The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

pH-dependence of the Quaternary structure of the Cyanide Dihydratase from *Bacillus pumilus*

by

Johann J. Eicher

Department of Molecular and Cell Biology
Faculty of Science
University of Cape Town

A thesis submitted in fulfillment of the requirements
for the degree of Master of Science

Cape Town
December 2007

PLAGIARISM DECLARATION

1. I know that plagiarism is wrong. Plagiarism is using another's work and to pretend that it is ones own.

2. Each significant contribution to, and quotation in, this essay/report/project/... from the work, or works of other people has been attributed, cited and referenced.

3. This thesis is my own work.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

SIGNATURE: 

DATE: 19-5-2006

ABSTRACT

Nitrilases are moderately ubiquitous nitrile/cyanide-degrading enzymes, found in both eukaryotes (animals, fungi, plants) and prokaryotes (archaea, bacteria) which catalyse the condensation and hydrolysis of a wide range of non-peptide nitrile substrates and are involved in nitrile-post-translational modification.

As Cyanide and related compounds are used extensively by humans in various industrial processes which, due to carelessness and inadequate waste-management systems, contribute significantly to the levels of toxic cyanide contamination in the environment nitrilases have been speculated to be useful for bioremediation amongst other things.

Nitrile/cyanide hydrolysing enzymes have a broad range of substrates and they function via four known pathways. Nitrilase and cyanide dihydratase completely hydrolyse nitriles and HCN respectively to yield the corresponding acid and ammonia without going via an amide intermediate. Nitrile hydratase and cyanide hydratase perform a single hydrolysis producing the corresponding amide and formamide, respectively.

The nitrilases are known to form extensive quaternary structures including dimers, spirals and rods/ helices. Generally microbial nitrilases exist as homo-oligomers having a large molecular weight (> 300 kDa). These enzymes are known to oligomerise under conditions of substrate activation (*Rhodococcus rhodocrous*) and pH change as is the case for the Cyanide dihydratase from *Bacillus pumilus C1* ($CynD_{pum}$) which exists as a terminating spiral of ~16 subunits above pH 6 but forms a long helical fibre below ~pH 6.

In this project the Cyanide dihydratase from strain 8A3 of *B. pumilus* was analysed using electron microscopy at pH of 5.4, 6 and 8. These data were reconstructed at pH 6 and pH 8 using the single particle reconstruction technique to resolutions of 29Å and 31Å respectively.

It is shown that at pH 6 the enzyme consists of 20 subunits (10 dimers) and at pH 8 22 subunits (11 dimers). These models show that $CynD_{pum}$ exists as an oligomeric spiral that terminates by decreasing the helical radius and tilting the terminal subunits toward the helical axis. Below pH 5.4 $CynD_{pum}$ from strain 8A3 does not extend into a fibre as in C1, this is explained to be due to the lack of 3 key histidine residues found on the C-terminal tail of $CynD_{pum}$ which point into the inner cavity of the spiral and become charged below pH 6 producing a repulsion preventing the termination of the spiral by narrowing of the helical radius and thus encouraging extension into the helical form.

For all the Bagginses and Boffins,
Tooks and Brandybucks,
Grubbs and Chubbs,
Burrowses and Hornblowers,
Bolgers and Bracegirdles,
Goodbodies, Brockhouses
and Proudfoots
who remain my unrelenting
educators in the art of
walking out of the front door.

Contents:

1. Literature review

1.1 Introduction

Nitrilases are reasonably ubiquitous nitrile/cyanide-degrading enzymes, found in both eukaryotes (animals, fungi, plants) and prokaryotes (archaea, bacteria) which catalyse the condensation and hydrolysis of a wide range of non-peptide nitrile substrates and are involved in nitrile-post-translational modification (Pace & Brenner, 2001). The nitrilases have evolved in response to the presence of cyanide compounds which are considered (along with other molecules such as ammonia, carbonyls, amines etc.) to have been part of the primordial soup that progressed through an intermediate group of nitriles into the amino acid alphabet of life. Thus cyanide/nitrile compounds have been an important part of nature before the inception of life (Commeyras *et al.*, 2004).

Inorganic and Organic cyanide compounds are found widely dispersed in nature in a range of forms including hydrogen cyanide (HCN), cyanide salts (e.g. KCN, NaCN), more complex metal-cyanides (of which the most abundant species is CN coordinated to iron, e.g. $[Fe(CN)6]^{-4}$, $[Fe(CN)6]^{-3}$), thiocyanates ($[SCN]^{-1}$, $[Fe(NCS)(H_2O)_5]^{2+}$) and various nitriles (e.g. cyanoglycosides, pesticides) (Ebbs, 2004; Kjeldsen, 1999). These compounds are commonly found in many organisms (mainly plants) as a synthesis product and in soil and water systems as a toxic byproduct of human activity (Baxter & Cummings, 2006). Cyanide compounds vary in toxicity from mildly toxic/distasteful to lethal depending on the compound's ability to release free cyanide which is a potent inhibitor of cytochrome-c-oxidase whilst some nitriles are mutagenic or carcinogenic (Banerjee *et al.*, 2002). This characteristic of the cyanoglycosides has been exploited in the production of anti-cancer drugs. Kousparou et. al. (Kousparou *et al.*, 2002) have created a novel anti-cancer treatment by combining the cyanoglycoside linamarin (as a harmless prodrug) with a fusion protein consisting of an immunoglobulin specific for a cancerous antigen (directing the enzyme to cancerous cells) and a beta-glucosidase,

linamarase, which hydrolyses linamarin resulting in the release of free cyanide at the site of the malignant cells.

Plants, fungi, insects and bacteria produce cyanoglycosides primarily as a defensive weapon and less commonly as a hormone (arthropods and plants). To confirm the repellant nature of these compounds Tattersall et. al. (Tattersall *et al.*, 2001) transferred a tyrosine-based cyanoglycoside synthesis pathway from the *Sorghum bicolor* plant to *Arabidopsis thaliana*. Resistance to the flea beetle *Phyllotreta nemorum*, a pest of the Brassicacaea family to which *Arabidopsis* belongs, was conferred to the new transgenic *A. thaliana*. The compound produced is known as 'dhurrin'. The major phytohormone indole acetic acid is synthesised from a nitrile intermediate, indole acetonitrile in the same plant family (Muller & Weiler, 2000). UV exposure has been shown to induce the production of a nitrile antibiotic effective against bacteria and various fungi in the fungus *Trichoderma harzianum* (Faull *et al.*, 1994). A species of caterpillar, *Calindoea trifascialis*, has the ability, when disturbed, to secrete a defensive substance consisting of, amongst other compounds, mandelonitrile which acts as an effective ant deterrent (Darling *et al.*, 2001). Bacteria have been an invaluable source of insight into antibiotics for the human race. *Micromonospora echinospora* is but one of a collection of bacteria that have been shown to produce nitrile-based antibiotics. *M. echinospora* produces a nitrile which acts against gram positive bacteria and drug-resistant strains of *Staphylococcus aureus* (Sugawara *et al.*, 1997).

Cyanide and related compounds are used extensively by humans in various industrial processes which, due to carelessness and inadequate waste-management systems, contribute significantly to the levels of toxic cyanide contamination in the environment. Coal is often degassed to produce what is known as 'town gas', a hydrogen/methane-rich substance used domestically in many countries for light and heat and to power gas-engines. The oxidised product is a gas rich in hydrogen sulphide and hydrogen cyanide which is purified by exposing it to bog iron ore to bind the cyanide in ferric metal complexes and thiocyanates. The disposal of the saturated ore is an obvious problem (Theis *et al.*, 1994). Cyanide salts are used in chemical baths in the electroplating

industry which can spill into the environment. Certain metal-cyanides are used as anti-clumping agents in road salts. These salts are used to de-ice roads in extremely cold conditions and the overlaying of roads and concomitant run-off exposes soil and water systems to potentially toxic cyanides. Probably the most extensive commercial use of cyanide is in the mining industry where metals such as gold are leached from ore and water-solubilised by binding to cyanide. This results in toxic collection pools (Kjeldsen, 1999; Baxter & Cummings, 2006; Akcil & Mudder, 2003). Some herbicides (e.g. bromoxynil) are nitrile-based and expectedly cause a range of negative symptoms upon prolonged exposure (Banerjee *et al.*, 2002). As of 2005 the Agency for Toxic Substances & Disease Registry ranked cyanide at 28th on its Priority List of Hazardous Substances and had documented 45 sites of known Completed Exposure Pathways which are areas of contamination linked to a receptor population (ATSDR, 2005b; ATSDR, 2005a). Taking these issues into consideration it is not difficult to see that there exists a myriad of potential applications for a nitrile-degrading enzyme.

Though Nitrile/cyanide hydrolysing enzymes have a broad range of substrates they function via four known pathways. Nitrilase and cyanide dihydratase completely hydrolyse nitriles and HCN respectively to yield the corresponding acid and ammonia without going via an amide intermediate. Nitrile hydratase and cyanide hydratase perform a single hydrolysis producing the corresponding amide and formamide, respectively. Generally these amides enter a subsequent amidase pathway releasing ammonia and converting the amide into its corresponding acid (Jandhyala *et al.*, 2003; O'Reilly & Turner, 2003) **(Fig. 1).** *(see also: 1.5 Structure and Catalysis)*

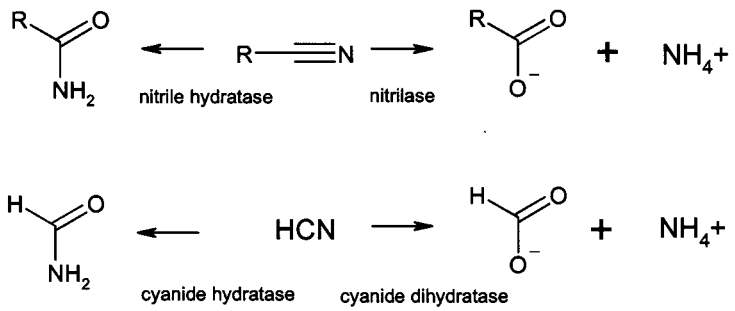

Figure 1. The various reactions of the nitrile-hydrolysing enzymes. Adapted from (O'Reilly & Turner, 2003).

Nitrilases, cyanide hydratases (CHT) and cyanide dihydratases (CynD) have been grouped into branch 1 of the "Nitrilase Superfamily" of nitrilase-related enzymes based on their sequence similarity. The family consists of 13 branches of which only members of the first branch, the nitrilases, function as classical nitrilases while eight branches exhibit amidase activity and several others are hybridised with other protein domains. All the proteins in this superfamily are distinguished by a conserved Glu-Lys-Cys catalytic triad which is absent in the Fe-S/heme-containing or Cobalt nitrile hydratases and thus regardless of its functional similarity to the cyanide hydratase it is excluded (Huang *et al.*, 1997). The nitrilase-related proteins also contain a conserved αββα-αββα sandwich fold and are generally multimeric (Pace & Brenner, 2001). These enzymes exhibit high substrate-specificity and enantioselectivity and do not require the harsh reaction conditions of traditional nitrile-converting protocols. These characteristics make the nitrilases good candidates for the pharmaceutical industry. Also, the fact that a number of bacteria produce nitrile-degrading enzymes makes them a viable alternative for bioremediation of many industrial and agricultural sites (Banerjee *et al.*, 2002).

1.2 Classification of the Nitrilases

Nitrilases are an important class of enzymes due to their fundamental role in cyanide hydrolysis of a wide range of substrates. Based on the sequence and structural similarity of 176 protein sequences Pace and Brenner (Pace & Brenner, 2001) classified the nitrilase enzymes into a superfamily with 13 branches/subfamilies. Branches were allotted according substrate-specificity of the enzyme and sequence similarity of the residues surrounding the conserved catalytic triad (esp. the residue adjacent the catalytic Cysteine residue on the C-terminal side). There are 4 kinds of reactions performed by the superfamily. The first is that performed by the nitrilases which hydrolyses a nitrile/cyanide to the corresponding acid and ammonia. This is performed in 2 hydrolysis steps via a thiomidate intermediate resulting in an acyl group attached to the enzyme which is subsequently released as the acid. The cyanide hydratases which only catalyse the first half of this reaction belong to the same subfamily. The second mechanism is the amidase reaction which involves the hydrolysis of an amide to a corresponding acid and ammonia. Thirdly, the carbamylase reaction, similarly to the amidase reaction, hydrolyses carbamides to an amine, ammonia and bicarbonate. The fourth and last mechanism is analogous to a reversed amidase reaction performed by the N-acyltransferases. This reaction attaches an acidic group to the N-terminal of a polypeptide.

Figure 2 shows a graphical representation of the 13 branches of the superfamily with domain fusions and residues adjacent the catalytic triad. Seven branches contain members with nitrilase domains fused to other domains fulfilling either a 'Rosetta Stone' function, whereby through a combined effort the domains fulfill a biochemical function or allowing for specific cellular localisation. An example of this fusion is the NAD^+ synthetase from *Mycobacterium tuberculosis*. NAD^+ synthetases catalyse the conversion of deamido-NAD to NAD^+ by attaching an amide group derived from glutamine or free ammonia. Pace et al. mention that *E. coli* NAD^+ synthetase has long been known not to be able to utilise glutamine as a nitrogen source whilst *M. tuberculosis* NAD^+ synthetase can. It was discovered that *M. tuberculosis* NAD^+ synthetase contained an N-terminal

nitrilase domain that *E. coli* NAD^+ synthetase lacked which transferred nitrogen by hydrolysing glutamine thus fulfilling the role of a glutamine amidotransferase (Bellinzoni *et al.*, 2005; Pace & Brenner, 2001). This finding confirms that protein domains can act in a modular fashion by contributing different functionalities.

The 13 branches are as follows (Pace & Brenner, 2001; Brenner, 2002):

Branch 1: Nitrilases

This is the true nitrilase branch inluding nitrilases, cyanide dihydratases and cyanide hydratases and is found in plants, animals, fungi and bacteria. These enzymes, first discovered in 1964, are involved in key processes such as the conversion of indole-3-acetonitrile to the very important phytohormone indole-3-acetic acid in plants. This process is catalysed by three similar nitrilases in *A. thaliana*: nit1, nit2 and nit3 which also convert 3-phenylpropionitrile to phenylpropionic acid (Vorwerk *et al.*, 2001). Whereas, the same plant uses nit4, a cyanoalanine hydratase, to hydrolyse an intermediate of cyanide detoxification, beta-cyano-L-alanine, to asparagine and aspartate (Piotrowski & Volmer, 2006; Piotrowski *et al.*, 2001). Generally the nitrilases fall into 3 categories depending on their preferred substrates: aromatic, aliphatic or arylacetonitrile.

Branch 2: Aliphatic amidases

This is the first group of 8 amidases. The members of this group are from bacteria and are specific for aliphatic amide substrates like asparagine and glutamine. Included in this branch is the aliphatic amidase from *Pseudomonas*.

Branch 3: Amino-terminal amidases

These amidases are very similar to Branch 2 but are specific for N-terminal residues. The N-end rule of ubiquitination determines the cellular life time of a protein before it is degraded by the fact that N-terminal residues differ in their affinity for ubiquitination. Pace and Brenner point out that the enzyme responsible for this in brewer's yeast, Nta1, belongs to this subfamily. N-terminal Asn and Gln residues are considered tertiary destabilising residues which, when hydrolysed to Asp and Glu by Nta1, become secondary destabilising residues. These signal a more urgent demand for protein degradation (Baker & Varshavsky, 1995).

Branch 4: Biotinidase

These enzymes have a preference for R-C=O(NHR') amines and include amongst others biotinidase, the vanins and GPI-80. Biotinidase is responsible for releasing biotin from a precursor molecule (e.g. biocytin → biotin + lysine). Vanins have an as of yet unknown substrate and both vanins and GPI proteins play an important role in the immune response. Interestingly, the domain fusions show that these enzymes have C-terminal domains that confer unique functional abilities much like the NAD^+ synthetases mentioned above, while one of the vanins even contains an additional partial N-terminal nitrilase domain **(Fig. 2).**

Note: branches 3,4 and 9 are the only branches able to act upon polypeptides.

Branch 5: β-ureidopropionases

These proteins are similar to the carbamylases and catalyse the conversion of N-carbamyl-beta-alanine to beta-alanine, a reaction important in the catabolism of pyrimidines (Rawls, Jr., 2006).

Branch 6: Carbamylases

These enzymes are important for the decarbamylation of D-amino acids which are not common to the protein alphabet of most life-forms but are used, for instance, extensively in bacterial peptidoglycan cell walls (van Heijenoort J., 2001).

Branch 7/8: Glutamine-dependent NAD^+ synthetases

Branches 7 and 8 represent an elegant example of combining simple biochemical functionalities, also known as "Rosetta Stone" fusions. The fusion of a Glutamine Amino Transferase domain to an NAD^+ synthetase domain in certain prokaryotes (branch 7) like *M. tuberculosis* and eukaryotes (branch 8) allows these essential catalysts to utilise glutamine as a source of ammonia/nitrogen.

Branch 9: Apolipoprotein N-acyltransferases

This branch is identified by a reversed amidase reaction transferring fatty acids to protein N-termini. Members of this group have been indicated to be important in various bacterial pathways especially flagellar assembly. FlgH, an essential structural protein in the flagellar subunit assembly of *Salmonella* species is a lipoprotein that requires modification by an apolipoprotein N-acyltransferase for correct functioning (Dailey & Macnab, 2002). These proteins have an N-terminal hydrophobic domain and occasionally a C-terminal dolichol phosphate mannose synthase domain which could be expected if this branch is primarily involved in modification of proteins, especially those found in cell walls.

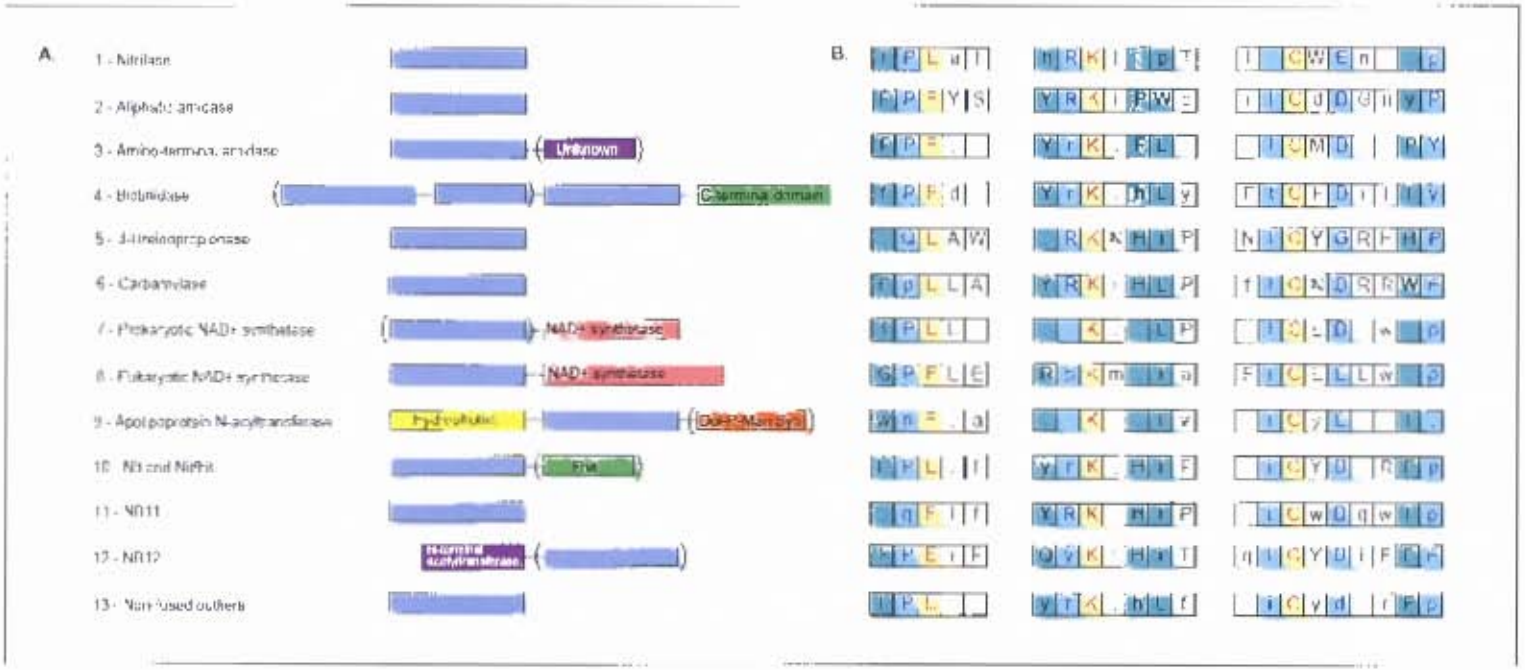

Figure 2. **A.** The 13 branches with domain-fusions. The nitrilase domain is represented as a purple block.; **B.** Adjacent residues to the catalytic triad residues. Uppercase letters represent 90 % or greater intrabranch consensus while lower case represent higher than 50 %. Interbranch consensus is depicted by a colour scheme: Red letters on yellow – conserved triad residues, dark blue letters on light blue – residues conserved in 9 or more branches, black letters on green – conserved in 6-8 branches (Pace & Brenner, 2001).

Branch 10: Nit

Nit is a plant/bacterial nitrilase homologue that was first found in the fly *Drosophila melanogaster* and nematode *Caenorhabditis elegans* as a fusion with the Fragile Histidine Triad protein (fhit), a mammalian tumour-suppressor protein (involved in cell death). Nit1 is found in mammals and is expressed separately to Fhit but they maintain similar expression patterns suggesting involvement in a similar pathway, a theory which is strengthened by the finding that these two proteins are fused in other organisms (Pekarsky *et al.*, 1998). The function of Nit in this context is as yet unknown.

Branches 11-13: The 'Sundries'

These branches were used to designate groups of proteins that have no type-member to identify them. Group 12 are however, distinctively fused to an N-terminal acyltransferase and branch 11 has as of yet 1 member that has characteristic properties, AguB. This enzyme from *Pseudomonas* is key in the metabolism of arginine to spermidine and is known as an N-carbamyl putrescine amidohydrolase (Nakada *et al.*, 2001).

1.3 Substrates, enantioselectivity and applications

In aid of identifying novel nitrilases and enantioselective trends amongst the nitrilases Robertson et al. (Robertson *et al.*, 2004) recently created a collection of cDNA libraries from 651 environmental metagenomic samples derived from soil and water systems including Southeast Asia, Central America, the Arctic, Antarctica and others. The resulting fragments were cloned and expressed in bacteria yielding up to 1 billion individual clones. These clones were put through a screening process to test for the ability to grow on selected nitriles as the sole nitrogen source. This process eventually yielded many putative nitrilases of which 137 were sequenced, expressed and underwent a phylogenetic analysis including 13 nitrilase sequences from plants, fungi and bacteria **(Fig.3)**.

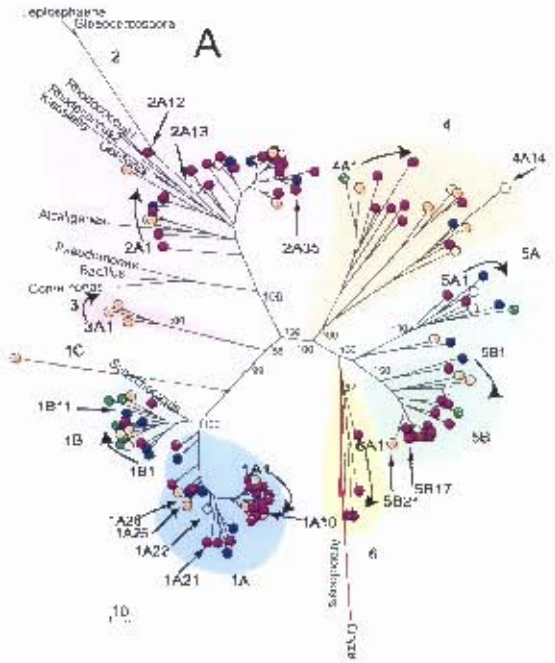

Figure 3. Maximum-likelihood phylogeny of 150 nitrilase sequences (137 novel, 9 bacterial, 2 plant, 2 fungal). Clades are sequestered by shading. Sequences are represented by a dot which is coloured according to geographical source: Southeast Asia - magenta, Central America – orange, Arctic – blue, Antarctica – lilac, Temperate regions – green, Other – white. The scale bar represents an expected number of substitutions in the protein sequence (Robertson *et al.*, 2004).

The sequences divided into 6 definite clades having an intraclade sequence identity of 75% with the 9 known bacterial sequences falling into clades 1 and 2. The various geographical locations yielded inconsistent sequence data. Some nitrilases were vastly different within the same clade (e.g. 2A13 and 4A14) while a few very similar nitrilases could be found in different clades (e.g. the spectrum of geographic locations in clade 1B) **(Fig. 3)**. Robertson et al. also postulate that the tendency for tropical samples to form clusters within clades identifies a possible series of accelerated evolution events afforded by the many microenvironments hospitable to microbial life in this type of ecosystem.

Nitrilases were then analysed to determine their substrate specificity and enantioselectivity by monitoring the various enzymes' abilities to hydrolyse 3 industrially important synthetic intermediates: 3-hydroxyglutaronitrile, an intermediate in the synthesis of the 3-hydroxy-3-methyl-glutaryl-CoA reductase inhibitor, cholesterol-reducing drug Lipitor (Atorvastatin) (DeSantis *et al.*, 2002); mandelonitrile, which hydrolyses into the two useful enantiomers of mandelic acid used in the synthesis of thrombin inhibitors amongst others (Inghardt *et al.*, 2002); and phenylacetaldehyde cyanohydrin, which also produces two valuable aromatic compounds. Clades 1A/B and 4 were especially selective for the hydrolysis of 3-hydroxyglutaronitrile to the S-enantiomer product, 4-cyano-3-hydroxybutiric acid. There remained a small cluster of enzymes in clade 1A specific for the R-enantiomer product. Whilst only ~35 % of the enzymes were reactive with mandelonitrile, these were primarily found in clade 2 and were almost all S-enantioselective. 50% of the enzymes were active on phenylacetaldehyde cyanohydrin with most of them being found in clade 1A/B and some having a representation in clade 4. Also, almost all of these representatives were S-specific. Conclusively, sequence similarity and enantioselectivity are shown to be correlated and the nitrilases have a wide substrate range. Also, due to the extraordinary diversity in the microbial populations caused by rapid evolutionary explosions, there exists a vast and potentially untapped resource for novel enzymes.

Microbes are known to produce many nitrilases and thus are able to colonise areas using nitriles as their sole nitrogen source. Indeed, nitrilases have been identified and

characterised in several species of bacteria including some key species: *Rhodococcus rhodochrous* which shows a very broad substrate range with preference for aromatic and aliphatic nitriles; *Bacillus pallidus*, showing preference for aromatic nitriles; *Pseudomonas fluorescens* which prefers arylacetonitriles; *Nocardia sp.*, preferring aromatic nitriles. Cyanide dihydratases have been found in *Alcaligenes xylosoxidans*, *Pseudomonas fluorescens*, *Pseudomonas stutzeri* and *Bacillus pumilus* (O'Reilly & Turner, 2003; Pace & Brenner, 2001). Fungi also express nitrilases (*Fusarium solani, Fusarium lateritium*) and cyanide hydratases (*Fusarium solani, Fusarium lateritium, Gloeocercospora sorghi, Trichoderma harzianum)* (Baxter & Cummings, 2006).

In a recent screen for novel nitrile-hydrolysing microorganisms, a large number of bacterial and yeast strains were isolated from environmental and industrial areas in South Africa (Brady *et al.*, 2006). 150 bacterial strains were identified as nitrilase-producing by their ability to grow on 3-hydroxy-3-phenylglycinonitrile. This number was reduced to 40 by further screening for the ability to grow on 3-hydroxy-3-phenylpropionitrile. These 40 strains were then submitted to further screening for a large range of substrate specificities by allowing them to grow on various nitrile-containing media. Most of the enzymes came from *Bacillus*, *Pseudomonas* and *Rhodococcus* genera with the latter producing the widest substrate specificity.

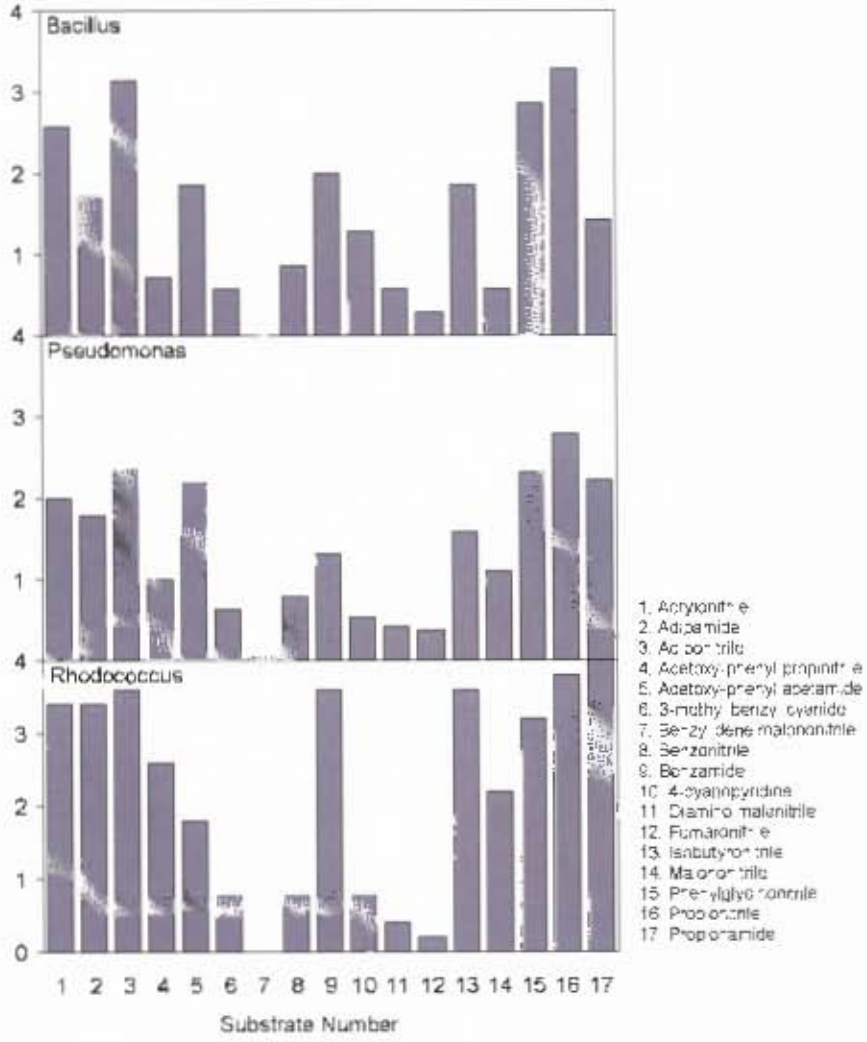

Figure 4. Average values of the ability of 3 bacterial genera to grow on a range of nitriles as their sole nitrogen source as presented in Table 1 of (Brady *et al.*, 2006). Growth values are represented as: 0 – no growth, 1 – poor growth, 2 – moderate growth, 3 – good growth, 4 strong growth. Note: where substrates weren't tested they were replaced with a value of 0.

Figure 4 shows a comparison of the abilities of each genus to grow on a range of nitrile substrates. Clear preference can be seen in all three genera for aliphatic nitriles (substrates 1-3), phenylglycinonitrile (15) and propionitrile (16). The *Rhodococci* showed an additional preference for benzamide (9), isobutyronitrile (13) and propionamide (17). Almost none of the strains could utilise benzylidene malonitrile (7) and fumaronitrile (12). The specificity of a series of bacterial whole-cell extracts was tested against a number of industrial enzymes. Of all the strains *Rhodococcus rhodochrous sp. 4* was shown to have the widest range of substrate specificity.

Due to the wide range of substrates accessible by the nitrilase superfamily and their unique enantioselectivity, they have become important players in the biocatalysis industry. Many existent and potentially novel synthetic applications exist for these enzymes as they also catalyse the hydrolysis of nitriles with fewer steps and under far less extreme reaction conditions than are currently being used in the fine chemicals industry (Mylerova & Martinkova, 2003). Some of the downsides are protein intolerance of high reactant concentrations and temperatures. Also, for the wide spectrum of substrates that these enzymes can hydrolyse, this is still a limiting factor when compared to the extensive substrate range of alkaline/acid hydrolysis

Enzymes are restricted in their choice of substrate primarily by the size and shape of the active site pocket into which the substrate docks. This has been studied in *Rhodococcus* biocatalysis systems and it was discovered that all the substrates unavailable to the specific strain were either larger than the pocket (228 $Å^3$) or imposed an intractable steric hindrance. Also, enzymes have narrow temperature/pH ranges which are gradually being overcome by the isolation of thermostable (>50° C) and wide pH range (2-13) strains of *Rhodococcus* (Mylerova & Martinkova, 2003).

One of the logistical issues facing scientists is the development of suitable mechanisms for the application of microbes or microbial enzymes to environmental issues and industrial processes. Do we use whole cell cultures, cell extracts or purified enzymes? Many techniques have been developed and been shown to be successful. Generally,

whole-cell systems are preferred due to the ease with which they are produced and the high cost of purifying enzymes for commercial use. Gel-immobilised cell systems have shown themselves to be successful. Common gel substrates include polyacrylamide, Ca/Br alginate, K-carrageenan, pectate and polyvinyl alcohol (Mylerova & Martinkova, 2003). Rhodococcus has been used in this manner to degrade acrylamide by immobilisation on polyacrylamide gels (Nawaz *et al.*, 1998). *Rhodococcus* has also been attached to other substrates: porous carbon (Maksimov *et al.*, 2007), polyurethane foams (Quek *et al.*, 2006) and silicone polymers (Roach *et al.*, 2004) amongst others. Another possibility is using lyophilised cells attached to a matrix as a biocatalyst. This has also been shown to be successful, with the addition of necessary lyoprotectants, in the case of *Rhodococcus* (Mylerova & Martinkova, 2003).

Both aerobic and anaerobic batch culture/reactor systems have been utilised extensively. Some simple approaches use a produced biomass to 'mop up' cyanide and metal cyanides with a higher affinity than activated charcoal. More complex systems involve suspended solution and fixed film reactors for the treatment of toxic liquids by hydrolysis of nitriles. A system employed at the Yellow Pine Mine site in Idaho, USA, involved isolating and enriching nitrile-utilising bacteria native to the site and redeploying the enriched cultures into the spent ore and tailing pond to leach out toxic metal-cyanides (Baxter & Cummings, 2006).

Due to the beneficial characteristics of the nitrilases and nitrilase-producing organisms a selection of novel biocatalytic applications can be devised for the enhancement of fine chemical synthesis by reducing the number of chemical intermediates necessary and removing the need for harsh reaction conditions. Also, these enzymes can be used for the protection of the environment through the bioremediation of toxic sites produced by various industries, not least of which is the mining community.

1.4 The genetics and evolution of the microbial nitrilases

In response to Robertson et al.'s work in classifying a collection of microbial nitrilase sequences from cDNA libraries into 6 distinct clades (Robertson *et al.*, 2004) (cf. 1.3, para. 1), Podar et al. went further and compared the genetic neighbourhood of some of these sequences to their enzymatic activity (Podar *et al.*, 2005). Originally, Robertson et al. classified 137 novel nitrilase sequences and 9 known bacterial sequences. With the release of certain genome sequences Podar et al. were able to include an additional 9 nitrilases from cultivated bacteria and subject the whole dataset, once again, to phylogenetic analysis, rooting the tree with 3 plant and 4 fungal sequences. As seen previously, all the cultivated bacterial sequences (now a total of 18) fell into clades 1 and 2 with up to a 70 % and 40 % identity respectively **(Fig. 5)**.

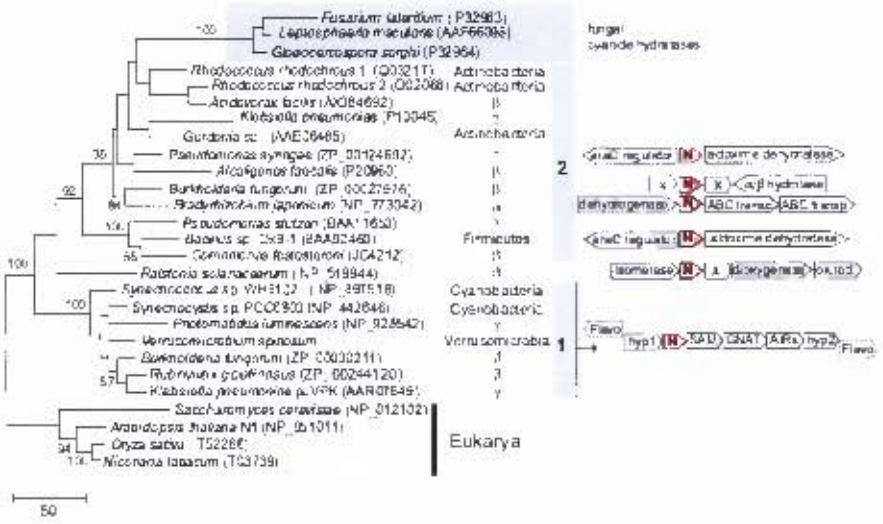

Figure 5. Maximum likelihood tree of 18 bacterial nitrilase sequences, 3 plant sequences and 4 fungal sequences from Podar et al. Phyla, subfamily and known gene clusters are shown to the right of the tree in that order for the bacterial sequences. (Podar *et al.*, 2005).

It is known that in the prokaryotes genes can be organised into 'clusters' or 'neighbourhoods'. These are orthologous collections of genes with related function (and their neighbours) loosely conserved between various organisms and are considered to be of higher order yet very similar and sometimes inclusive of operons/regulons in which genes with very strongly related functions are grouped under the same transcriptional control mechanisms (Rogozin *et al.*, 2002). Thus, between different organisms there might be some genes missing or the order of the genes might differ. In this study, all of the subfamily 1 genes were found to belong to the Nit1C gene cluster. This cluster was hypothesised to be an operon due to the fact that 6 of the genes are found on the same DNA strand and are transcribed in the same direction. Only 4 clusters were available for subfamily 2. Of these, 2 were identical and encoded the AraC regulatory protein but are not organised on the same strand **(Fig. 5)**.

1.5 Structure and Catalysis

The two initial crystal structures of the nitrilase family which have formed the basis for many assumptions about the rest of the superfamily are the NitFhit protein from the nematode *C. elegans* and *N*-Carbamoyl-D-amino Acid Amidohydrolase (carbamylase) from *Agrobacterium.*

NitFhit

The Fhit protein is a tumour-supressor in humans and its absence is a major contributor to the acquisition of lung cancer. It is expressed as a single domain peptide in mammals and fungi and as two domains in flies (*D. melanogaster*) and worms (*C. elegans*) with Fhit on the carboxy terminal side. The N-terminal domain is a Nit protein with 22 % sequence similarity to plant/bacterial nitrilases and 48 % similarity to the mammalian Nit orthologues which are expressed separately to Fhit. These two proteins have been implicated in a combined function and indeed in mice their mRNA expression patterns have been shown to be similar (Pekarsky *et al.*, 1998).

NitFhit is a 440 amino acid protein with a molecular weight of ~50 kDa but on a column the native eluent has a molecular weight of ~200 kDa indicating that it is a tetramer. The N-terminal Nit domain consists of 5 α-helices and 13 β-strands which are arranged in an αββα sandwich fashion. This is composed of 2 α-helices, a sheet of 6 β-strands, another sheet of 6 β-strands and closes off with another 2 α-helices. These Nit domains are arranged into a tetramer as shown in **Figure 6** via the interaction of 2 interfaces. The A-interface is considered to be the standard interaction forming 2 dimers and is mediated by the two proximal helices 3 and 4 and their symmetry-equivalents. Thus, in the dimer the sandwich fold is extended to become an 8-layered α-β-β-α-α-β-β-α fold. Strands 11 and 12, with their symmetry-equivalents in an antiparallel conformation, form the B-interface for the interaction of 2 dimers (Sewell *et al.*, 2005). This extends the width of the β-sheet portion of the sandwich to 12 strands. The active site, as with all members of the superfamily, consists of a Glu-Lys-Cys catalytic triad. Glu54 is found immediately after

the C-terminal end of β-strand 2 (NS2), Lys127 in the same position relative to NS5a and also Cys169 relative to NS7 (Pace *et al.*, 2000).

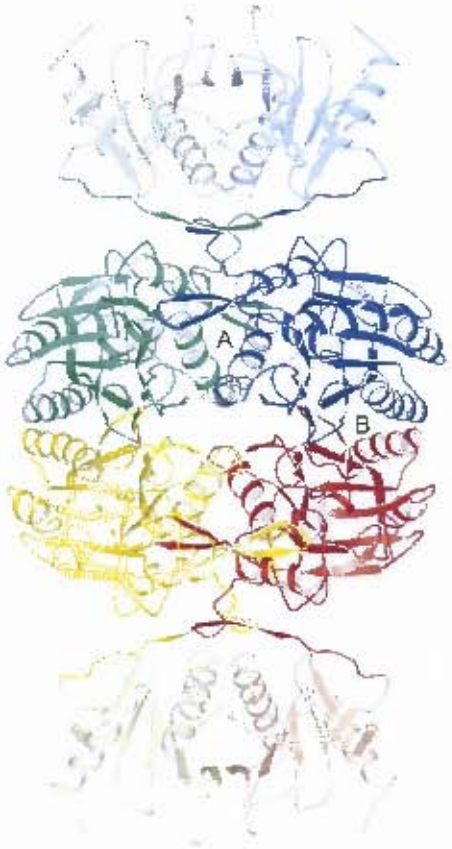

Figure 6. The tetrameric structure of Nit-Fhit. A and B interfaces are indicated (Pace *et al.*, 2000).

N-carbamoyl-D-Amino acid Amidohydrolase

D-NCAase or DCase has been found in several bacteria and was crystallised from *Agrobacterium.* This enzyme catalyses the hydrolysis of *N*-carbamoyl-D-Amino acids to D-amino acids; an important constituent of bacterial cell wall peptidoglycan. Like Nit, DCase also exists as a tetramer in the crystal structure but is most likely a dimer/trimer in nature according to biochemical assays. The conserved α-β-β-α sandwich fold and the catalytic triad (Glu46-Lys126-Cy171) are retained. The A-interface, between members of the dimer, is mediated by α-helices 5/6 (which, with the exception of certain insertions/deletions, are identical to helices 3/4 in Nit) and the C-terminal tail which interacts with NH5 on the opposite monomer (Nakai *et al.*, 2000).

Catalytic triad

The C-E-K catalytic triad is well conserved amongst the nitrilases and is used as the primary criterion for entry into the nitrilase superfamily (Pace & Brenner, 2001). Nakai et al.'s catalytic mechanism formed the basis for understanding catalysis in the nitrilase family. According to the basic mechanism, Glu46 acts as a general base, extracting a proton from Cys171 and activating it as a nucleophile either directly or via an abridging water molecule. The thiol group is then able to make a nucleophillic attack on the carbonyl/amide carbon the substrate (in this case an *N*-carbamoyl-D-amino acid) forming a tetrahedral intermediate. The transiently negative carbonyl oxygen interacts with the positive amide group of Lys126 which stabilises the complex. After this, the same carbonyl oxygen resumes its double bond nature forcing the amine group to adopt the excess proton from Glu46 and leave as ammonia to be replaced by a water molecule resulting in the acyl-enzyme intermediate. Once again, Glu46 acts as a general base and activates the water molecule which subsequently performs an attack on the carbonyl carbon forming a second tetrahedral intermediate which is released as an *N*-carboxy-amino acid which eventually collapses to the D-amino acid (Nakai *et al.*, 2000).

Jandhyala et al. add that a similar mechanism takes place in the Cyanide dihydratases and Cyanide hydratases (Jandhyala *et al.*, 2005). Initially, the nucleophilic attack on the amide carbon by cysteine causes the cyanide nitrogen to adopt a proton from lysine forming the thioimidate intermediate. Lysine then replaces its lost proton by attacking and activating a water molecule which subsequently attacks the amide carbon forming a tetrahedral intermediate. At this junction either the C-S bond is broken releasing formamide, as in the Cyanide hydratases or the C-N bond is broken releasing ammonia as in the case of the Cyanide dihydratases. Once ammonia has been released, the pathway is identical to the nitrilase mentioned above including nucleophillic attack by an activated water molecule and eventually the release of formate.

After the structure of DCase was solved by Nakai et al., Chen et al. performed a series of mutational studies to determine the specifics of binding and catalysis (Chen *et al.*, 2003). It is known from the crystal structure that the substrate is bound in a pocket constituted of 4 loops each harbouring a number of key residues that interact with the substrate. The most striking mutation performed was that of mutating the catalytic cysteine (C172) to a serine (C172S) and exposed the enzyme to *N*-carbamoyl-D-*p*-hydroxyphenylglycine (HPG) for crystallization with a substrate. This mutation reduced the enzyme's activity to less than 0.1 %. Eleven interactions were identified between the substrate carboxyl group and the pocket, five of which were hydrogen bonds. A key finding was that Arg^{175} and Arg^{176} play an important role in providing substrate specificity for fitting of carboxyl-containing substrates. Asn^{173} was also implicated in correct docking of the substrate by blocking part of the outer binding pocket with its large side chain helping to select for D-enantiomer substrates. Fourteen interactions were found between the pocket and the substrate carbamoyl moiety, seven of which were hydrogen bonds identified in the C172S structure. On the floor of the binding pocket S172 was found at the C-terminal side of beta sheet B7 with its hydroxyl group sticking into the pocket pointing directly at the carbamoyl carbon of the substrate. Glu146 hydrogen bonds with the substrate carbamoyl group and in conjunction with His144 represent the beginning of an H-bond network in the pocket which serves to fix the geometry of Lys127 **(Fig. 7)**.

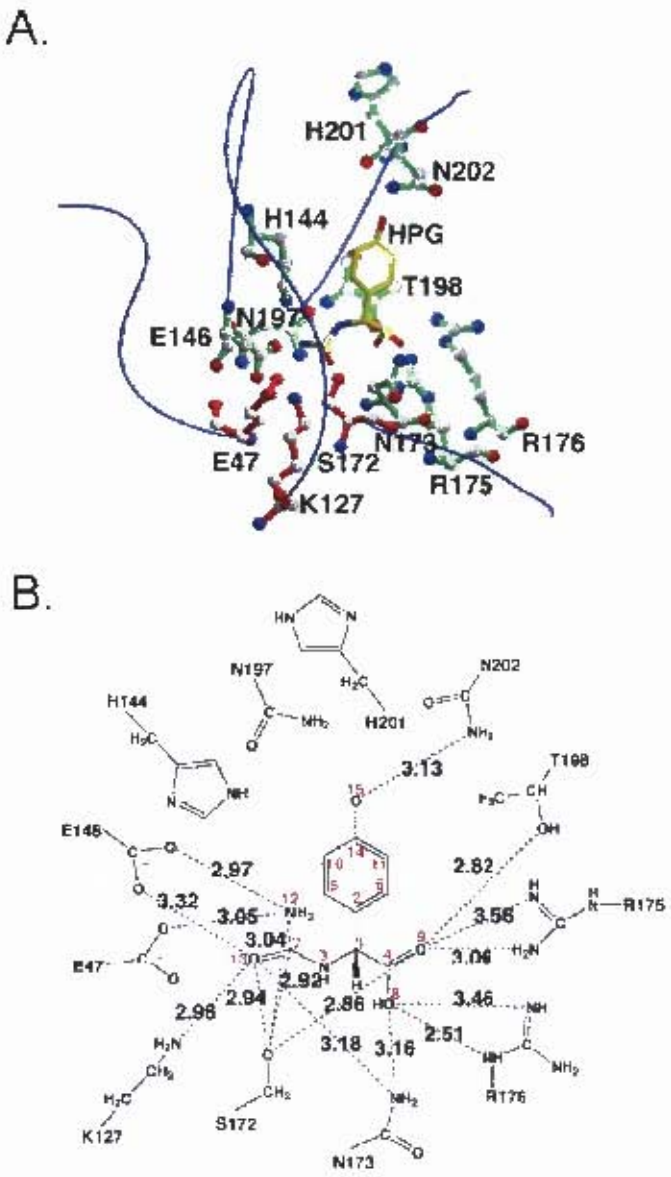

Figure 7. The C172S mutant of D-NCAase. **A.** The substrate (HPG) bound in the active site. Constituent loops are represented as blue ribbons. **B.** Line-drawing of HPG interacting with key residues in the binding pocket (Chen *et al.*, 2003).

Quaternary structure

The nitrilases are known to form extensive quaternary structures including dimers, spirals and rods/helices. Generally microbial nitrilases exist as homo-oligomers having a large molecular weight (> 300 kDa) (Scheffer, 2006). The cyanide dihydratases from *P. stutzeri* and *B. pumilus* C1 are known to form terminating spirals of length 14 and 16/18 subunits respectively (Jandhyala *et al.*, 2003; Sewell *et al.*, 2003).

A cyanide dihydratase ($CynD_{pum}$) was isolated from the bacterium *B. pumilus* in 1993 (Meyers *et al.*, 1993).

Berman produced the two initial negative-stain reconstructions of *B. pumilus* CynD. The first reconstruction being of the recombinant $CynD_{pum}$ (expressed in *E. coli*) at pH 6 and the second being the native $CynD_{pum}$ at pH 8 (already mentioned). Both structures were reconstructed with a global dyad symmetry imposition based on assumptions about the assembly of the spirals (Berman, 2003).

The pH 6 structure was shown at a resolution of 3.2 nm to consist of a 16-subunit terminating spiral. A terminal region of partial density attached to both ends of the spiral was noticed. Berman postulated that this indicates partial occupancy by an extra pair of terminal subunits which was challenged by Scheffer *(see later)*. The radius was shown to decrease from 5.3 nm at the centre to 4.6 nm at the end thus inducing termination.

The native pH 8 structure was solved to a resolution of 3.4 nm and consisted of a similar terminating spiral of 18-subunits with identical symmetry to the recombinant form. With a central radius of 6.45 nm this structure is wider with a more prominent extrusion of density on the outside surface.

The structure of the enzyme from the C1 strain was further refined for publication (Jandhyala *et al.*, 2003) and was shown by negative stain electron microscopy to be a spiral of 18 dimers at pH 8 with a length and width of 18.5 nm and 9.5 nm respectively at

a resolution of 3.2 nm **(Fig. 8)**. The spirals exhibit twofold symmetry about the A-interface of the central dimer with pseudo-dyad symmetry in the centre of each dimer. These spirals are found only above pH 6 below which (~pH 5.4) the enzyme associates to form extended helices with left-handedness shown by unidirectional shadowing.

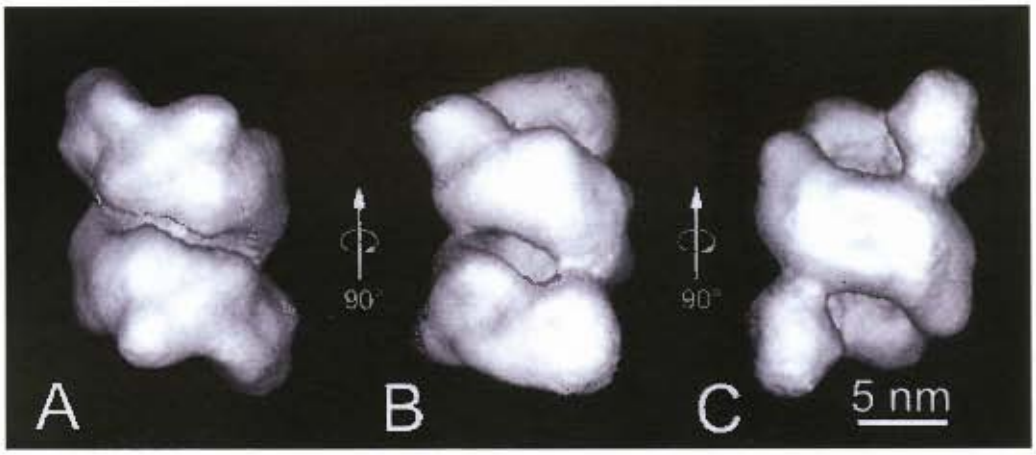

Figure 8. Single particle reconstruction of CynDpum at a resolution of 3.2 nm. The three views show 90 ° rotations. A,C show the rear and front views of the twofold axis of symmetry respectively (i.e. C shows the A interface) (Jandhyala *et al.*, 2003).

The enzyme has recently been characterised (Jandhyala *et al.*, 2005). Optimal pH for $CynD_{pum}$ was shown to be around 7-8 with a sharp falloff in activity above 8. Interestingly, the falloff below 7 showed a 'kink' in the profile representing a slight increase in activity as opposed to the steady rate of decrease found in the CynD of *P. stutzeri* ($CynD_{stu}$) and the CHT of *G. soyghi*. This has been suggested to be as a result of multimerisation of the enzyme which is known to happen at a pH of 5-6 in the C1 strain of *B. pumilus* (Jandhyala *et al.*, 2003). In fact, it is known that the nitrilases from *R. rhodocrous* J1 and NCIB11216 exhibit an increased activity with certain substrates upon oligomerisation and Jandhyala et al. therefore speculate that the inclusion of the terminal subunits in helix formation locks them into an active conformation identical to the other subunits (Jandhyala *et al.*, 2005).

$CynD_{stu}$ has been shown by the same technique to be a 14-subunit spiral with the same twofold symmetry and a length of 18 nm and diameter of 12 nm (Sewell *et al.*, 2003). The radius decreases near the ends by about 1 nm. To explain this phenomenon, the technique of cylindrical projection allowed angular displacements of subsequent subunits in the spiral structure to be determined. The projection in **figure 9** clearly shows that subsequent pseudo-dyad axes have larger angular displacements and are thus closer to the helical axis which accounts for the radial decrease upon termination in the spiral.

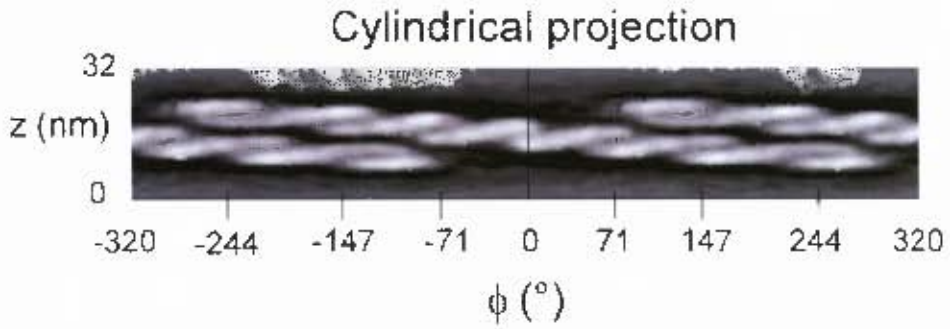

Figure 9. Cylindrical projection of the cyanide dihydratase from *P. stutzeri* spanning a rotation of 640 °. Angular displacement (φ) is plotted against rise (z). The global dyad axis is marked in the center at an angular displacement of 0 ° and subsequent pseudo-dyad axes (A-surfaces) are indicated along the x-axis (Sewell *et al.*, 2003).

Thus Sewell et al. postulate that there are two determining factors in spiral termination/extension: the decreasing terminal diameter of the spiral by about 2 nm near the ends and the concomitant increase in subunit tilt by 12° inwards toward the helical axis. These two factors serve to close up the spiral preventing the addition of more subunits. The terminating interaction is created by the last dimer tilting in toward the helical axis and forming the E-surface *(see below)*.

Also, it is generally accepted that around 10 subunits (5 dimers) are required to complete a single turn after which dimers are added simultaneously to opposite ends and thus increasing the subunit number by increments of 4 at any one time. Thus we would expect structures of 10, 14, 18, 22 etc subunits and not 12, 16, 20. This assumption can be

considered a guideline at best based on the 16 subunit reconstruction by Berman (Berman, 2003), Scheffer's theory about combined models (Scheffer, 2006) and data presented later in this thesis.

Scheffer proposed two possible explanations for the observed monomeric terminal densities in Berman's 16 subunit pH 6 model. The conclusion was made that the presence of these densities was most likely due to flexibility in the terminal dimer. The alternative hypothesis being that Berman's reconstruction was performed on a heterogeneous sample of 14- and 18-subunit particles producing a perceived 16-subunit density. Indeed, docking of a homology model to the density indicates that the optimal placement of the terminal dimer is in fact in a helical orientation with respect to its previous partner. This is not unexpected bearing in mind the tendency of $CynD_{pum}$ to form extended fibres. $CynD_{hu}$ which has a well defined termination forming the E-interaction does not exhibit this orientation. Thus, $CynD_{pum}$ appears to have a weaker terminating interaction which favours extension (**Fig. 10**).

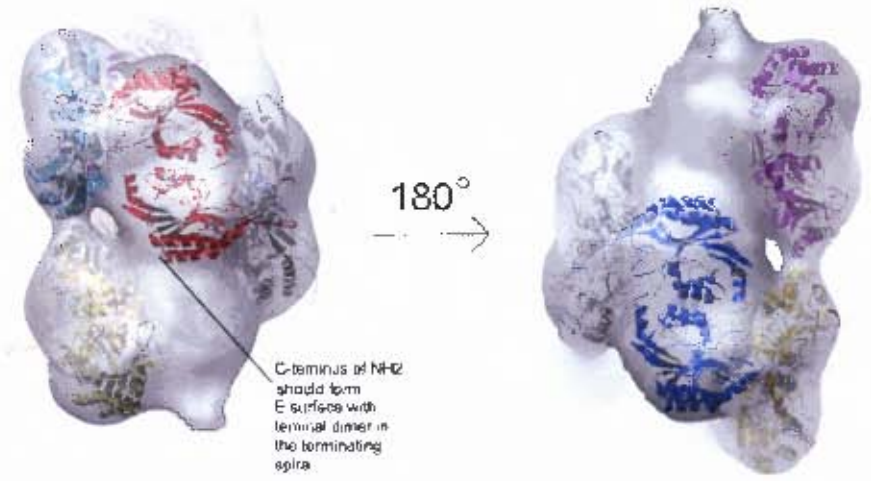

Figure 10. Docked model of '16-subunit' *B. pumilus* CynD showing terminal density. Adapted from (Scheffer, 2006).

3D reconstructions of nitrilases from *B. pumilus, P. stutzeri* and *G. sorghi* have allowed annotation of important contact surfaces (Sewell *et al.*, 2005; Scheffer, 2006):

(refer to sequence alignment: ***fig. 15)***

Several mutation experiments have been performed to identify important residues involved in oligomerisation and to elucidate whether this form of association is in fact necessary for activity *(**Table 1** for all mutations).*

A-surface

This surface represents the dimerisation interface and the global axis of symmetry of known spirals, consisting of helices NH3/4 and includes the C-terminal extension seen in $CynD_{pum}$ and $CynD_{stu}$.

C-surface

Allowing the sequential interaction of dimers along the spiral, this surface consists of residues found in two insertions at the N-terminal of NH2 and on the loop joining NS9/10 that are not present in crystal structures of non-extending nitrilases. An NS9/10-loop deletion mutant (mutant 5, Δ219-233) shows no activity which immediately lends weight to its suspected involvement in oligomerisation and the enhancement of activity upon oligomerisation. Additionally, crystal structures of the amidase enzyme show extended C-terminal arms of two monomers locking in with each other to form the dimer. This phenomenon is clearly demonstrated by the Amidase from *Geobacillus pallidus* (**Fig. 11**). The 'arm' wraps around the side of the opposite monomer forming a short α-helix before folding back upon itself and reentering the A-interface or dimerisation surface. This extension stretches as far as V226 in the opposite monomer of the amidase which would be the equivalent of M220 in $CynD_{pum}$, the beginning of the second unique insertion sequence (between NS9/10) constituting the C-surface. The similarity of the amidases to the cyanide dihydratases *(elaborated below: 1.6 The Amidase)* suggests that the C-terminal extension

performs a similar function and thus not only plays an important role in dimerisation at the A-surface but also oligomerisation at the C-surface (Agarkar *et al.*, 2006).

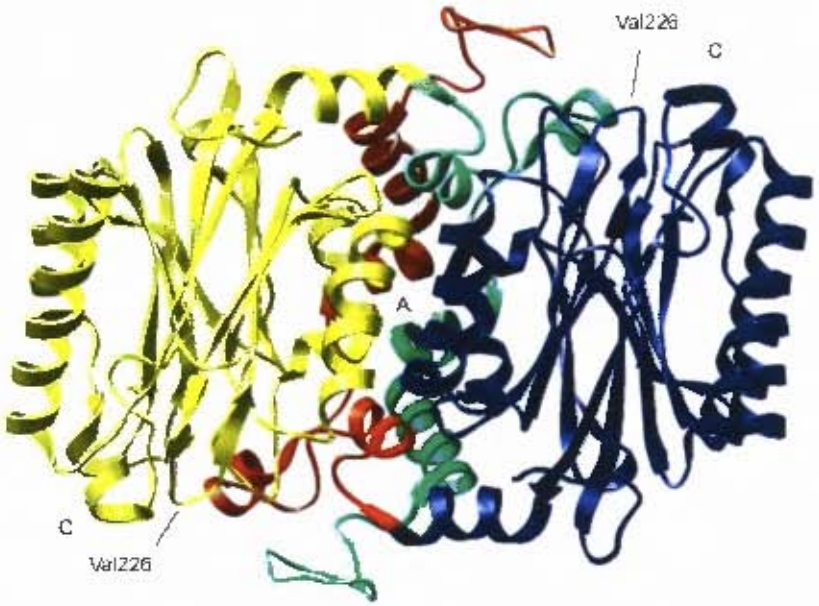

Figure 11. Crystal structure of Amidase from *Geobacillus pallidus* RapC8. Monomers are shown in yellow and blue with c-terminal extensions in red and green respectively. A and C surfaces are annotated.

D-surface

An interaction noticed as forming across the groove of the spiral in CynD. This surface is implicated between the first and sixth dimers of a single turn. This relationship is likely mediated by an interaction between E90 and R94 (NH3) or E90 and K93 with their symmetry partners on the opposite dimer (**Fig. 14**). Two interesting mutations were recently created by Van der Vyver in our lab in an attempt to disrupt the D-surface and thus hypothetically implicate it as the primary stabilising interaction in the multimeric forms of CynD ***(refer to Table 1)***.

$CynD_{stu}$ remains in the form of a terminating spiral regardless of pH. To test the hypothesis that the D-surface interaction is the primary stabilising force in the extended fibrous form of the nitrilases, the D-surface of $CynD_{pum}$ was replaced with the equivalent sequence of $CynD_{stu}$ (mutant 6). Contrary to expectations, the new mutant retained its ability to form fibres below pH 6 and counter intuitively showed an inclination toward short fibre formation at higher pHs too (**Fig. 12**).

pH 5.4 **pH 6** **pH 8**

Figure 12. Negative stain electron micrographs of $CynD_{pum}$ MB3799 mutant at three different pH values. Unpublished; courtesy of Miss Leandri van der Vyver.

It is also commonly expected that the helix adjacent to NH3, NH1, contributes to the D-surface. A glutamate residue (E31) was identified as having the potential to form salt bridge across the groove and mutant 7 was created from mutant 6 by additionally mutating this residue to a lysine (identical to $CynD_{stu}$). Having opposite charge the new mutant would thus be expected to repel any available like-charged side chain provided by a participating symmetry-related subunit. Electron micrographs of the mutant at pH 5.4 revealed fibre formation. Moreover, the mutant was seen to form even longer fibres than the previous mutant above pH 6 (**Fig. 13**).

A number of residues are suspected to contribute to this interface: On NH1 23E, 34D and 35E; 72E and 75K are found on NH2 which is adjacent but more likely to form part of the C-surface; E79 is a very likely candidate for a salt-bridge and is situated on the loop between NH2 and NH3; 87K in NH3 should also be investigated though it shows no possible partner.

pH 5.4 **pH 6** **pH 8**

Figure 13. Negative stain electron micrographs of $CynD_{pam}$ MB3799-2 mutant at three different pH values. Unpublished; courtesy of Miss Leandri van der Vyver.

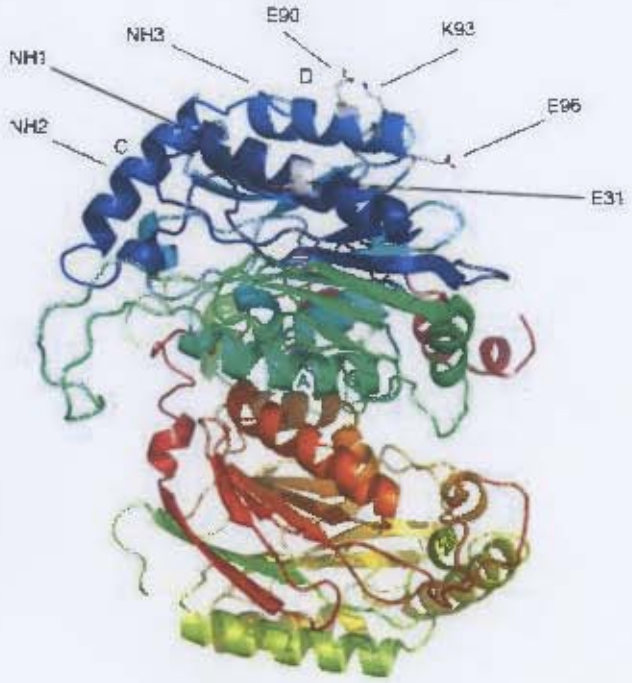

Figure 14. $CynD_{pum}$ homology model. Key D-surface residues, surfaces and secondary structure elements are indicated.

E-surface

This surface has been implicated in $CynD_{Nit}$ as being responsible for terminating the spiral. The interaction is thought to be mediated by residues found in NH2 on one side of the spiral and in NS12 on the other: residues E266 and D268 (**Fig. 10**) in the 12th β-strand potentially form a salt bridge with the alternate half of a D-surface interaction across the groove (i.e. residues 92**RKNK**95 on the C-terminal end of α-helix 2) and thus formation of the E-surface results in the ipso facto exclusion of terminal subunit addition which requires new D-surfaces.

C-terminal extension

The C-terminal extension (beginning at residue 291) is one of the insertions seen in CynD which is absent in non-extending nitrilases, potentially implicating it in

fibre formation. Mutations 1-3 served to remove increasing lengths of the C-terminus and test for activity. It is shown by these mutations that after losing the residues C-terminal of 293 (out of 330) $CynD_{pum}$ begins to lose activity and is completely inactivated by removal of residues after 279. As the relationship between activity and fibre formation has yet to be completely elucidated it is too early to draw a definite conclusion as to the role of the C-terminal extension while structural modelling data implicates it in the formation of the C and A surfaces.

Table 1: Mutations of the Cyanide Dihydratase from *B. pumilus* and their effects on activity

	Mutant	Surface	Change and location	Activity
1	Δ303	A	V303gtg → stop	Full activity
2	Δ293	A	M293atg → stop	Partial activity
3	Δ279	A	Y279tat → stop	Inactive
4	Y201D/A204D	A	Y201tat → Dgac A204gcg → Dgac	Inactive
5	Δ219-233	C	MKEMICLTQEQRDYF deleted E235gaa → Naac	Inactive
6	MB3799	D	90EAAKRNE → AAARKNK	Full activity
7	MB3799-2[b]	D	90EAAKRNE → AAARKNK E31K	Full activity

(Sewell *et al.*, 2005)(b. pers. comm. Miss Leandri van der Vyver)

Figure 15. Sequence alignment of cyanide dihydratases from B. pumilus C1, P. stutzeri and four homologues: 1f89, Nit3 from *S. cerevisiae*; 1j31, PH0642 from *Pyrococcus horikoshii*, 1ems, NitFhit from *C. elegans*; DCase from *Agrobacterium*. Secondary structural components are according to 1ems and are indicated below the alignment. Interaction surfaces are represented above the alignment and are also positioned according to 1ems. Key residues for interacting surfaces are highlighted in black, catalytic triad residues in grey (Sewell *et al.*, 2005).

Scheffer produced two reconstructions of the cyanide dihydratase fibres from *B. pumilus* at pH 5.4 (Scheffer, 2006). These were performed with 3701 negative-stain images and used two different starting models: a cylinder and a model of the cyanide hydratase from *G. sorghi*. The models converged at a resolution of 19Å, determined by generating a Fourier Shell Correlation between the two models and using the '0.5 cut-off criterion'. The fibres were unequivocally shown to have one-start left-handedness by a shadowing experiment.

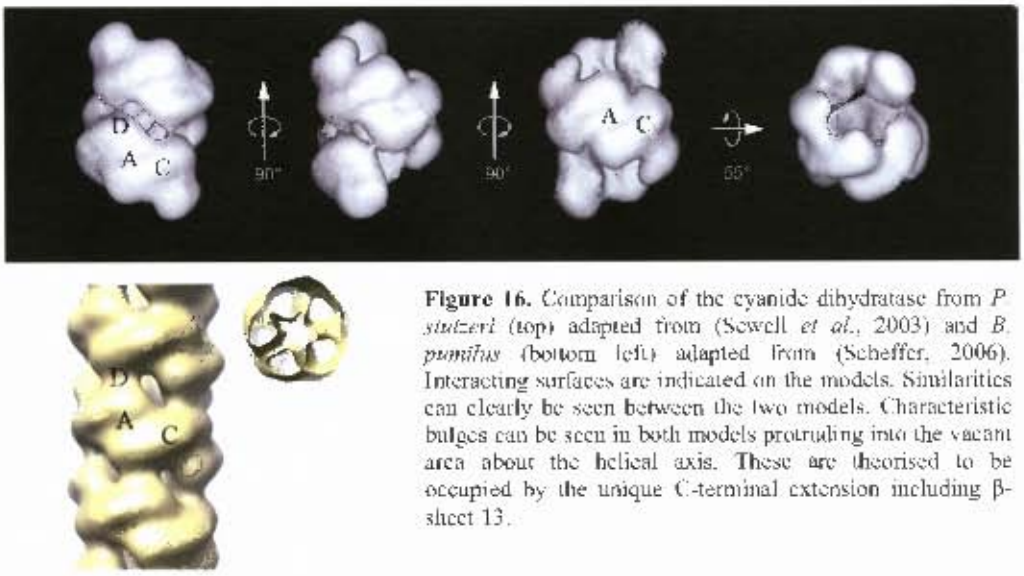

Figure 16. Comparison of the cyanide dihydratase from *P. stutzeri* (top) adapted from (Sewell *et al.*, 2003) and *B. pumilus* (bottom left) adapted from (Scheffer, 2006). Interacting surfaces are indicated on the models. Similarities can clearly be seen between the two models. Characteristic bulges can be seen in both models protruding into the vacant area about the helical axis. These are theorised to be occupied by the unique C-terminal extension including β-sheet 13.

The reconstruction shares fundamental structural characteristics with the CynD from *P. stutzeri* including: left-handedness, global dyad axis on the A-surface, internal bulges and intersubunit interactions including the C and D surfaces (**fig. 16**).

The *B. pumilus* density was fitted with a dimer homology model based on previously solved structures. The insertions that are unique to CynD were modelled as loops in the same general area of the structure to fill up the necessary density. This was specifically important during docking as the dimers could face inward, that is, with the C-terminal extension protruding in toward the helical axis or outward. It was shown that docking the models with β-strand 13 of the C-terminal extension facing inward better filled the available density especially the bulges on the cavity side of the structure. Scheffer postulates that due to its inherent variability and absence in non-extending nitrilases the C-terminal extension most likely plays an important role in oligomerisation.

1.6 The Amidase

Our group recently published the crystal structure of the aliphatic amidase from *G. pallidus* RAPc8 (Kimani *et al.*, 2007). This stereoselective enzyme catalyses amides to the corresponding carboxylic acid and ammonia by amide-hydrolysis and acyl-transfer. The amidase subunit shares the nitrilase α-β-β-α fold with CynD and exists as a hexamer (of 3 dimers) in solution. The amidase also shares the same insertions and deletions with CynD which makes it an excellent template for homology modeling especially as one of the most important insertions, C-terminal extension, performs a homologous function by acting as an interlocking 'arm' holding adjacent subunits together. Identically to CynD the amidase forms the A-surface in between two monomers forming a dimer (and maintaining the α-β-β-α:α-β-β-α sandwich seen in other nitrilases) and the C-surface in between dimers. The A-surface comprises helices α5/6 (residues 172-186, 197-210) and the C-terminal residues 273-340 (specifically α8, α9, α10 and α11) which form an interlock and exhibit a stabilizing force on the active site holding Cys166 in place. The C-surface is created by the interaction of α7 (260-269) and the N-terminal loops of adjacent subunits. The CEK nitrilase catalytic triad is conserved in RAPc8 (Glu59, Lys134, Cys166) and the small active site pocket shows specificity for short aliphatic substrates.

1.7 Aim

The previous work done on the cyanide dihydratase from *B. pumilus* C1 leaves a collection of unanswered questions, not least of which is how the quaternary structure undergoes such dramatic alteration upon pH change. The role of this project was to determine the mechanism and cause of these transitions and to integrate the results of the other students into a coherent model.

The bacterium from which CynD was isolated was *B. pumilus 8A3*. This strain is similar to *B. pumilus C1* (the bacterium used in previous projects) but it was noticed that the histidine residues found on the C-terminal extension of $CynD_{pum}$ in C1 are absent in 8A3 **(fig. 17)**. Elucidating the structure of $CynD_{pum}$ from strain 8A3 at various pHs traversing the point of structural transition from terminating spirals to extended fibres (which is ~pH 6.0, the pKa of histidine) would thus shed light on the transition and possibly implicate the aforementioned histidines as key players in this process. This hypothesis is explained in *3.2.4 Histidines and net-charge*.

Negative-stain micrographs of *B. pumilus* 8A3 were analysed at 3 different pH points: 8, 6 and 5.4. Single particle reconstructions were performed at pH 8 and 6 achieving resolutions of 31Å and 29Å respectively. Finally, a homology model was docked into the densities.

```
C1    1   MTSIYPKFRAAAVQAAPIYLNLEASVEKSCELIDSAASNGAKLVAFPEAFLPGYPWFAFI 60
8A3   1   MTSIYPKFRAAAVQAAPIYLNLEATVQKSCELIDSAASNGAKLVAFPEAFLPGYPWFAFI 60

C1   61   GHPEYTRKFYHELYKNAVEIPSLAIQKISEAAKRNETYVCISCSEKDGGSLYLAQLWFNP 120
8A3  61   GHPEYTRKFYHELYKNAVEIPSLAIQKISEAAKRNETYVCISCSEKDGGSLYLAQLWFNP 120

C1   121  NGDLIGKHRKMRASVAERLIWGDGSGSMMPVFQTEIGNLGGLMCWEHQVPLDLMAMNAQN 180
8A3  121  NGDLIGKHRKMRASVAERLIWGDGSGSMMPVFQTEICNLGGLMCWEHQVPLDLMAMNAQN 180

C1   181  EQVHVASWPGYFDDEISSRYYAIATQTFVLMTSSMYTESMKEMICLTQEQRDYFETFKSG 240
8A3  181  EQVHVASWPGYFDDEISSRYYAIATQTFVLMTSSIYTESMKEMICLTQEQRDYFETFKSC 240

C1   241  RTCIYGPDGEPISDMVPASTEGIAVAEIDVERVIDYKYYIDPAGHYSNQSLSMNFNQQPT 300
8A3  241  RTCIYGPDGEPISDMVPASTEGIAYADIDVERVIDYKYYIDPAGHYSNQSLSMNFNQQPT 300

C1   301  PVVKELNEQKNEVFTYEDIQVQECILEEKV 330
8A3  301  PVVKQLNDNKNEVLTYEAIQVQNGMLEEKV 330
```

Figure 17. Key sequence differences between *B. pumilus* strain C1 and 8A3. Differing residues highlighted in yellow; suspected key histidines are underlined.

1.8 Introduction to SPR methodology

The Single Particle Reconstruction Method

The electron microscope is a powerful and versatile tool aiding visualisation of objects not accessible by conventional light microscopy. The development of powerful computation systems allows working through large mathematical calculations in vastly reduced times to generate 3D reconstructions of 2D electron micrographs. Here follows a description of the protocol used in this project:

1.8.1 Imaging in the Electron Microscope

The electron microscope employs a series of electromagnets as lenses, focussing a beam of electrons (similar to glass lenses and photons in the light microscope) onto a specimen. The basic principle in a light microscope is that photons passing through a specimen are scattered by specific variations in refractive index (e.g. biological material) while unscattered photons (those not interacting with specimen) pass through to generate an image by producing contrast. This is analogous to electron microscopy in which electrons interact with a specimen containing differences in electrostatic potential. Images are formed in the Transmission Electron Microscope (TEM) when electrons experience 'elastic scattering': colliding electrons retain the same energy level but are deflected from their original trajectory and produce high frequency information about the sample. Inelastic scattering involves energy transfer from an electron to a sample atom and a deflection of trajectory. This form of interaction contributes to noise in the image.

The objective lens is found immediately below the specimen and is thus the first to receive image data from the specimen or 'object' being visualised. This lens is convergent and produces a magnification of the image on a distant image plane. However, there exists a point between the lens and the image plane where the unscattered radiation converges on a point and scattered radiation creates an in-plane diffraction pattern. This is known as the back focal plane and the pattern produced is the Fourier

transform of the image. A Fourier transform converts an image into reciprocal space. That is distances are inverted (e.g. 5 Å in real space = 1/5 Å in reciprocal space) and the image is decomposed into its constituent frequencies. As both high resolution information and the majority of noise are found at high frequencies a balance needs to be achieved between these 2 factors that maximises resolution. Inverting this transform will regenerate the original image (Dorset, 1995).

1.8.2 Staining

In electron microscopy staining serves to protect the integrity of the sample (which will subsequently be exposed to a vacuum) and provide contrast. In this project the negative-staining method was used in which a 2% uranyl acetate mixture that interacts strongly with electrons is applied to molecules which are fixed on a carbon coated copper grid. This stain coats the molecule and thus produces a 'shell' of contrast allowing us to reconstruct the outer shape of the molecule. This is a fairly easy technique and involves far less preparation than cryo-electron microscopy but potentially sacrifices an amount of resolution. The two shortcomings of this technique are the distortion of a flexible molecule when air-dried on a grid (as opposed to embedding in ice) and the meniscus effect creating variation in the thickness of the stain, thus potentially rendering one side of the protein 'invisible' to the microscope. This problem of partial concealment can be overcome if the particle has suitable symmetry or by using the double-carbon coating method.

1.8.3 Alignment and Averaging

The primary hurdle which needs to be overcome in Single Particle Reconstruction is reducing noise. In microscopy 'noise' denotes 'any contributions to the image that do not originate with the object' (Frank, 2006). The kind of noise encountered in the TEM is stochastic noise and thus averaging of a set of aligned images (of the same projection) should increase the Signal-to-Noise ratio dramatically as the random components cancel each other out and the signal component is reinforced.

1.8.4 The Projection Theorem

The projection theorem states that the 1D projection of a 2D function can be obtained as the inverse 1D Fourier transform of a line normal to the direction of projection passing through the origin of its 2D Fourier transform (Frank, 2006). This means that the Fourier transform of a 1D projection of a 2D image will occupy a central line in the full 2D Fourier transform of the original 2D image and thus the transforms of two projections of an image from different angles will share a central point in Fourier space. This theorem applies in 3 dimensions as well, so that the Fourier transform of a 2D projection of a 3D volume occupies a central slice in Fourier space of the transform of the volume. Because Fourier slices will all share a common line (a line that is identical in the absence of noise) that passes through the centre of Fourier space, having 3 or more projection-transforms allows us to determine the orientation of all projections relative to each other. By implication this means that by collecting an adequate number of views of a 3D particle with a TEM one can 'populate' Fourier space if the angles of the Fourier slices relative to each other are known (**fig. 18**).

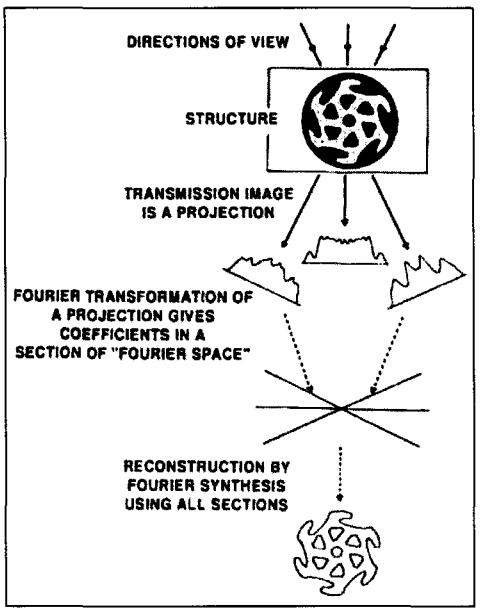

Figure 18. The Projection Theorem. Fourier transforms of projections of a 3D volume are used to approximate the original structure (Jinghua, 2007).

Note that the central core of the 3D Fourier transform (representing lower frequencies) will be filled much more rapidly than the peripheral high frequencies. As the high frequencies contain the fine structural information, resolution is limited by the number of projection orientations that can be obtained.

Chapter 2: Materials and Methods

2.1 Sample preparation and Microscopy

Four microlitres of purified enzyme solution was pippetted onto a fresh glow-discharged grid previously coated with a thin carbon support film under vacuum. In order to reduce precipitation between phosphate buffer and uranyl acetate, grids were subjected to two successive water washes followed by staining with 2% uranyl acetate. At each step, excess sample, wash and stain were blotted. Grids were air-dried before electron microscopy. The salt concentration in the buffer was reduced by a 5–10-fold dilution with distilled water. All staining was carried out at room temperature. Micrographs for image processing were recorded slightly under focus on Kodak S0163 film under low-dose conditions on a JEOL 1200EX II transmission electron microscope operating at 120 kV.

2.1.2. Picking, Filtering and Masking

Micrographs were picked using Boxer (Ludtke *et al.*, 1999), Ximdisp (Smith, 1999) and Signature (Chen & Grigorieff, 2007) in 256 x 256 pixel boxes at a sampling rate of 2 Å.pix^{-1}. 19 395 images were captured at pH 8 and 6732 at pH 6. These image stacks were subsequently scaled down to 128 x 128 pixels (sampling rate of 4 Å.pix^{-1}), filtered to exclude frequencies outside of the range 1/16 Å^{-1} to 1/256 Å^{-1}, normalised and finally the images were masked with a circular Gaussian mask having a falloff function half-width of 15 **(Fig. 19)**.

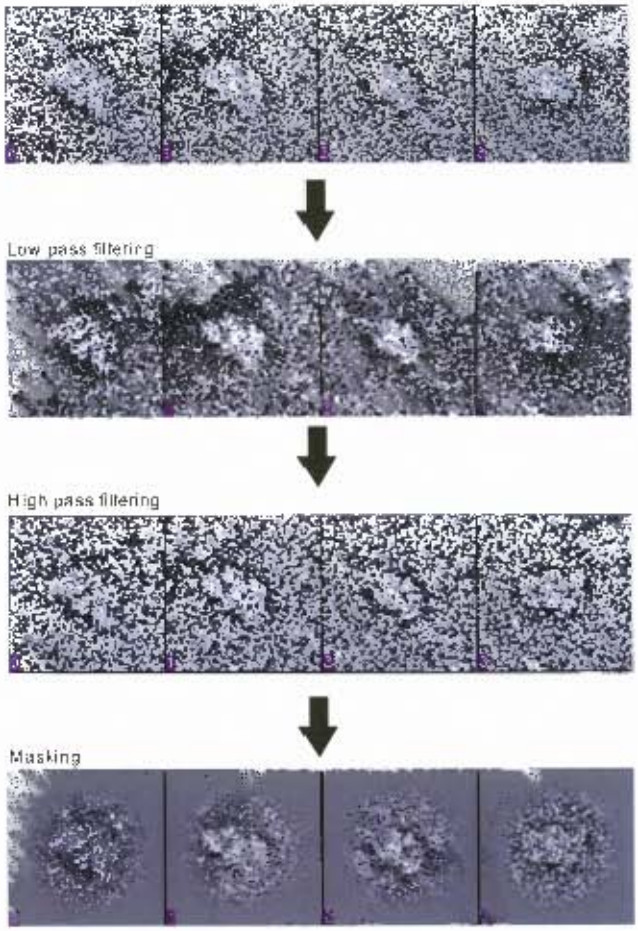

Figure 19. Images were band-pass filtered to a maximum resolution of 16Å to remove excess noise, normalised and masked.

Chapter 3: Results and Discussion

3.1 Micrographs

The negative stain micrographs reveal important information about the pH effects on CynD **(fig. 20)**. At pH 8 one can clearly see the homogeneous "elephant-like" shapes of the 22-mer terminating spiral (note that the dark spots are aggregations of stain and were not picked). The pH 6 micrograph is slightly more heterogeneous and a number of GroEL particles are identifiable. The spirals are in the majority over a small number of transition structures of differing size. As the majority of these structures are smaller than the spiral they are possibly attributable to either protein degradation or a retarded and weaker spiral assembly process. The majority of particles observed at pH 5.4 are of 2 kinds, the first being long irregular strands and the second being small roughly tetramer-sized particles. It is clear that the small particles are aggregating to form the irregular strands as various transition stages of assembly can be seen on the micrograph (**fig. 20 bottom slide**).

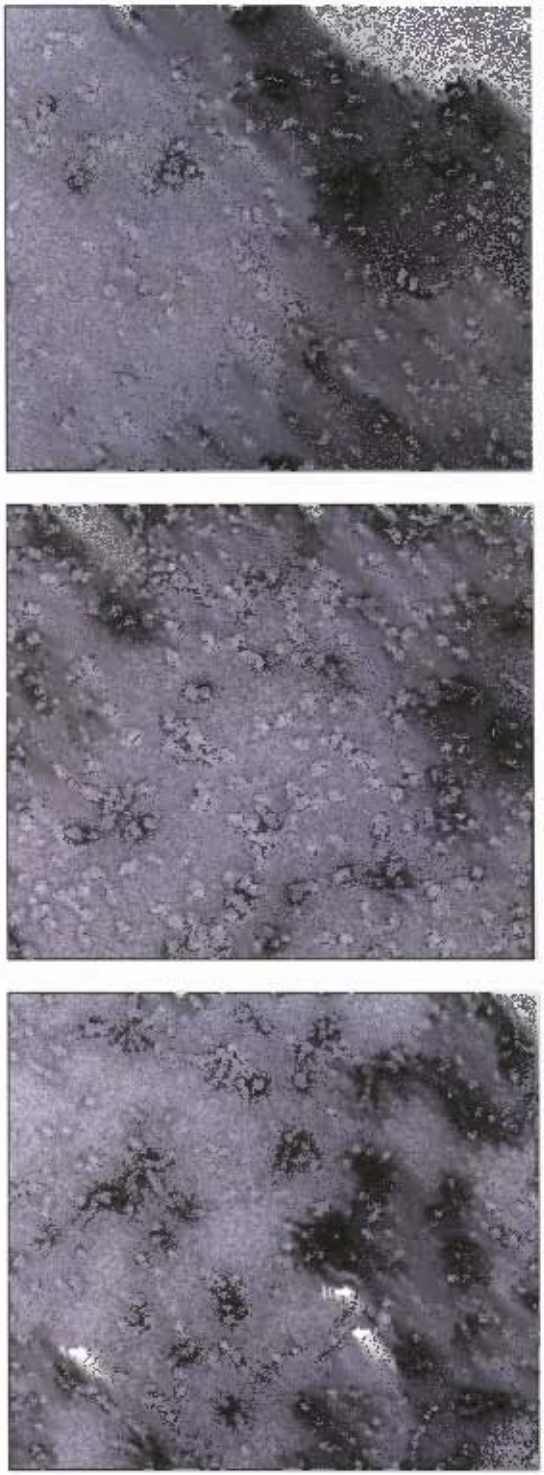

Figure 20. Negative stain Micrographs of *B. pumilus 8A3* CynD at 3 pHs. Top: pH 8, 'elephant-like' terminating spirals are clearly visible, darker conglomerates are excess uranyl acetate stain. **Middle:** pH 6, spirals are visible amongst a collection of heterogeneous particles and some GroEL. **Bottom:** pH 5.4, consists primarily of aggregated strands and tetramer-sized subunits. **A.** Individual subunit. **B.** Aggregated strand showing none of the regularity typical of CynD fibres. **C.** Terminal end of extending strand appears to be currently being augmented by tetramer-sized subunits.

3.2 Reconstructions and docking

Note: for all sequence information refer to Figure 30

3.2.1 Eliminating contribution of contaminating GroEL

The micrographs were contaminated with GroEL particles at a ratio of ~5%. Projections of a 13Å electron micrograph reconstruction of GroEL (Braig *et al.*, 1994) were used to "trap" projections of GroEL. GroEL is a cylindrical structure and thus has 2 preferred major orientations on a flat surface either upright or lying on its side. 4 unique projections were selected based on these preferred orientations and the inherent symmetry of the molecule. The image stacks were aligned to these projections using SPIDER (Frank *et al.*, 1996) and sorted in order of correlation to the projections. GroEL images were then manually deleted from the stacks **(Fig. 21).**

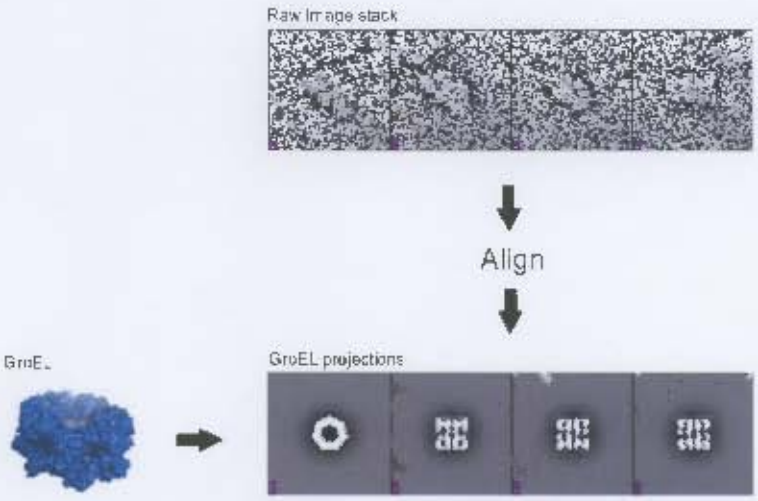

Figure 21. A 2.8Å model of GroEL was used to generate a trap-class and remove GroEL projections from the image stacks.

3.2.2 Iteration

Images were aligned to 84 projections of a starting model of a cylinder with dimensions 30 x 64 pixels (120 x 256 Å, volume of 2.89 x 10^6 Å^3) using SPIDER and subsequently divided into classes based on correlation to a specific angular projection. Within classes an arbitrary cut-off of 2 standard deviations from the mean correlation was used to exclude bad images and the remainder were averaged. These 84 averages were backprojected to recreate a density which was once again projected to produce a template for the next round of alignment (fig. 22).

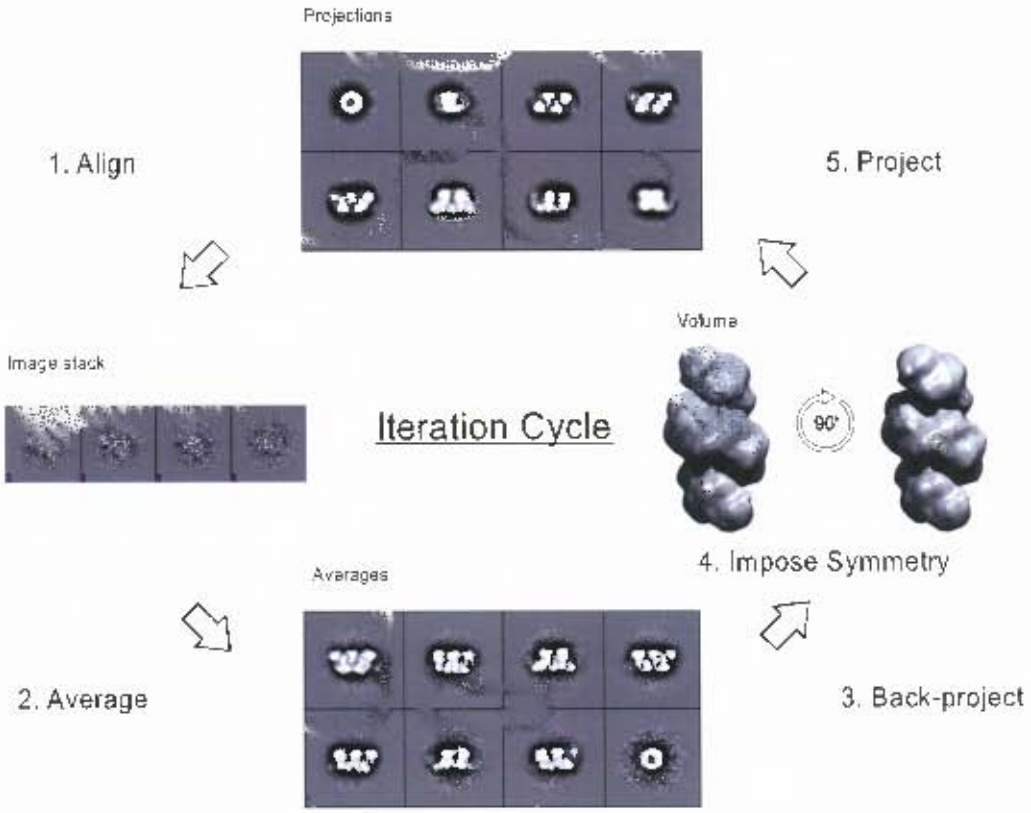

Figure 22. The iteration protocol used in reconstruction volumes of CynD. Images were aligned to projection of a model and separated into classes; class averages were then back-projected to create a volume; 2-fold symmetry was imposed on the volume; projections were made and the alignment procedure would start again.

Initially, only a narrow range of contrast (1.77, 1.99) and a high smoothing factor (0.9994) were used in the SPIDER routine for back-projection to create a crude model from the two stacks. This model was used as a basis for removal of noisy projections that showed very low correlation and blank images. After excluding these images the stacks were reduced to 7811 and 2599 images for pH 8 and pH 6 respectively.

Strong evidence has been presented (Sewell *et al.*, 2003) suggesting that CynD exhibits a 2-fold axis of symmetry perpendicular to the spiral axis which is due to it being constructed of dimers. Thus, during iteration of the images the 2-fold symmetry was imposed. Iteration of these images and the generation of densities proceeded as follows:

Both stacks eventually converged upon a terminating spiral structure **(fig. 24 and fig. 28 for class averages)**. At pH 8 the structure had 22 subunits whilst the pH 6 structure had 20 subunits and resolutions of 31Å and 29Å respectively (determined by Fourier Shell Correlation using the 0.5 Correlation Coefficient cut-off) **(Fig. 23) (Frank, 2006)**. The molecular weight for the 330 residue monomer is 37,295 Da (74,590 Da dimer). Thus, the calculated molecular weights for the 22- 20-mer spirals are 820,490 Da and 745,900 Da respectively. Fischer et al. concluded that average protein density is not independent and is actually a function of molecular weight (Fischer *et al.*, 2004). For high molecular weight proteins (> 30 kDa) the proposed average density is 1.41 g/cm^3. Using this value the volume of a dimer was calculated as being 87,850 Å^3 and thus a 20-mer 1.757 x 10^6 Å^3 and 22-mer 1.932 x 10^6 A^3. After manually inspecting the volumes produced using UCSF Chimera (Pettersen EF FAU - Goddard *et al.*,), density thresholds were set to final volumes of 1.575 x 10^6 Å^3 and 1.656 x 10^6 Å^3 for the pH 6 20- and pH 8 22-mer respectively **(Figs. 25/6)**. This is justifiable as low resolution single particle reconstructions such as these produce an outer shell of the molecule. The actual volume occupied by the protein is thus probably significantly less. Images were represented in all classes for both pHs but showed definite trends. This indicates a preferential orientation of the molecule on the carbon grid **(Fig. 27).**

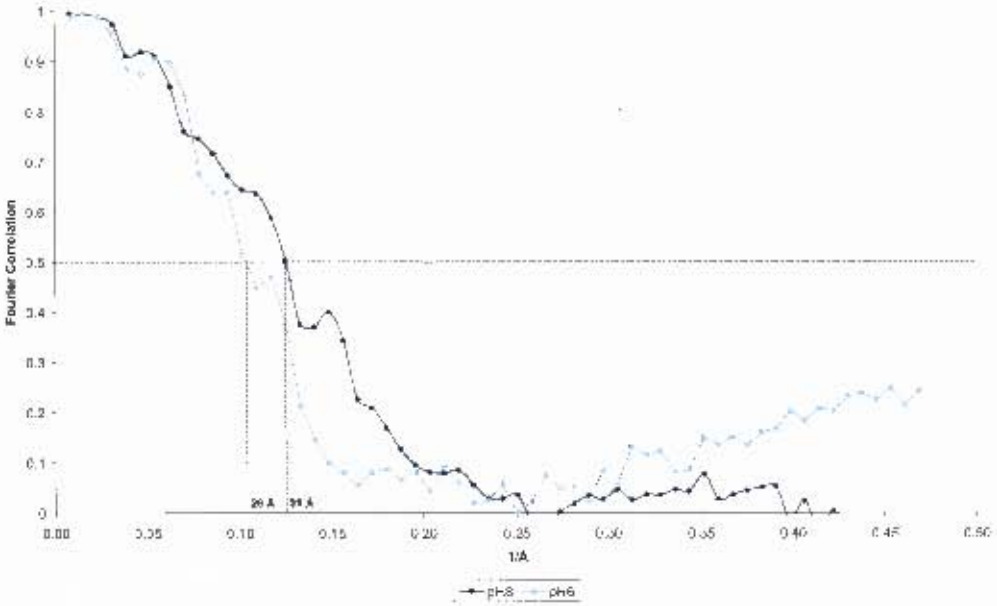

Figure 23. Fourier Shell Correlation of pH 6 and pH 8 reconstructions. Estimated resolution based on 0.5 correlation cut-off is indicated on the graph.

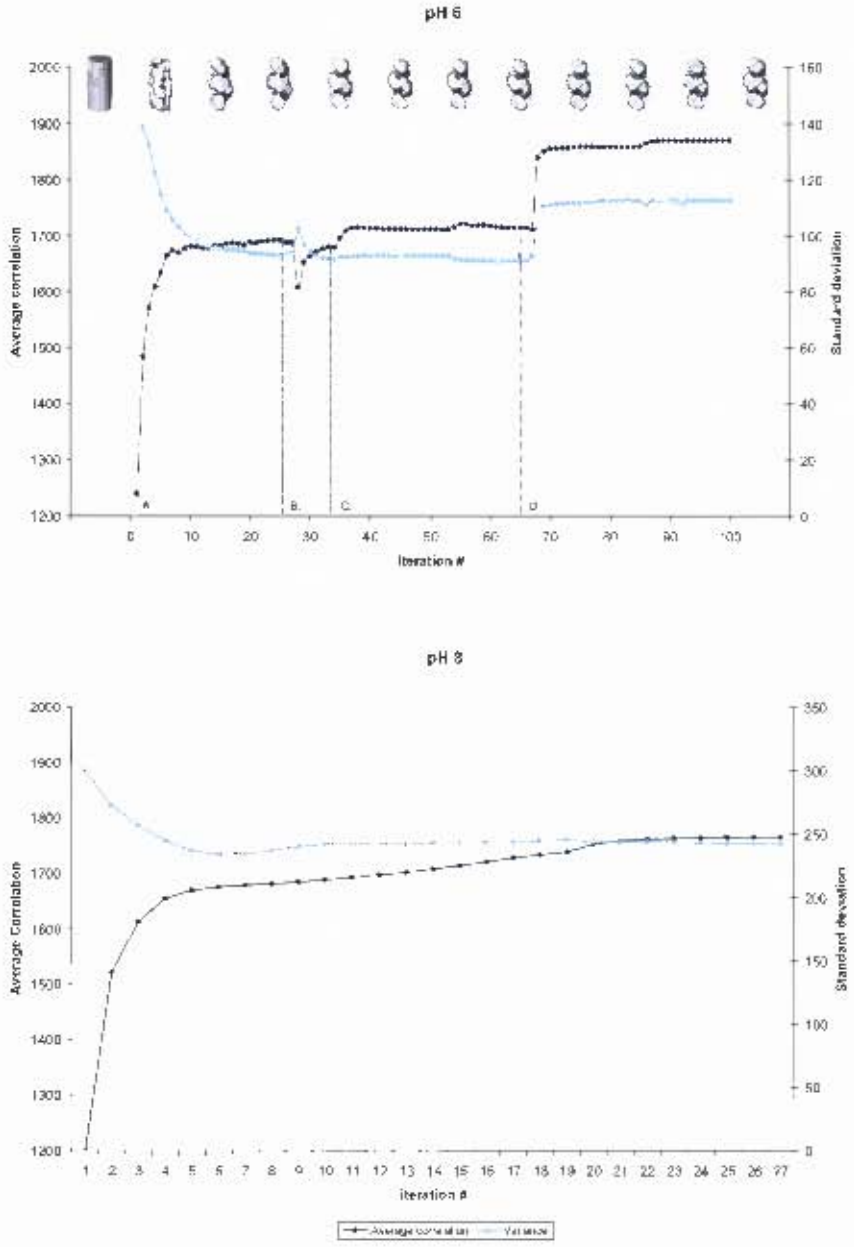

Figure 24. Convergence of reconstructions. Iteration number is plotted against Average Correlation and Standard Deviation for datasets at pH 6 and pH 8. The pH 6 plot is divided into 4 different construction 'stages' (see *Table 2*).

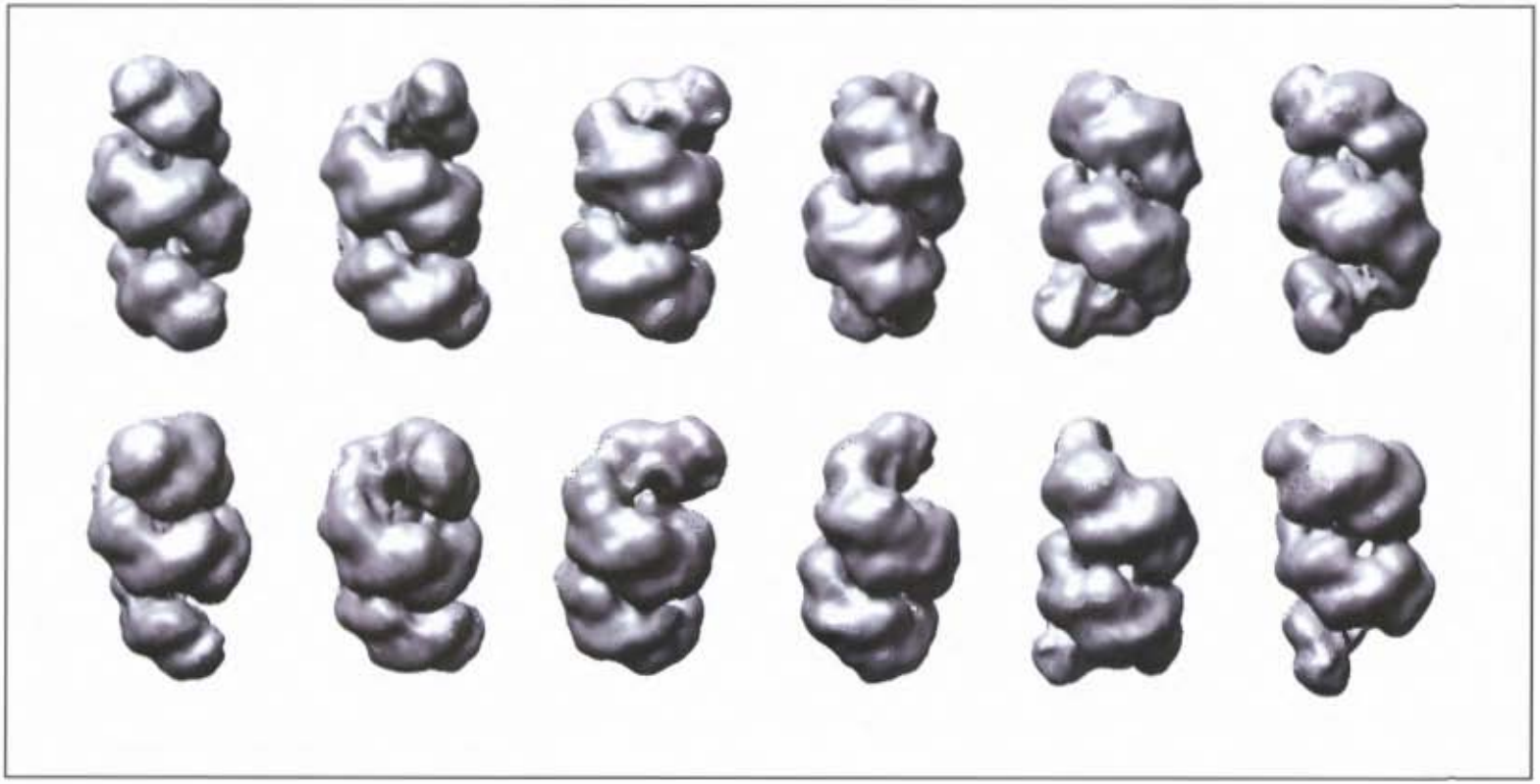

Figure 25. 29 Å resolution reconstruction of CynD at pH 6. Top row is rotated by 60° increments; bottom row is initially tilted by 30° then rotated by 60° increments.

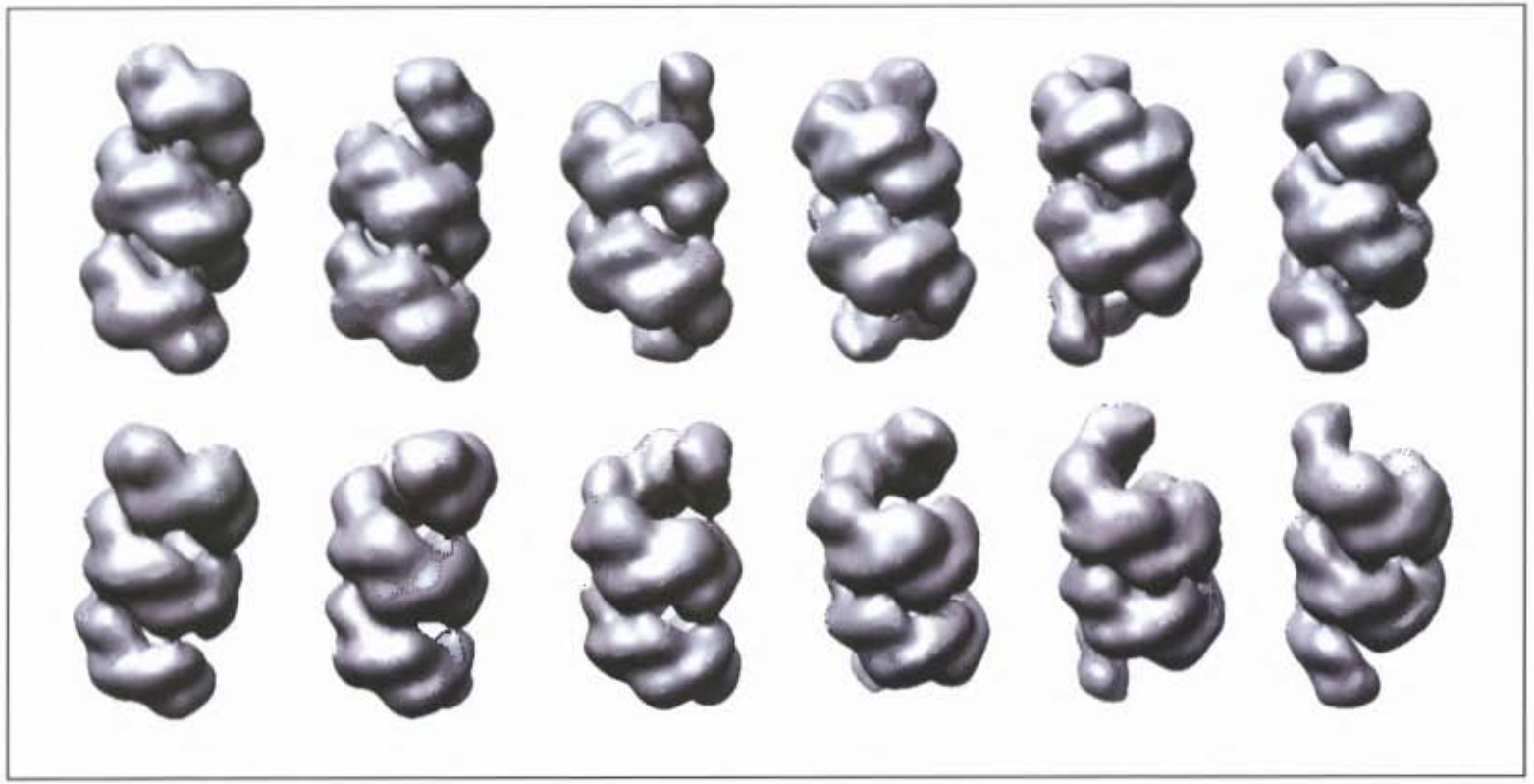

Figure 26. 31 Å resolution reconstruction of CynD at pH 8. Top row is rotated by 60° increments; bottom row is initially tilted by 30° then rotated by 60° increments.

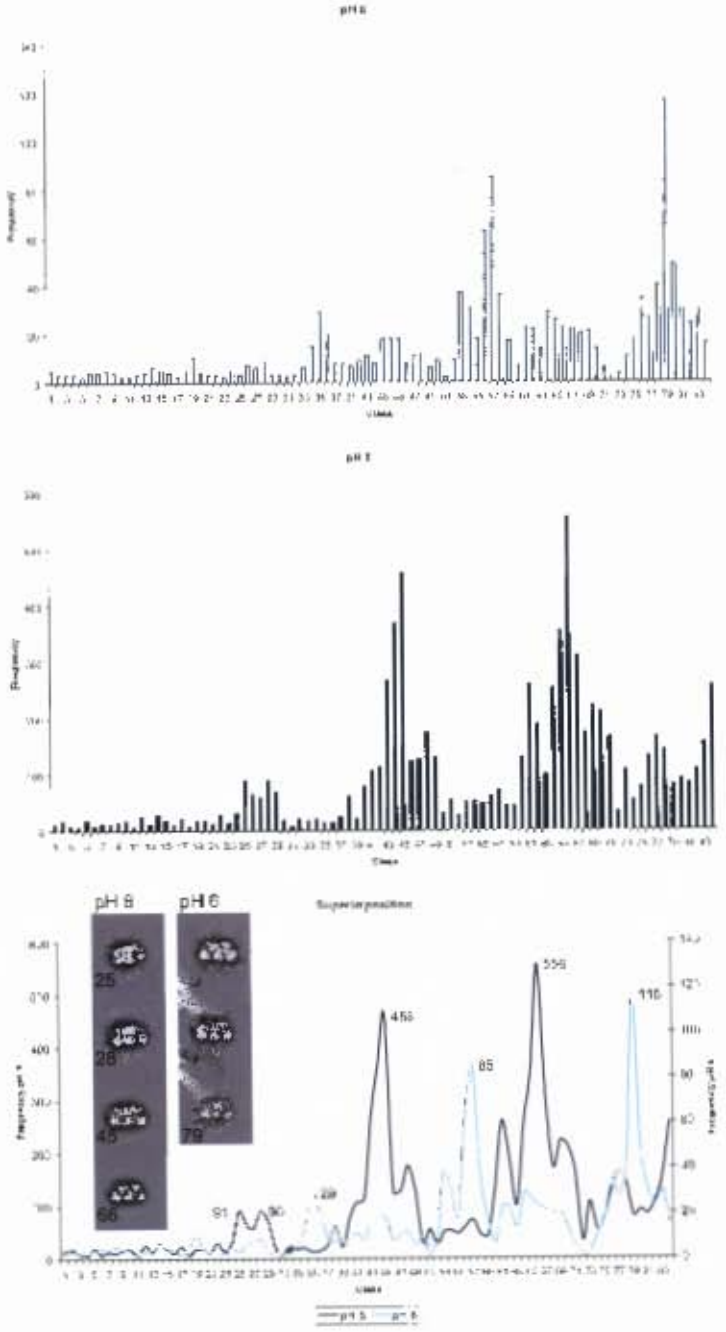

Figure 27. Distribution of Class Frequencies. Top: pH 6, Middle: pH 8, Bottom: pH 6 and pH 8 class frequencies superimposed. Peak values are labeled and indicate a clear preference for two spatial orientations at each pH. The classes being 25, 28, 45 and 66 at pH 8; 35, 57 and 79 at pH 6 (insert).

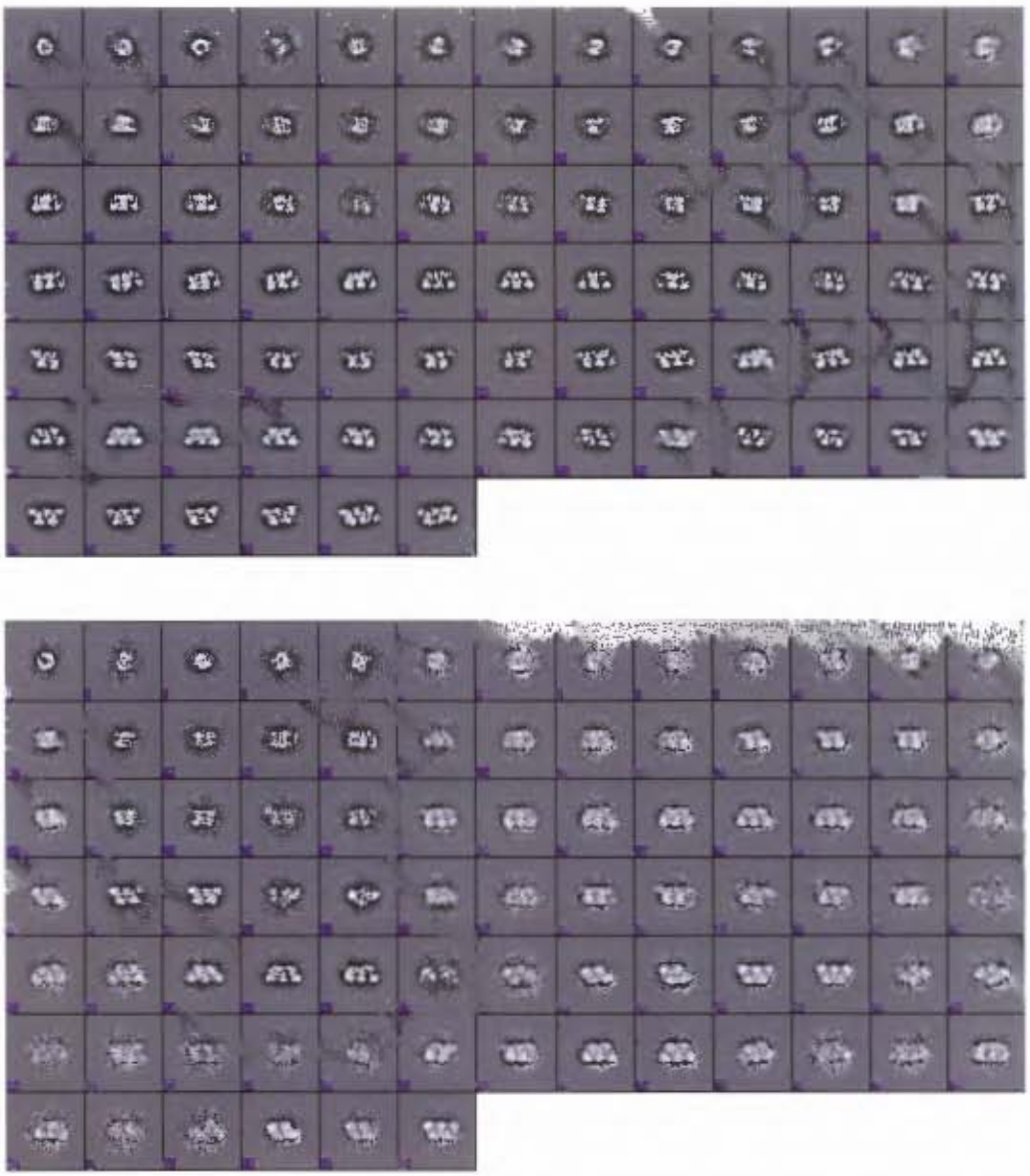

Figure 28. Class averages. **Top.** pH 8 after 27 rounds of iterations. **Bottom.** pH 6 after 100 rounds of iterations. *Note: numbering is according to EMAN convention beginning at 0 instead of 1.*

3.2.3 A stage-based approach to pH 6

The pH 6 dataset required 4 different stages of refinement (labelled as **A → D** in *Table 2* and ***Fig. 24***). The first stage, A, produced a very low resolution model due to the very narrow range of contrast and high smoothing factor used in back-projection. Stage B was identical to Stage A with the exception that 2-fold symmetry was imposed. After this Stage C involved halving the image set (2599 → 1299 images) based on correlation to the last model in Stage B. This step dramatically improved resolution. The final step included all available contrast in back-projection and excluded all smoothing effects. This stage (D) also improved resolution as can be seen by an increase of ~200 (~10%) in the correlation at which the iterations converged.

When this reconstruction was repeated from scratch with the reduced image set and only the parameters (**Table 2**) from Stage D the data did not converge. In fact the resulting model was egg-like in shape and did not resemble a spiral.

Table 2: Iteration cycle

		pH 8	pH 6			
Subunits		22	20			
Resolution		31 Å	29 Å			
Images		7811	2599			
Starting model		120 x 256 Å cylinder	120 x 256 Å cylinder			
Total Iterations		27	100			
Construction Stage		**A**	**A**	**B**	**C**	**D**
Images		7811	2599	2599	1299	1299
Iterations		27	25	8	32	35
Search radius and increment		For 3 rounds: 12, 3 Subsequently: 10, 1	For 3 rounds: 12, 3 Subsequently: 10, 1	10, 1	10, 1	10, 1
Back-projection	**Radius**	35	35	35	35	35
	Contrast	-1.77, 1.99	1.77, 1.99	1.77, 1.99	1.77, 1.99	-1.77, 1.99
	Smoothing	0	0.9994	0.9994	0.9994	0
	Symmetry	Yes	No	Yes	Yes	Yes

*(See **figure 25** for a visualisation of stages of construction)*

3.2.4 pH 8 converging on 2 volumes

A similar protocol was used for the initial reconstruction of the pH 8 dataset resulting in a 20 subunit volume indistinguishable from the pH 6 volume. However, when repeated with only the Stage D parameters, the pH 8 dataset converged on a 22 subunit volume. This reconstruction was repeated with two different starting models, the initial cylinder and a conglomeration of 3 overlapping spheres. The resulting volumes remained identical at 22 subunits.

To exclude the possibility of a heterogeneous sample plots were made of the frequency of images at each correlation value. Scheffer (Scheffer, 2006) notes that a good indicator of heterogeneity is a bimodal distribution of such a plot. This would be caused by two image sets derived from different molecules being found to have different average correlation values. **Figure 29** shows a unimodal, indeed almost Gaussian, distribution in the pH 8 dataset. Scheffer's situation may have arisen from heterogeneity along the fibre, particularly in the D-surface interaction.

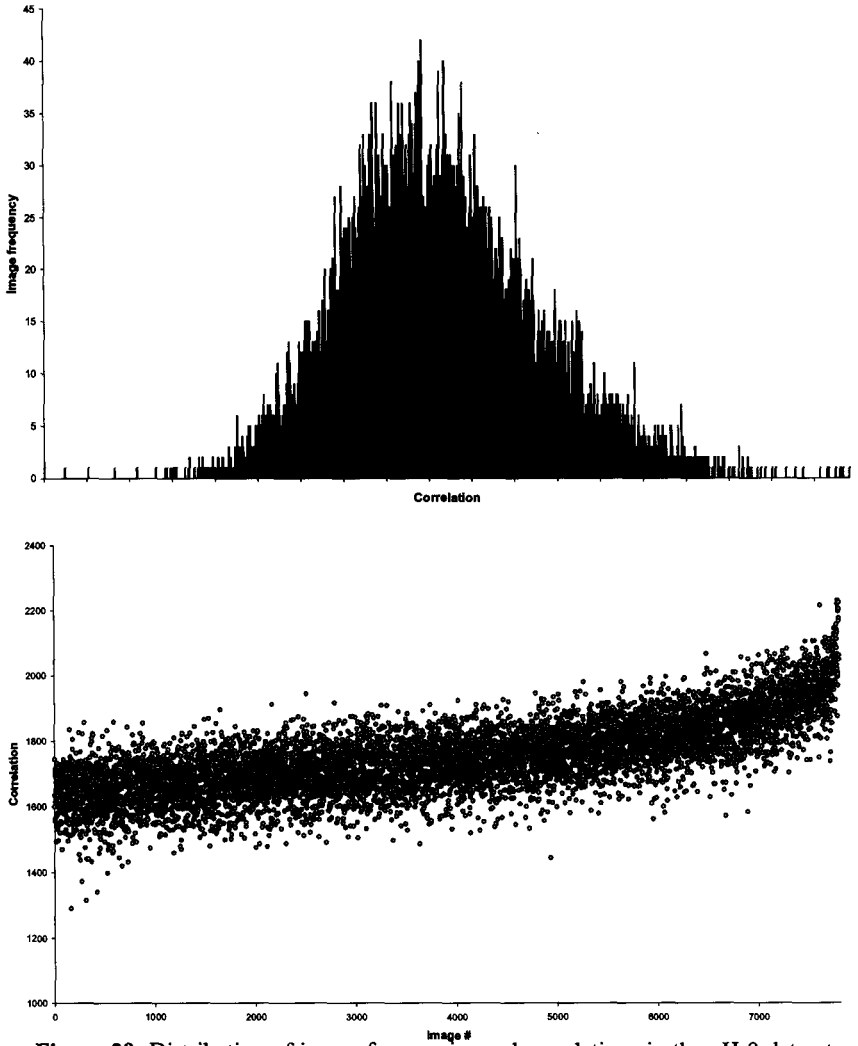

Figure 29. Distribution of image frequencies and correlations in the pH 8 dataset. Top: Number of images per Correlation value. Bottom: correlation coefficient of each image. These graphs demonstrate a unimodal distribution of images suggesting that only one species of molecule exists and thus a homogeneous sample.

Though this is a good indicator this criterion is not entirely conclusive. To further clarify this issue and to exclude the possibility of anomalous effect from enforcing symmetry both the pH 6 and pH 8 models were allowed to continue for 10 iterations without the symmetry impositions. The result was that both models stayed essentially the same as a 20- and 22-mer at pH 6 and 8 respectively.

Another indicator is partial density. If the terminal subunits were seen to be of partial density, that is, they disappeared soon upon increasing the density threshold, it could be postulated that they are a result of bad alignment. This could take place if a there is a ratchet-like alignment of the oligomeric spirals leaving a fraction of images with subunits extending beyond the real terminal subunit whilst still contributing to the density of the central subunits. This is not the case with the pH 8 volume. In fact when increasing the density threshold in Chimera the terminal subunits of the 22-mer are of the last pieces of density to remain when all other density has disappeared. The pH 6 dataset was reconstructed with the pH 8 22-mer volume as a starting model. The resulting volume converged upon a low resolution 22-mer with partial density for terminal subunits. This suggests that the particle is in fact a 20-mer.

The aligned datasets were classified using a k-means clustering algorithm which attempts to find a natural clustering of data by treating data-attributes points in vector space (Macqueen, 1966). If the aligned datasets were in fact 'ratchetting' one would expect to find that individual datasets clustered into at least 2 distinct classes. This is not the case therefore it is safe to deduce that the pH 8 and pH 6 particles consist of 22 and 20 subunits respectively.

3.3 Homology modelling

Homology modelling was done using Modeller version 9 (Sali & Blundell, 1993). Modeller assigns a Probability Density Function to the position of each atom and attempts to satisfy a set of spatial restraints to predict by *ab initio* modelling the structure of a selected target protein sequence against a template.

In the case of $CynD_{pum}$, the most important areas from an evolutionary perspective are the insertions that are missing from other nitrilases. Because they are missing it is difficult to provide a complete template for modelling purposes. The C-surface (proposed to be the interacting surface between dimers along the helix/spiral) is comprised primarily of 2 insertions in the sequence **(Fig. 30)**: the first insertion being residues 54-71, the second being 221-235. There is also a significant C-terminal extension.

Initially the secondary structure was predicted using PSIPRED (McGuffin *et al.*, 2000) online server. These predictions were included as constraints in the modelling process.

Three template sequences were chosen:

1J31 - Hypothetical protein PH0642 from *Pyrococcus horikoshii*

1ERZ - N-Carbamyl-D-Amino Acid Amidohydrolase from *Agrobacterium tumefaciens*

2PLQ - Amidase from *Geobacillus pallidus*

The three templates share the nitrilase domain whilst the third exclusively contains an elongated C-terminal tail which was included to model the C-terminal extension of $CynD_{pum}$. As no sequence was available for the first two insertions, they were dealt with as follows:

Insertion 1: According to PSIPRED predictions this insertion was constrained to 2 alpha helices. The first helix spans residues 55-58, the second 62-76 (effectively an N-terminal extension of A2 in the other sequences).

Insertion 2: This insertion had no defined secondary structure and was eventually excluded from the model.

The extension was modelled successfully, and similar to the amidase the tail wrapped around its symmetry-related partner, interacting at the C-surface and returned to contribute to the A-surface (see section on *Quaternary structure*). To model the A-surface correctly dimer templates were used. ProSA (Wiederstein & Sippl, 2007) is a program which recognises geometrical errors in protein models and assigns them an overall Z-score based on correctness. The Z-score for this homology model was -3.89 which is within the range of scores for database crystal structures with the same number of residues.

The resulting model had 12 α-helices and 10 β-strands **(fig. 30)**. For the sake of consistency, the secondary structural annotations have been generalised for the all the homologues. This means that β6, 10, 12, 13 and α-8 aren't present in $CynD_{pum}$ 8A3 but have been included in the numbering scheme. Also based on the modelling and for greater resolution, α2 and α5 have been split into 2 helices each (thus 'NH2 of the C-surface' is actually referring primarily to α2b). It should also be noted that the secondary structural elements in the second insert are not highlighted as that specific insert was not modelled. Therefore β10 could very well be present in nature. PSIPRED predicted that α2 extends backward through the first insertion sequence which is not the case for the homologues.

Figure 30. Sequence alignment of CynD and homologues. 1j31, 1erz and 2plq were used to model Cyn from *B. pumilus* 8A3. Members of the catalytic triad are highlighted in grey. Helices are highlighted in red, strands in yellow and are annotated below the sequence. Insertions 1-3 are boxed. Surfaces are annotated above the sequence. Alignment of the solved template structures was performed using ALIGN (Cohen, 1997) and edited manually. *Note:* 1EMS – *Helicobacter pylori* formamidase, 1ERZ – *Agrobacterium sp.* Amidohydrolase, 1UF5 – *Agrobacterium sp.* Amidohydrolase C171A/V236A mutant. *Geobacillus pallidus* Hydrolase.

3.4 Docking

Models were docked manually using Chimera and aided by the 'Fit model in map' function to optimise the docking. Model-fitting was aided by the high definition of dimers in the density. Agreeing with Scheffer (Scheffer, 2006) the C-terminus was docked facing inward toward the cavity of the spiral. This agrees well with the available density. The pH 8 22-mer was fitted with the A-surface coinciding with the global 2-fold axis of symmetry. This axis coincided with the C-surface in the pH 6 20-mer model.

3.4.2 C-surface

As has been discussed, the pH 6 particle consists of 20 subunits with an axis of symmetry about the C-surface whilst the pH 8 22-mer is symmetrised about the A-surface. The C-surface is comprised of:

Residues	Secondary Structure	Sequence (Hydrophobic – red, Hydrophilic – blue)
62-76	α 2	HPEYTRKFYHELYKN
139	α 4	L
281-282	α 8	DP
295-307	α 9 – loop	FNQQPTPVVEQLN
325-329	α11	MLEEK

The first and third sequences are provided by one monomer, the second and last two by its partner and thus there is a pseudo-dyad symmetry about the interface (**fig. 31**). The first sequence (α2) is contributed by an insertion relative to homologous sequences. The last two sequences are in the C-terminal extension which wraps around the opposite monomer and folds back to form part of the A-surface. Thus 3 of the sequences that create the C-interface are insertions in *B. punilus*. Also, the insertion that was not modelled appears to belong to the C-surface too and it consists of the sequence: 219-EMEEMICLTQEQREVTETFESGHT-242. This insertion is known to be important

as a Δ219-233 deletion mutant is inactive, potentially implicating the C-interaction and consequently oligomerisation as a requirement for activity (**Table 1**). Also, it has been shown that the C-terminal deletion Δ293, which would delete α9/10/11 destroys most of the enzyme's activity. Extending the deletion to residue 279, thus excluding α8 as well, totally destroys activity (**Table 1**). Considering that an increase in pH is noticed below pH 6 (Jandhyala *et al.*, 2005) (i.e. upon oligomerisation) and that the C-surface is required for activity it is speculated that oligomerisation directly determines activity.

The fact that the majority of residues comprising the C-surface are insertion sequences suggests that forming the C-interface is an evolutionary property of *B. pumilus*. The other nitrilases known to share these three insertions are the Cyanide dihydratase from *P. stutzerii* ($CynD_{stu}$) and the Cyanide hydratase from *G. sorghi* (CHT). Both of these proteins have the ability to extend by forming the C-interface. The fact that $CynD_{stu}$ remains as a terminating spiral (Sewell *et al.*, 2003) and CHT remains as an extended fibre (Scheffer, 2006) regardless of pH indicates that the degree of extension is not dependent upon the C-surface.

Thus, the interaction between two dimers is mediated equally by hydrophobic and hydrophilic residues in the helices α2, α4, α8, α9 and α11, one loop and an unknown structure of which 4 of these sequences are insertions.

Figure 31. Interdimer junction of adjacent C-surfaces. The interacting regions across the C-surface are highlighted with different colours: Green - 62-76 (α2), Red - 295-307 (α9), Orange - 325-329 (α10). The beginning and end of the unmodelled insertion sequence is highlighted in cyan (E218) and yellow (C243).

3.4.3 D-surface

The D-surface is the sequence that creates the inter-groove bond likely mediated by 90-EAAKRNE-96 (NH3) and potentially E31 (NH1) between dimers i=1 and i=6. The D-interaction is considered to be the primary stabilising factor upon elongation as it holds the extending helix/spiral together in the direction of extension.

Three types of inter-groove bonds are noticeable in the pH 6 reconstruction **(fig. 32).** Two of these are D-interactions, the first of which is opposite to the middle C-surface and lies on the global axis of symmetry **(fig. 32c/f)**. This structure will be designated D1 and is most likely mediated by the predicted 90-EAAKRNE-96 sequence in NH3 as the two symmetry-related NH3 helices containing this sequence lie neatly opposite each other while NH1, containing E31, is twisted away from the interaction. The second D-interaction (D2, **fig. 32b/e)** can be seen if the molecule is rotated by 90° from the symmetry axis. Because of symmetry there are 2 of these interactions in the spiral and

they are structurally almost identical to the first kind of D-interaction. The last interaction involves the E-surface and is dealt with in its own section below.

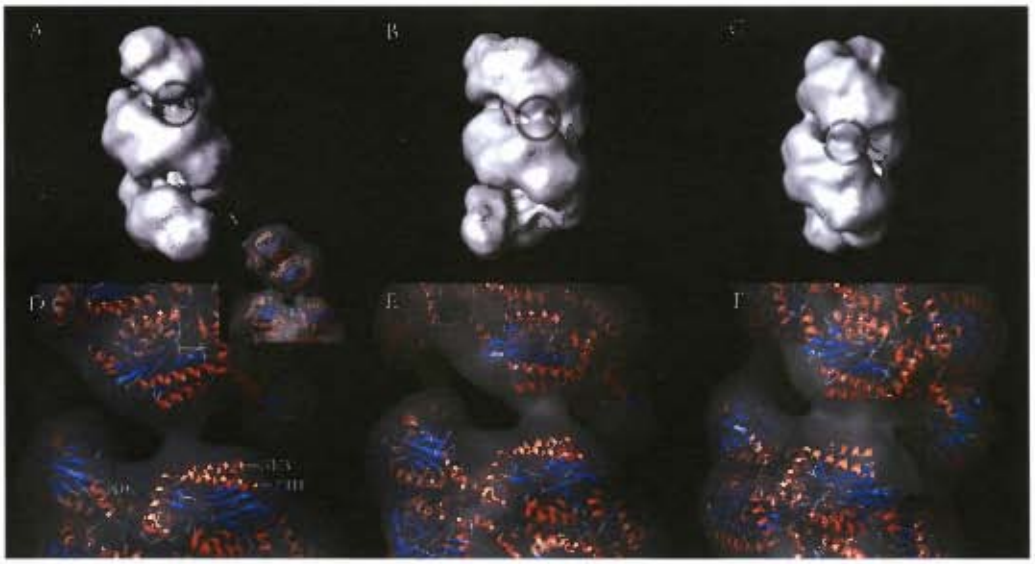

Figure 32. The inter-groove interactions in CynD at pH 6. A/D: E-surface (D3); B/E: D2-surface viewed at 90° to the axis of symmetry; C/F: D1-surface interaction viewed along the axis of symmetry from the 'rear' of the molecule. Major contributing helices are labeled in D. X – inset, low thresholding generates an additional E-interaction labeled E2.

The pH 8 structure has 4 types of inter-groove interactions: 2 involving the D-surface, 1 unknown interaction and 1 involving the E-surface **(fig. 33)**. The symmetry-axis of the 22-mer runs through the A-surface (between two monomers) of the central dimer. The D-interactions at the 'rear' of the molecule **(D1, fig. 33c/f)** appear to be shifted with respect to predicted interacting sequence in NH3. This is not seen in the other D-interactions along the 'side' of the molecule **(D2, fig 33b/e)** which are essentially identical to those seen at pH 6. This might implicate some other suspect residues found in NH2: 72F and 75K, as well as E79 which is found on the loop joining NH2 and NH3. Rather, what seems to be happening is that there are a number of variations of the D-interaction. This is observed as a translation which is mild in D1, greater in D2 and greatest in E (which is henceforth annotated as D3). If these translations are increasing it would indicate and increase in $\Delta\varphi$ (translation per subunit) and thus a decrease in helical radius.

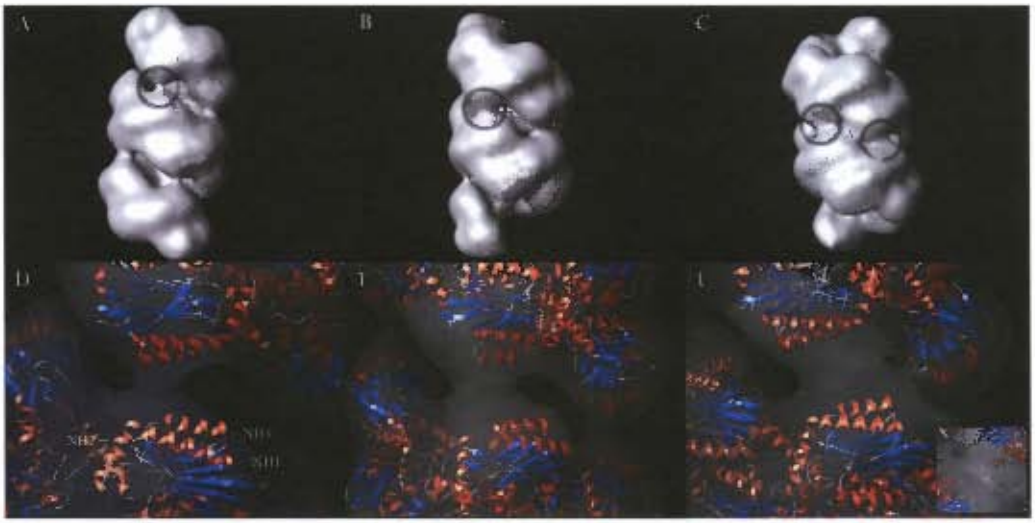

Figure 33. The inter-groove interactions in CynD at pH 8. A/D: E-surface (D3); B/E: D2-surface viewed at 90° to the axis of symmetry; C/F: D1-surfaces and hypothetical interaction viewed along the axis of symmetry from the 'rear' of the molecule. Major contributing helices are labeled in D.

To further elucidate the interactions at the D-surface, a helix was generated using a program written by Sewell called HELIXBUILD using the 8Å3 $CynD_{pum}$ homology model. The helical parameters were optimised by docking the model into Scheffer's map of *B. pumilus C1* CynD fibre at pH 5.4 using COLORES which is part of the SITUS package (Wriggers *et al.*, 1999) **(fig. 34)**. COLORES yields a comparable correlation coefficient for each docking experiment allowing one to track the optimisation of a model when adjusting for fitting into a density volume. The generated helix had the following parameters: $\Delta\varphi$ = -76.67, ΔZ = 15.87, r = 44.5 Å. This helix was also docked into Van der Vyver's map of *B. pumilus C1* CynD fibre at pH 5.4 and yielded a high correlation value. NH3 was shown to be the dominant contributor to the D-interaction as is the case with terminating spirals.

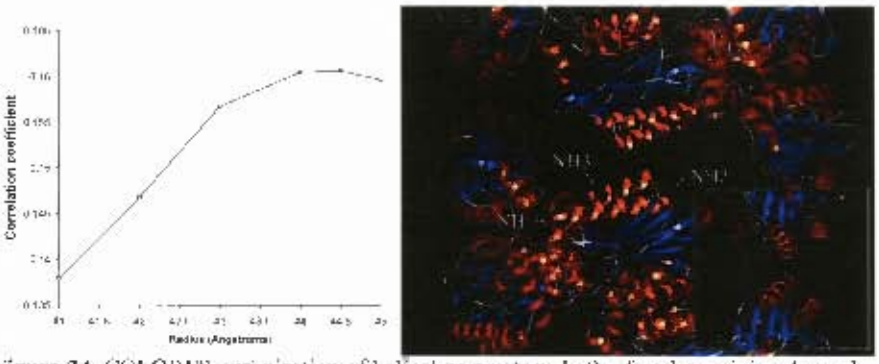

Figure 34. COLORES optimization of helical parameters. Left: after determining Δφ and Δz the optimal radius was elucidated by successive docking and correlation. Right: the D-surface of the generated helix with side view as inset showing that NH3 contributes the primary interaction.

Mutants MB3799 and MB3799-2 **(Table 1)** which served to make the D-surface positively-charged in the hopes of disrupting it (and consequently disrupting oligomerisation) formed more regular helices. The explanation for this might be that there are multiple salt-bridges that could form to create the D-surface and by deleting glutamate residues one is selecting against heterogeneity amongst the particles. Indeed, Scheffer discovered heterogeneity in her reconstruction of $Cynl2_{pum}$ C1 at pH 5.4 by noticing a bimodal distribution of image frequency to correlation coefficient (Scheffer, 2006). Separating the two species (one a higher average correlation than the other) and reconstructing individually yielded 2 helices with the following parameters:

High Correlation	-	$\Delta\varphi = -76.67$ $z = 15.87$
Low Correlation	-	$\Delta\varphi = -76.2$, $z = 15.91$

The primary difference being half a degree in Δφ, the rotation per subunit and at the D-surface a translation of about 1Å for NH3 relative to it's symmetry-related partner (NH3'). This translation is indicative of a switch between two possible binding states as shown in **figure 35**. This translation could push E96 into a position which favours interaction with R94 over K93 explaining the higher correlation coefficient in the species with four salt-bridges as opposed to only two. The low correlation species could possibly be stabilised by an additional interaction also this hypothesis suggests different D-surface

interactions in the spirals e.g. the translated D1 **(fig. 33c/f)** vs. the untranslated D2 interaction **(fig. 33b/e)**.

It must be noted that the pH 8 density shows an unknown interaction in between the two D1 interactions **(fig. 33x)** which may be an artefactual product of thresholding. The expected residues involved in this interaction would be found in the loop between NS4 and NS5, NH3 and the loop between NS5 and NS6.

The results of mutant MB3799 **(table 1)** indicate that removing E90, E96 and exchanging K93 with R94: 90EEAAKRNE96 → 90AAARKNK96 acts to stabilise the extension of $CynD_{pam}$. This can be explained by the presence of 3 acidic residues in NH1, namely, E31, D34 and E35 and an interaction with K94 **(fig. 36)**. This would also explain the continued stability of MB3799-2 in which the suspected bridge-former E31 was mutated to K. Homology modelling of the helix shows that residue D34 is in closer proximity to K94, thus future work would be to mutate the remaining two residues in NH1.

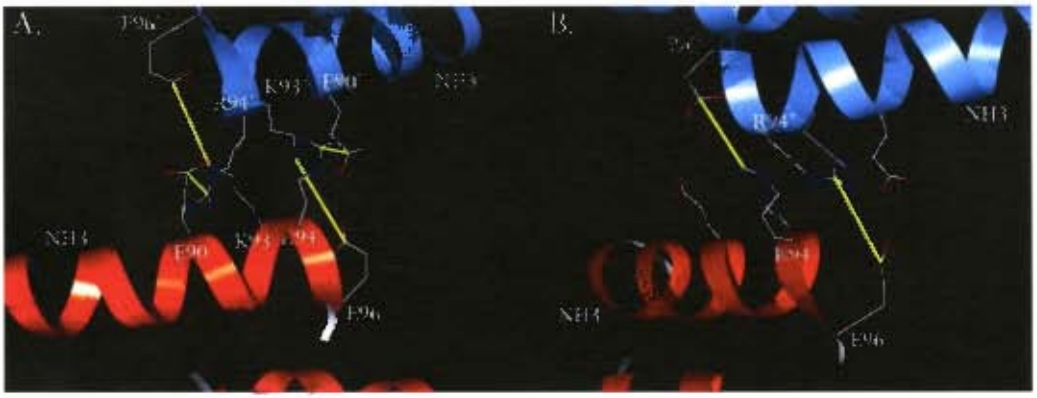

Figure 35. The D-interaction of two possible species of CynD 8A3 due to a translation of NH3 relative to NH3'. **A.** The high correlation model with salt-bridges between E90-R94 and K93-E96. **B.** The low correlation model with a single salt-bridge between R94-E96.

Figure 36. Three possible stabilizing salt bridges in MB3799. L94 potentially interacts with either E31', D34' or E35' which are on NH1' of the opposite dimer.

Figure 37. The E-surface terminating interaction at **pH** 6. Traditionally, residues D267 and D269 are considered to form this interaction by forming salt-bridges with the D-surface residues but are too far away in this model (indicated by circles).

3.4.4 E-surface (D3)

The E-surface forms the terminating interaction at the ends of the spiral. This interaction is thought to be formed by D267 and D269 (E266 and D268 in $CynD_{sta}$) interacting with the D-surface residues. Neither the pH 6 nor the pH 8 models show this to be plausible **(fig. 37)**. The E-interaction at both pH 8 and pH 6 is almost indistinguishable from the D-interaction with the exception that the terminal subunit at pH 6 is tilted counter-clockwise along its own axis of symmetry (A-surface) to the extent that lowering the thresholding creates a second E-interaction designated E2 **(fig. 32x)**. This tilting of the terminal dimer in conjunction with the concomitant narrowing of the helical diameter would hypothetically block the addition of another dimer. This interaction is most likely mediated by residues in NH7 (K277, D281), NH11 (E327, E328) and the loop joining NH4 and NS7 (E155). These residues are directed towards the D-surface of the opposite dimer. Thus this surface could be considered as one of the contributors to termination in conjunction with the tilting of the terminal dimer and the narrowing of the helical diameter.

3.5 Helical Radius

To determine whether terminal narrowing of the helical radius was in fact taking place, the helical radius of both reconstructions was analysed by assigning a centre-of-gravity (COG) to each subunit and measuring these distances from an assigned helical axis (**fig. 38**). Initially the two models were aligned (CHIMERA). The resulting subunit radii produced a similar profile due to the almost identical electron density maps (**fig. 39 Top/Middle**). Terminal subunits narrow and appear to 'close-off' the spiral possibly preventing further extension at pH 8 (penultimate dimers at pH 6). Both models also demonstrate a narrowing near the centre of the spiral. Terminal narrowing at pH 6 is mediated primarily by the penultimate subunits. The terminal subunits protrude away from the helical axis into solution. This could be evidence of a termination mechanism as the terminal subunits are directed away from the ideal position to accommodate an additional dimer.

The diameter of the spirals with respect to the terminal COGs is 207.4 Å for the 20mer at pH 6 and 179.2 for the 22mer at pH 8 (**fig. 38 B/E**). The helical radii of the COG for the dimers (A-interaction) were produced by averaging the subunit radii (**fig. 39 Bottom**). These profiles were also similar in both reconstructions with the exception that the penultimate dimers at pH 6 have a much narrower radius than pH 8. These results are indicative of an increase in $\Delta\varphi$ and a decrease in helical radius. As the pH drops from 8 to 6 CynD potentially experiences a reconfiguration of interactions along the spiral particularly a translation of the residues that make up the D-surface. This reconfiguration results in the loss of a terminal dimer and termination at 20 rather than 22 subunits. In the case of C1 a further reconfiguration must take place causing the extension into a fibre probably mediated by the key Histidine residues which become charged below pH 6 (*see below:* ***3.6 Histidines and net-charge***).

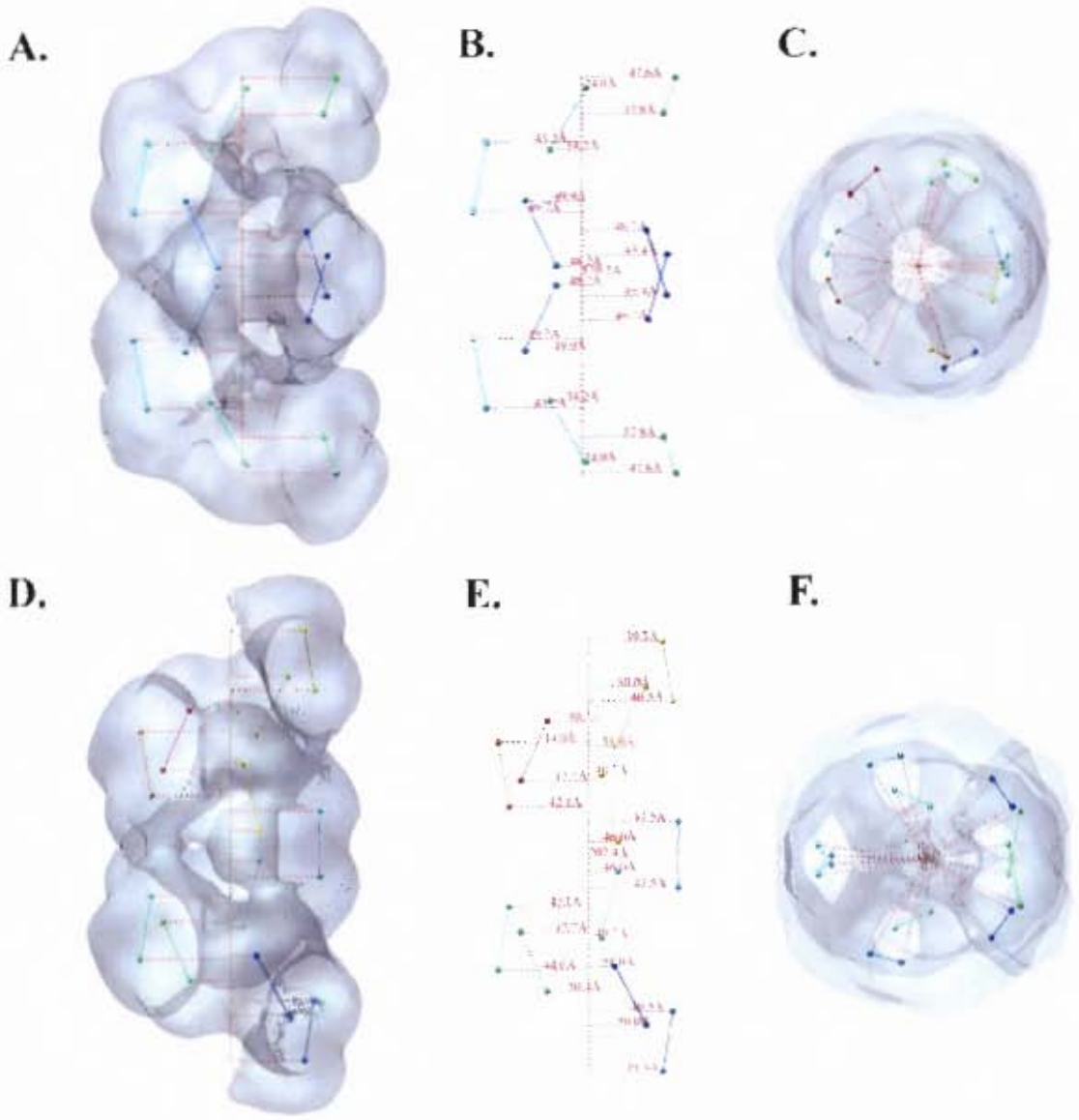

Figure 38. The Helical Radius of $CynD_{pum}$ at pH 8 and pH 6. Individual subunits were assigned a Center-of-gravity (COG) to calculate the helical radius at points along the spiral. **pH 8** (A, B, C), **pH 6** (D, E, F). **A/D**. Helical axis in electron density with subunit COGs represented as spheres. Note: each subunit is colored differently, COG-pairs the same. **B/E**. Helical radii of individual COGs (distances in Å). **C/F**. Head-on view of helical axis with electron density surrounding.

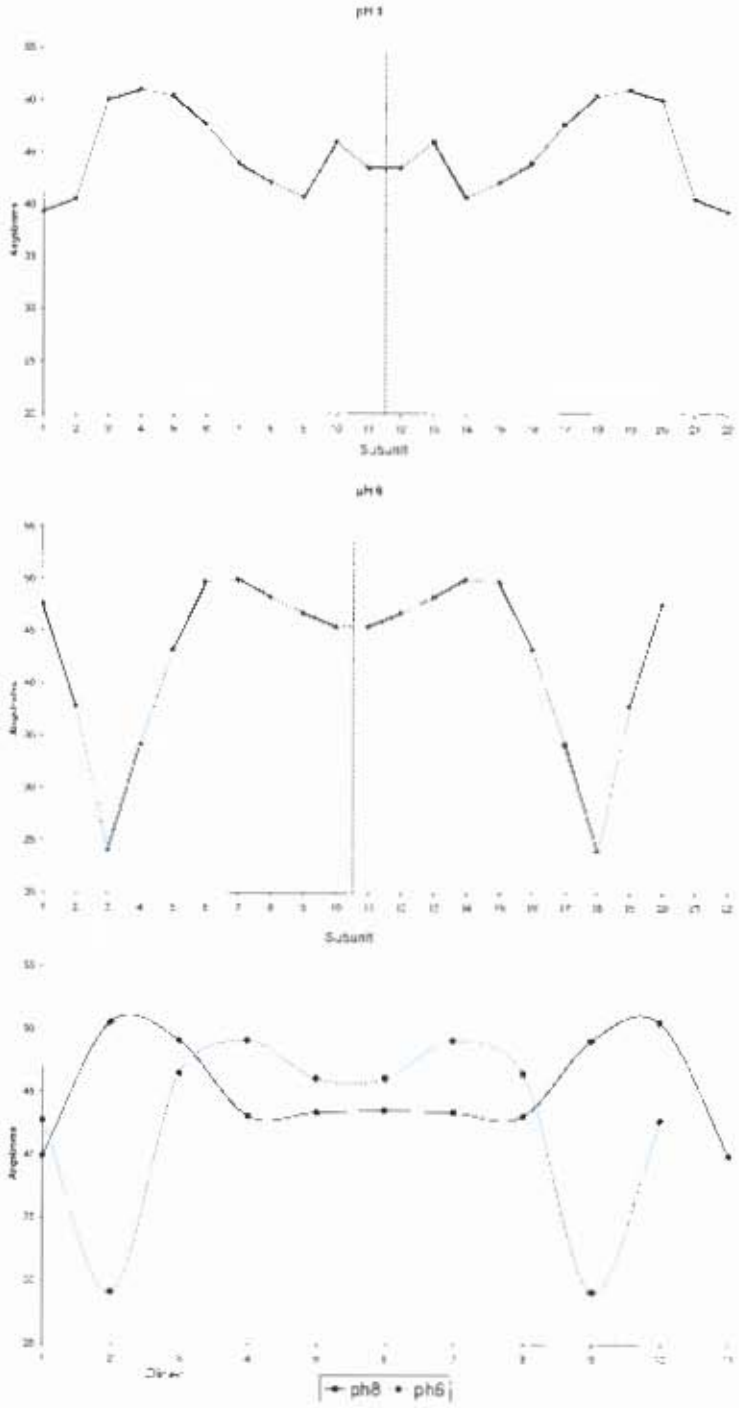

Figure 39. Helical radius of $CynD_{pum}$ at pH 8 and pH 6.

Top. The individual radii per subunit at pH 8. Penultimate subunits exhibit radial narrowing whilst the terminal subunits protrude away from the helical axis. Note: bold lines represent dimers; axis of symmetry is indicated with a central dashed line.

Middle: Terminal subunit 1 narrows significantly more than terminal subunit 20.

Bottom: Average radial distance per dimer, equivalent to point of 2-fold symmetry in the dimer (A-surface) Both reconstructions are similar and exhibit terminal narrowing.

3.6 Histidines and net-charge

Though $CynD_{pum}$ and $CynD_{stu}$ show ~70 % sequence similarity there are some key differences which could account for the vastly different oligomeric natures of the enzymes (Berman, 2003) **(fig. 17)**. *B. pumilus* has a number of histidines which are absent in *P. stutzeri* and are of interest due to the pKa of the histidine side chain which is ~6; the same as the pH transition point of the terminating spirals to extended fibres in $CynD_{pum}$.

These histidines could act by contributing an ionic force either by repelling or attracting to stabilise the helical form of the enzymes as they make the transition below pH 6 and become positively charged. These histidines are found in the C-terminal extension which has been repeatedly modelled as facing the open space in the middle of the fibre **(fig. 41)**. It is postulated therefore that the positively charged histidines repel each other in the fibre's hollow discouraging formation of the terminal E-interactions as a result of narrowing of the diameter of the fibres and tilting of the subunits. Thus, below pH 5 there would be space for the addition of additional dimers at the ends of previously terminating spirals and fibre formation would ensue.

To determine whether the presence of the histidines or lack thereof has an effect on the structure upon pH shift, the charges in the area of the C-terminal extension were calculated **(Table 3, fig. 40).** The data indicate that the histidines make the most significant contribution to the net charge of the inner core of $CynD_{pum}$ C1. Consequently, the net charge of this region in $CynD_{pum}$ C1 increases from 0 to +1.7 when pH is decreased from 8 to 6, whilst the charge of $CynD_{pum}$ 8A3 remains almost inert at 0. Further decreasing the pH to 5.4 results in a net charge increase to almost +3 in $CynD_{pum}$ C1 whilst pH 8 remains low at +0.32.

The termination of $CynD_{pum}$ elongation is most likely not a robust process and the terminal spiral subunits are thought to exist in a state of fragile equilibrium (Scheffer, 2006). This shallow energy well implies that the spirals require small environmental

changes to extend into fibres. The charge data supports this hypothesis as increased charge seems to correlate with fibre extension below pH 6 in C1. Both C1 and 8A3 appear to decrease in size from pH 8 to pH 6 then C1 extends into a fibre as the core charge exceeds a threshold value upon decrease to pH 5.4. As 8A3 lacks these histidines, the core charge remains close to 0 and no extension is observed. An undetermined cooperation of factors at various interaction surfaces which modifies the fragile termination equilibrium is most likely causing the spirals to shorten briefly upon pH decrease before extension. It is probable that fibres would increasingly extend as pH decreases until core-charge repulsion (and concomitant extension) is balanced out against the lack of charges in mandatory residues (e.g. Glutamate in the D-surface will have an effective neutral charge below pH 4.5 and gradually become ineffective as a contributor to salt-bridging).

Table 3: Charge distribution of the C-terminal extension of $CynD_{pum}$

	pH	C1	Quaternary structure	8A3	Quaternary structure
Histidines only	8	+0.03 (+0.01)	Spiral (18mer)	-1.00	Spiral (22mer)
3His (1His)	6	+1.68 (+0.56)	Spiral (14/18mer)	-0.97	Spiral (20mer)
(In 8A3 charge is calculated for the replacement residues)	5.4	+2.49 (+0.83)	Fibre	-0.89	Aggregate
	8	+0.01	Spiral (18mer)	0.00	Spiral (22mer)
Net charge on all residues	6	+1.72	Spiral (14/18mer)	+0.08	Spiral (20mer)
	5.4	+2.80	Fibre	+0.32	Aggregate

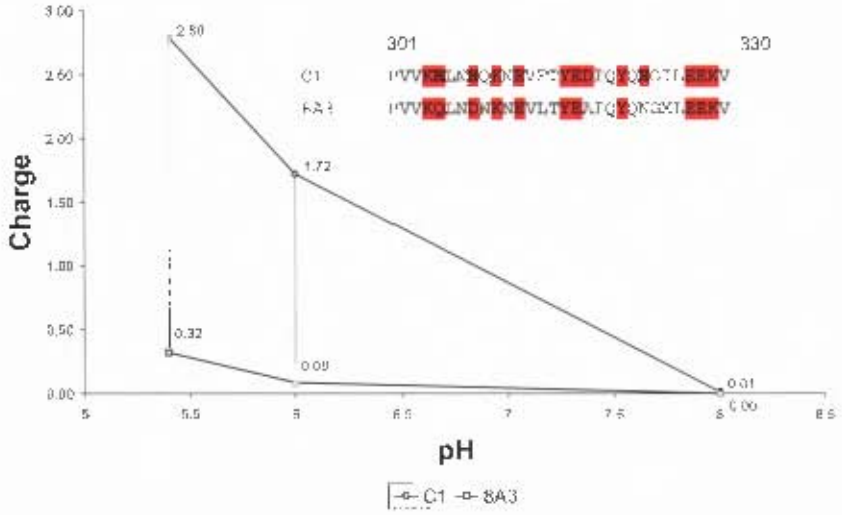

Figure 40. The net charge of the C-terminal extension at 3 pH values. The sequence that was selected is printed (charge-contributing residues highlighted in red, histidines bold):

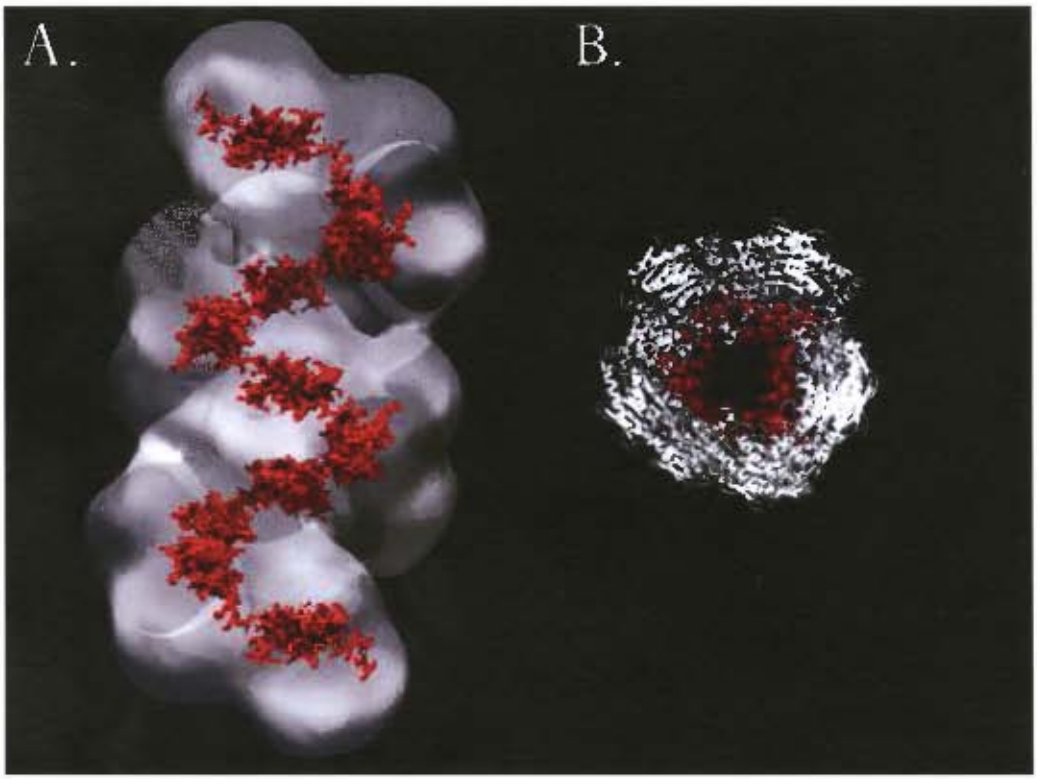

Figure 41. Core charged residues of $CynD_{Pam}$. Residues in the Carboxyl tail are highlighted in red. A. View perpendicular to the molecule. B. View down the core.

4. Conclusions

The structure of the Cyanide Dihydratase from *Bacillus pumilus 8A3* was elucidated at pH 8, 6 and 5.4 by the Single Particle Reconstruction method from Electron Micrographs. $CynD_{pum}$ from *B. pumilus* C1 is known to exist as an 18-mer (9 dimer) spiral at pH 8 and a 14/18-mer (7/9 dimer) spiral at pH 6 but spontaneously extends into a helical fibre at pH 5.4. $CynD_{pum}$ 8A3 exists as a 22-mer (11 dimer) spiral at pH 8 and a 20-mer (10 dimer) spiral at pH 6 but forms asymmetrical fibrous aggregates at pH 5.4.

Homology modelling reveals a number of surface interactions key to oligomerisation. The foremost of these is found between the subunits of an individual dimer and is known as the A-surface.

The second interaction is created by the C-surface between dimers and is the primary interaction necessary for extension. This surface is comprised of a mix of hydrophobic/hydrophilic interactions contributed by residues found in α2, α4, α8, α9 (with the subsequent loop), α10 (found in the unmodelled 2nd insertion sequence) and α11.

The D-surface is postulated as being essential for spiral/helix formation as the primary interaction across the helical groove. This interaction forms between a single dimer and a symmetry-related partner 5 dimers further along the spiral/helix. In 8A3 at pH 8, 3 potential D-interactions are identified of which one interaction is most likely artefactual. There exist 2 of each of the other D-interactions and these are primarily mediated by residues in α3. The pH 6 model shows 2 types of D-interaction of which D1 exists as a single and the D2 a double. Likewise, these interactions are mediated by α3. Helical modelling using SITUS and optimised docking to $CynD_{pum}$ C1 pH 5.4 produces a helix ($\Delta\varphi$ = -76.67, ΔZ = 15.87, r = 44.5 Å) which allows identification of a likely heterogeneity in the quaternary structure. Two alternate forms of D-surface are postulated, the High Correlation species (2 salt bridges) and the Low Correlation species (a single salt bridge). This is supported by Scheffer's results (Scheffer, 2006) and the fact

that deletion mutants of 8A3 which served to remove glutamine residues and disrupt the D-surface (and therefore oligomerisation) produced more regular helices by selecting against heterogeneity. Docking of the helical model of 8A3 into Van Der Vyver's reconstructed MB3799/MB3799-2 fibres yields a high correlation and allows identification of a potential stabilising interaction between L94 (α3) and D34' (α1').

The final interaction is termed D3 (previously the E-surface) and is formed by the terminal dimer. This surface is considered to contribute to termination by tilting and narrowing of the helical radius thus blocking the addition of another terminal dimer. Modelling shows D3 to be a slightly translated D-interaction which could include residues found in α2 and the loop joining α1 and α2. At pH 6 a second "E surface" is noticed (E2) adjacent to the initial one (between the terminal dimer and the dimer 4 places along). This interaction could be formed by residues K227 and D281 (α7), E327 and E328 (α11) and E155 (in the loop joining α4 and β7).

It is very likely that all three forms of D-interaction are merely translations of the same interaction due to an increase in $\Delta\varphi$ and a concomitant narrowing of the helical radius resulting in termination.

An increasing net charge of the unique C-terminal extension which lines the vacant core of $CynD_{pum}$ is shown to correlate with helical extension in C1. However, the net charge in 8A3 remains close to 0 due to the lack of key histidine residues. It is therefore proposed that histidines found in the C-terminal region of C1 become positively charged as pH decreases and exceed a threshold value in between pH 6 and pH 5.4 which causes them to repel each other. This cross-core repulsion concomitantly disrupts the formation of the E-surface interaction seen in the terminating spirals by preventing the narrowing of the helical radius and the subsequent tilting of terminal dimers.

5. References

Agarkar VB, Kimani SW, Cowan DA, Sayed MF, Sewell BT. 2006. The quaternary structure of the amidase from *Geobacillus pallidus* RAPc8 is revealed by its crystal packing. *Acta Crystallograph.Sect.F.Struct.Biol.Cryst.Commun.* **62:** 1174-1178.

Akcil A, Mudder T. 2003. Microbial destruction of cyanide wastes in gold mining: process review. *Biotechnol.Lett.* **25:** 445-450.

ATSDR. 2005 ATSDR CEP Site Count Report. **http://www.atsdr.cdc.gov/cep/05cep.html** . 2005a. Ref Type: Electronic Citation

ATSDR. 2005 CERCLA Priority List of Hazardous Substances. **http://www.atsdr.cdc.gov/cercla/05list.html** . 2005b. Ref Type: Electronic Citation

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

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

Baxter J, Cummings SP. 2006. The current and future applications of microorganism in the bioremediation of cyanide contamination. *Antonie Van Leeuwenhoek* **90:** 1-17.

Bellinzoni M, Buroni S, Pasca MR, Guglierame P, Arcesi F, De RE, 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 MN. 2003. *Quaternary structures of the cyanide dihydratases of Bacillus pumilus C1 and Pseudomonas stutzeri AK61.* M. Sc. Thesis. University of Cape Town.

Brady D, Dube N, Petersen R. 2006. Green Chemistry: Highly selective biocatalytic hydrolysis of nitrile compounds. *South African Journal of Science* **102:** 339-344.

Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB. 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8 A. *Nature* **371:** 578-586.

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

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

Chen JZ, Grigorieff N. 2007. SIGNATURE: a single-particle selection system for molecular electron microscopy. *J.Struct.Biol.* **157:** 168-173.

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

Commeyras A, Taillades J, Collet H, Boiteau L, Vandenabeele-Trambouze O, Pascal R, Rousset A, Garrel L, Rossi JC, Biron JP, Lagrille O, Plasson R, Souaid E, Danger G, Selsis F, Dobrijevic M, Martin H. 2004. Dynamic co-evolution of peptides and chemical energetics, a gateway to the emergence of homochirality and the catalytic activity of peptides. *Orig.Life Evol.Biosph.* **34:** 35-55.

Dailey FE, Macnab RM. 2002. Effects of lipoprotein biogenesis mutations on flagellar assembly in *Salmonella*. *J.Bacteriol.* **184:** 771-776.

Darling DC, Schroeder FC, Meinwald J, Eisner M, Eisner T. 2001. Production of a cyanogenic secretion by a thyridid caterpillar (*Calindoea trifascialis, Thyrididae, Lepidoptera*). *Naturwissenschaften* **88:** 306-309.

DeSantis G, Zhu Z, Greenberg WA, Wong K, Chaplin J, Hanson SR, Farwell B, Nicholson LW, Rand CL, Weiner DP, Robertson DE, Burk MJ. 2002. An enzyme library approach to biocatalysis: development of nitrilases for enantioselective production of carboxylic acid derivatives. *J.Am.Chem.Soc.* **124:** 9024-9025.

Dorset DL. 1995. *Structural Electron Crystallography*. Plenum Press, New York.

Ebbs S. 2004. Biological degradation of cyanide compounds. *Curr.Opin.Biotechnol.* **15:** 231-236.

Faull JL, Graeme-Cook KA, Pilkington BL. 1994. Production of an isonitrile antibiotic by an UV-induced mutant of *Trichoderma harzianum*. *Phytochemistry* **36:** 1273-1276.

Fischer H, Polikarpov I, Craievich AF. 2004. Average protein density is a molecular-weight-dependent function. *Protein Science* **13:** 2825-2828.

Frank J. 2006. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. New York: Oxford University Press.

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

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

Inghardt T, Johansson A, Svensson A. New Mandelic acid derivatives and their use as Thrombin inhibitors. PCT/SE02/01557[WO 03/018551 A1], 1-67. 2002. Ref Type: Patent

Jandhyala D, Berman M, Meyers PR, Sewell BT, Willson RC, Benedik MJ. 2003. CynD, the cyanide dihydratase from *Bacillus pumilus*: gene cloning and structural studies. *Appl.Environ.Microbiol.* **69:** 4794-4805.

Jandhyala DM, Willson RC, Sewell BT, Benedik MJ. 2005. Comparison of cyanide-degrading nitrilases. *Appl.Microbiol.Biotechnol.* **68:** 327-335.

Jinghua. Jinghua's cryo EM page. Website . 2007. Ref Type: Electronic Citation

Kimani SW, Agarkar VB, Cowan DA, Sayed MF, Sewell BT. 2007. Structure of an aliphatic amidase from *Geobacillus pallidus RAPc8. Acta Crystallogr.D.Biol.Crystallogr.* **63:** 1048-1058.

Kjeldsen P. 1999. Behaviour of Cyanides in Soil and Groundwater: A Review. *Water, Air, & Soil Pollution* **115:** 279-308.

Kousparou CA, Epenetos AA, Deonarain MP. 2002. Antibody-guided enzyme therapy of cancer producing cyanide results in necrosis of targeted cells. *Int.J.Cancer* **99:** 138-148.

Ludtke SJ, Baldwin PR, Chiu W. 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J.Struct.Biol.* **128:** 82-97.

Macqueen JB. 1966. Some methods for classification and analysis of multivariate observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley* **1:** 281-297.

Maksimov AI, Maksimova I, Kuznetsova MV, Olontsev VF, Demakov VA. 2007. [Immobilization of *Rhodococcus ruber* strain gt1, possessing nitrile hydratase activity, on carbon sorbents]. *Prikl.Biokhim.Mikrobiol.* **43:** 193-198.

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

Meyers PR, Rawlings DE, Woods DR, Lindsey GG. 1993. Isolation and characterization of a cyanide dihydratase from *Bacillus pumilus* C1. *J.Bacteriol.* **175:** 6105-6112.

Muller A, Weiler EW. 2000. IAA-synthase, an enzyme complex from *Arabidopsis thaliana* catalyzing the formation of indole-3-acetic acid from (S)-tryptophan. *Biol.Chem.* **381:** 679-686.

Mylerova V, Martinkova L. 2003. Synthetic Applications of Nitrile-Converting Enzymes. *Current Organic Chemistry* **7:** 1279-1295.

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

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

Nawaz MS, Billedeau SM, Cerniglia CE. 1998. Influence of selected physical parameters on the biodegradation of acrylamide by immobilized cells of *Rhodococcus sp. Biodegradation.* **9:** 381-387.

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

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

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

Pekarsky Y, Campiglio M, Siprashvili Z, Druck T, Sedkov Y, Tillib S, Draganescu A, Wermuth P, Rothman JH, Huebner K, Buchberg AM, Mazo A, Brenner C, Croce CM. 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.

Pettersen EF FAU - Goddard T, Goddard TD FAU - Huang C, Huang CC FAU - Couch G, Couch GS FAU - Greenblatt D, Greenblatt DM FAU - Meng E, Meng EC FAU - Ferrin T, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis.

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

Piotrowski M, Volmer JJ. 2006. Cyanide metabolism in higher plants: cyanoalanine hydratase is a NIT4 homolog. *Plant Mol.Biol.* **61:** 111-122.

Podar M, Eads JR, Richardson TH. 2005. Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study. *BMC.Evol.Biol.* **5:** 42.

Quek E, Ting YP, Tan HM. 2006. Rhodococcus sp. F92 immobilized on polyurethane foam shows ability to degrade various petroleum products. *Bioresour.Technol.* **97:** 32-38.

Rawls JM, Jr. 2006. Analysis of pyrimidine catabolism in *Drosophila melanogaster* using epistatic interactions with mutations of pyrimidine biosynthesis and beta-alanine metabolism. *Genetics* **172:** 1665-1674.

Roach PC, Ramsden DK, Hughes J, Williams P. 2004. Biocatalytic scrubbing of gaseous acrylonitrile using *Rhodococcus ruber* immobilized in synthetic silicone polymer (ImmobaSil) rings. *Biotechnol.Bioeng.* **85:** 450-455.

Robertson DE, Chaplin JA, DeSantis G, Podar M, Madden M, Chi E, Richardson T, Milan A, Miller M, Weiner DP, Wong K, McQuaid J, Farwell B, Preston LA, Tan X, Snead MA, Keller M, Mathur E, Kretz PL, Burk MJ, Short JM. 2004. Exploring nitrilase sequence space for enantioselective catalysis. *Appl.Environ.Microbiol.* **70:** 2429-2436.

Rogozin IB, Makarova KS, Murvai J, Czabarka E, Wolf YI, Tatusov RL, Szekely LA, Koonin EV. 2002. Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res.* **30:** 2212-2223.

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J.Mol.Biol.* **234:** 779-815.

Scheffer MP. 2006. *Helical structures of the cyanide degrading enzymes from Gloeocercospora sorghi and Bacillus pumilus providing insights into nitrilase quaternary interactions*. M. Sc., University of Cape Town.

Sewell BT, Berman MN, Meyers PR, Jandhyala D, Benedik MJ. 2003. The cyanide degrading nitrilase from *Pseudomonas stutzeri* AK61 is a two-fold symmetric, 14-subunit spiral. *Structure.* **11:** 1413-1422.

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

Smith JM. 1999. Ximdisp--A visualization tool to aid structure determination from electron microscope images. *J.Struct.Biol.* **125:** 223-228.

Sugawara T, Tanaka A, Imai H, Nagai K, Suzuki K. 1997. YM-47515, a novel isonitrile antibiotic from *Micromonospora echinospora subsp. echinospora. J.Antibiot.(Tokyo)* **50:** 944-948.

Tattersall DB, Bak S, Jones PR, Olsen CE, Nielsen JK, Hansen ML, Hoj PB, Moller BL. 2001. Resistance to an herbivore through engineered cyanogenic glucoside synthesis. *Science* **293:** 1826-1828.

Theis TL, Young TC, Huang M, Knutsen KC. 1994. Leachate characteristics and composition of cyanide-bearing wastes from manufactured gas plants. *Environmental Science & Technology* **28:** 99-106.

van Heijenoort J. 2001. Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology* **11:** 25R-36R.

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

Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35:** W407-W410.

Wriggers W, Milligan RA, McCammon JA. 1999. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J.Struct.Biol.* **125:** 185-195.