasicrystals, and Penrose tilings. Proc. Natl. Acad. Sci. U. S. A *93*, 14267-14270.

Swango, K.L., Hymes, J., Brown, P. and Wolf, B. (2000). Amino acid homologies between human biotinidase and bacterial aliphatic amidases: putative identification of the active site of biotinidase. Mol. Genet. Metab *69*, 111-115.

Tietz Video and Imaging Processing Systems. EM-MENU. 2003. Germany.

Trus, B.L. and Steven, A.C. (1984). Diffraction patterns from stained and unstained helices: consistency or contradiction? Ultramicroscopy *15*, 325-335.

Van Heel, M., and Schatz, M. (2005). Fourier shell correlation threshold criteria. J. Struct. Biol. *151*, 250-262.

Van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy *21*, 111-123.

Vorwerk, S., Biernacki, S., Hillebrand, H., Janzik, I., Muller, A., Weiler, E.W. and Piotrowski,M. (2001). Enzymatic characterization of the recombinant Arabidopsis thaliana nitrilase subfamily encoded by the NIT2/NIT1/NIT3-gene cluster. Planta *212*, 508-516.

Wajant, H. and Effenberger, F. (2002). Characterization and synthetic applications of recombinant AtNIT1 from Arabidopsis thaliana. Eur. J. Biochem. *269*, 680-687.

Wang, P., Matthews, D.E. and VanEtten, H.D. (1992). Purification and characterization of cyanide hydratase from the phytopathogenic fungus Gloeocercospora sorghi. Arch. Biochem. Biophys. *298*, 569-575.

Wang, P., Sandrock, R.W. and VanEtten, H.D. (1999). Disruption of the cyanide hydratase gene in Gloeocercospora sorghi increases its sensitivity to the phytoanticipin cyanide but does not affect its pathogenicity on the cyanogenic plant sorghum. Fungal. Genet. Biol. *28*, 126-134.

Wang, P. and VanEtten, H.D. (1992). Cloning and properties of a cyanide hydratase gene from the phytopathogenic fungus Gloeocercospora sorghi. Biochem. Biophys. Res. Commun. *187*, 1048-1054.

Wang, W.C., Hsu, W.H., Chien, F.T. and Chen, C.Y. (2001). Crystal structure and site-directed mutagenesis studies of N-carbamoyl-D-amino-acid amidohydrolase from Agrobacterium radiobacter reveals a homotetramer and insight into a catalytic cleft. J. Mol. Biol. *306*, 251-261.

Watanabe, A., Yano, K., Ikebukuro, K. and Karube, I. (1998). Investigation of the potential active site of a cyanide dihydratase using site-directed mutagenesis. Biochim. Biophys. Acta *1382*, 1-4.

Williams, R.C. and Fisher, H.W. (1970). Electron microscopy of tobacco mosaic virus under conditions of minimal beam exposure. J. Mol. Biol. *52*, 121-123.

Wolf, B. and Jensen, K. (2005). Evolutionary conservation of biotinidase: implications for the enzyme's structure and subcellular localization. Mol. Genet. Metab *86*, 44-50.

Yang, S., Yu, X., Galkin, V.E. and Egelman, E.H. (2003). Issues of resolution and polymorphism in single-particle reconstruction. J. Struct. Biol. *144*, 162-171.