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

THE CRYSTAL STRUCTURE OF A MUTANT NITRILE HYDRATASE

SAMUEL KOJO KWOFIE

University of Cape Town

A **minithesis** submitted in partial fulfilment of the requirements for the degree of

Master of Science in the Molecular and Cell Biology Department,

Faculty of Science at the

University of Cape Town.

Supervisor: Prof. D. A Cowan

Co- supervisors: Prof. B. T. Sewell & Dr. Muhammed F-R Sayed

November 2006

CRYSTAL STRUCTURE OF A MUTANT NITRILE HYDRATASE

SAMUEL KOJO KWOFIE

KEYWORDS

Nitrile hydratase

Wild type

Mutant

Mutagenesis

Protein expression

Protein purification

Crystallization

X-ray crystallography

Normal mode analysis

Cavity calculations

Substrate specificity

Aromatic substrate

Aliphatic substrate

University of Cape Town

ABSTRACT

Nitrile hydratases can be used as industrial biocatalysts. They catalyze the conversion of nitriles to their corresponding amides. These industrial biocatalysts have recently gained attention due to the economic production of industrial commodities such as acrylamide and nicotinamide (Thomas *et al.,* 2002). Nitrile hydratases (NHases) are metalloenzymes made up of α and β subunits, and exist mostly as heterotetramers $(\alpha\beta)_2$. They are categorized into Co-containing and Fe-containing types depending on their co-factor requirements.

In this work, a βF55LF52G mutant of *Geobacillus pallidus* RAPc8 nitrile hydratase has been engineered for aromatic specificity and purified to homogeneity. Crystals with approximate dimensions of 0.3mm x 0.2 x 0.2mm, and which were grown in 1.1M sodium citrate and 0.1M Tris (pH 7.2), were used to collect diffraction data on an in-house X-ray source at a resolution of 2.5Å. The bravais lattice, space group and unit cell dimensions were determined as primitive tetragonal, $P4_12_12$ and a = b = 108.5Å, c = 81.15Å; $\alpha = \beta = \gamma = 90.00^{O}$ respectively. Molecular replacement was carried out via Molrep (CCP4 Suite, 1994), using the crystal structure of the wild type (Tsekoa, 2005) as the starting model. The correlation coefficient and R-factor were 0.63 and 0.373 respectively for the starting model. The mutant structure has been refined to R-factor of 0.236 and R_{free} of 0.343 using Refmac5 (Murshudov *et al.,* 1997) and O (Jones *et al.,* 1991).

Preliminary assessment of the mutant structure was done to gain initial insights into aromatic substrate specificity. Alignment of the mutant and wild type NHases showed no unusual side or main chain modifications. Normal mode analysis indicated the

University of Cape Town

presence of intrinsic flexibility within the mutant. This was consistent with previous observations made with other NHases (Bishop and Sewell, 2006)

The conformational changes created movements that may be described as shear motions within both the wild type and mutant NHases. Flip-flop substrate entry mechanism has been identified as one of the possible mechanisms for the entry of substrate into the active site of the mutant.

Though *Geobacillus pallidus* NHase has activity on aliphatic and heteroaromatic substrates, no activity has been observed on homoaromatic substrates (Pareira *et al.*, 1998). Other work done in parallel with this study has shown that the mutant engineered in this study has activity on benzonitrile with a K_m of 136 μm (Kowlessur, Unpublished data). Mutant active site pocket dimensions computed using CASTp Server (Liang *et al.*, 1998), indicated an increase in volume and surface area by 26% and 28% respectively. These pockets volume increments may be associated with homoaromatic substrate recognition.

University of Cape Town

DECLARATION

I declare that *Crystal structure of a mutant nitrile hydratase* is my own work, that it has not been submitted before for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged as complete references.

Samuel Kojo Kwofie November 2006

Signed: Signed by candidate

ACKNOWLEDGEMENTS

"My God loves me and His love will never end, He rest within my heart for my God loves me!" Oh, for a thousand tongues to sing my redeemer's praise. My God, I dedicate this work absolutely to you. Thank you my Blessed Virgin Mary for continuing to intercede on my behave. I constantly pray for Popes John Paul II and Benedict XVI for keeping the faith alive. To St. Ignatius de Loyola I say "ad morem dei gloriam". I also pray for Bishops Sam, Martin Darko and Essien for letting me know my God. To Fathers Benah, Michael Hagan and Paul Maher, continue to win souls for Christ. My grandmother Mary Amoah has been an eternal inspiration in my academic work, the candle that she burn will never die. I say thank you my grandma. I thank my Mum Agnes Aba Quarshie, Dad Samuel Kinglsey Kwofie, Aunt Theresa Quartey and Margin Enimil. God bless my Samuelle and Harriet for all the support through this work.

Many thanks to:

Faustie, Gifty, Uncle Obigie, Uncle Muda, Akeku, Paa Kwesi, Maame Bella, Naana Ngede, Daada Ben, Antie Mabel, Uncle Ataa, Bro Yaw, Maama Louisa, Brother Menya, Kweku Gyima, Kwesi, Laud Asibon, Antie Joana, Antie Soanke, Antie Nana, Uncle Paul and Sis. Emelia.

Structural Biology group: Serah Kimani, Itai Chitapi, Margo Scheffer, Felix Aduse-i-danso, James Onyemata, Jane Watermeyer, Jason, Tim, Ndoriah, Edwin, Takalani, Jeremy.

Electron Microscopy Unit: Sean, Miranda, Arvin, Muhammed Jafer and Ozlem.

University of Cape Town

Estiene (SANBI), Vicky (Bath), Terri Jett (Butler), Noncedo (UWC), Tumelo (UP), Brother Johnnie (UWC), Paul Abidemi (UWC), Rachel Dielle (UWC) and John Poole (UWC).

David Pugh (UWC) and Jasper Rees (UWC) for their tremendous support through this course.

This work is incomplete without the acknowledgement of ARCAM members especially Samson Muyanga, Heide, Jennifer, Parik, Okoh, William and Venod.

My appreciation goes to Tsepo Tsekoa for his continuous guidance through this work. My greatest appreciation goes to my supervisors Don Cowan, Trevor Sewell and Muhammed Sayed. This work would not have been possible without their mentorship and support. I now know structural biology through the inspiration of Trevor.

My final thanks go to the Carnegie Corporation of New York and the National Research Foundation of South Africa for funding my masters program.

University of Cape Town

Contents

University of Cape Town

University of Cape Town

University of Cape Town

University of Cape Town

Chapter 1: Literature review

1.0 INTRODUCTION

There are two main catalytic pathways for the metabolic conversion of nitriles into their corresponding carboxylic acid moeities. The first is the direct catalytic conversion of nitriles into their corresponding carboxylic acid moeities by the nitrilase. The second is the conversion of nitriles into their corresponding amides by nitrile hydratase and their subsequent catalytic conversion to carboxylic acids by amidase.

Application of enzymes for industrial bio-transformation has led to a significant increase in percentage production rate, low cost of production, minimal energy usage, and a decrease in harmful and toxic emissions (Thomas *et al.,* 2002; Shimuzu, 2001; Haveline *et al.,* 1998). The above reasons have led to a huge interest in the study of industrial enzymes such as nitrile hydratase. Mitsubishi Rayon of Japan has successfully used cells of *R. rhodochrous* J1 for the bio-catalytic conversion of acrylonitrile to acrylamide, whilst Lonza Guangzhou Fine chemicals in China has chemoenzymatically synthesized nicotinamide from 3-methyl-1,5-diaminopentane using *R. rhodochrous* J1 cells (Thomas *et al.,* 2002).

Most NHases are heterodimers and are made of non-homologous α and β subunits. The β-subunit is conventionally assigned as the largest in size and the molecular weights of both subunits range from 22 to 28 kDa. NHases are metalloenzymes containing either cobalt (Co) or iron (Fe) co-factors in their active sites. The iron

University of Cape Town

containing NHases are made up of a single low-spin nonheme Fe^{+3}, whilst the cobalt containing NHases have a single non-corrin Co^{+3}. The metallic co-factors of NHase are internally located between the αβ- dimer and the crystal asymmetric unit consists of a αβ dimer per one asymmetric unit (Huang *et al.,* 1997). The two αβ- dimers in the heterotetramers are bound to each other in two-fold axis symmetry (Kobayashi and Shimuzu, 199). From the crystal structures of NHases solved so far, their protein fold is conserved.

Though *Geobacillus pallidus* RAPc8 NHase is reported to have broad substrate specificity, it does not transform homoaromatic substrates such as benzonitrile and benzyl cyanide (Pereira *et al.,* 1998). Indeed benzonitrile acts as a non-competitive inhibitor (Tsekoa, 2005). Aromatic interactions between the aromatic residues around the active site of the enzyme and aromatic substrates may account for the lack of activity on homoaromatic substrates. This hypothesis constituted the background and reason for this work.

The aim of this work was to engineer the βF55LF52G mutant and determine the crystal structure to get insight into the interactions between aromatic residues and the aromatic substrates.

1.1 Nitrile-metabolizing Enzymes

Aldoxime metabolizing microbial strains converts aldoximes to their corresponding carboxylic acid derivatives (Fig. 1.1) in the presence of both aldoxime dehydratase and nitrile-hydrolyzing enzymes. Aldoxime dehydratase catalyses the conversion of

University of Cape Town

aldoximes into nitriles (Kato *et al.*, 2000). The nitrile derivatives are then converted into their respective carboxylic acid moieties by nitrile-hydrolyzing enzymes.

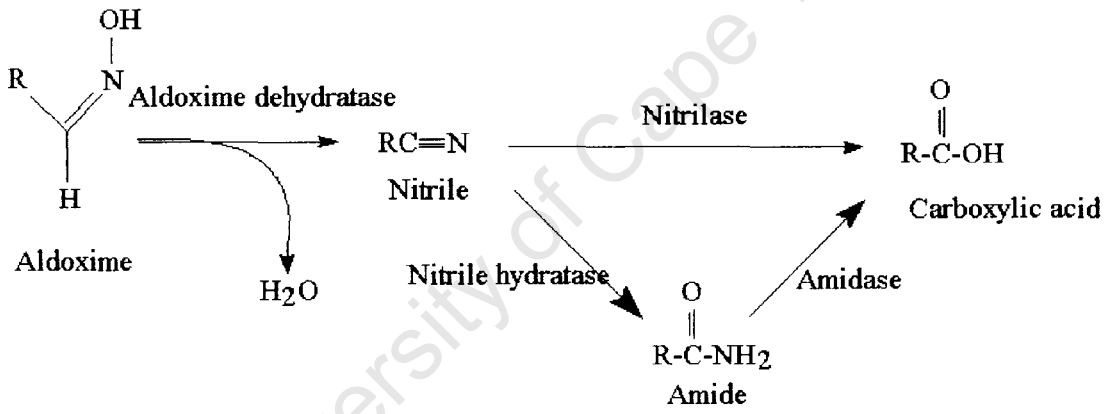


Fig. 1.1. Metabolic pathways for aldoximes via nitriles (redrawn from Kato *et al.*, 2000).

University of Cape Town

1.2 Mechanisms of reactions

NHases catalyse the hydration of nitriles to their corresponding amides. Three main reaction mechanisms (Fig. 1.2), where a metal ion act as a Lewis acid have been proposed (Huang *et al.*, 1997). (1) The nitrile substrate is activated for hydration after binding to the metal ion. Water molecule held in place by Y72 (Data not shown) subjects the nitrile carbon atom to nucleophilic attack (Fig. 1.2a). (2) The nitrile carbon is subjected to nucleophilic attack by metal-bound hydroxide ion (Fig. 1.2b). (3) Water activated by a metal-bound hydroxide subjects the nitrile carbon atom to nucleophilic attack (Fig. 1.2c). The slow rates of ligand exchange in trivalent metal ions make mechanism 3 the most probable model. Mechanism 3 does not involve ligand exchange (Huang *et al.*, 1997). The above reason was supported by similar rates of hydration of certain nitriles by some Co-type and Fe-type NHases (Mascharak, 2002).

University of Cape Town

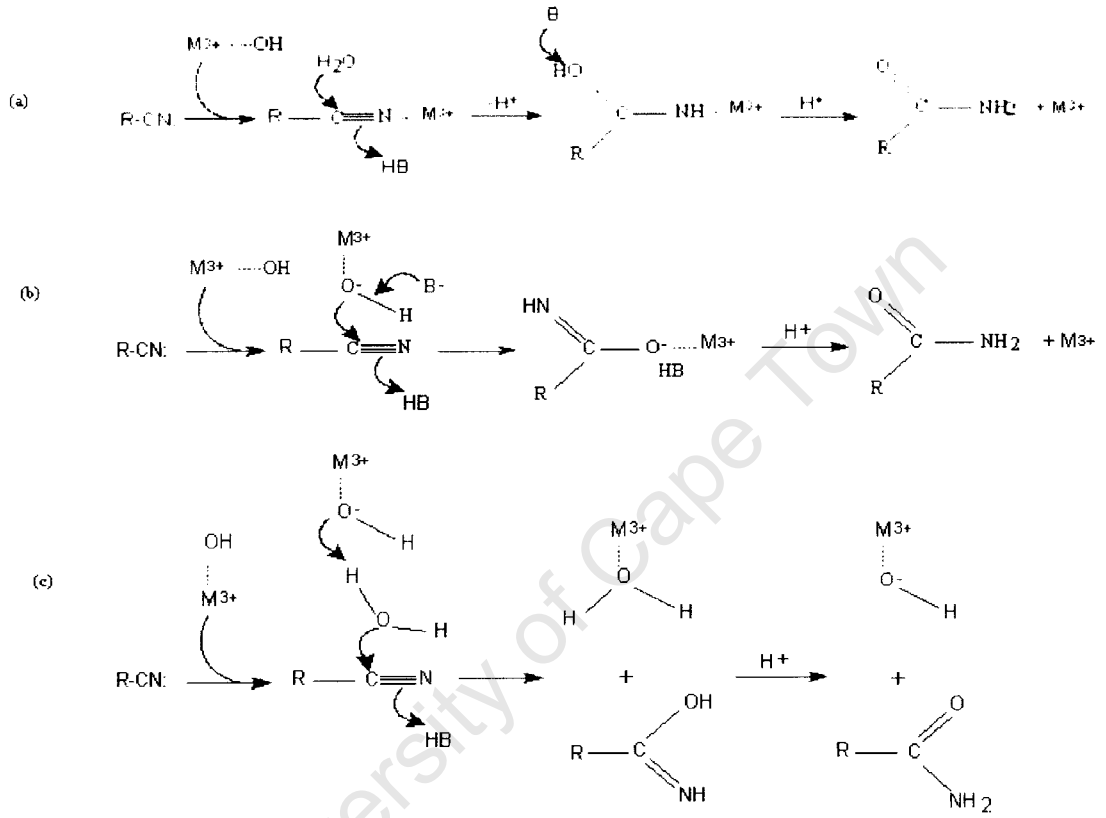


Fig. 1.2. NHase reaction mechanism: (a) direct binding of nitrile to metal ion (inner sphere model); (b) co-factor bound hydroxide ion subject nitrile carbon to nucleophilic (second inner sphere model); (c) activated water subject nitrile carbon to nucleophilic attack.

University of Cape Town

1.3 Molecular biology of NHases

1.3.1 Gene structure of NHases

The genes of nitrile hydratases consist of α and β subunits that are coded within 2 separate and adjacent open reading frames (ORFs). These two ORFs are normally separated by about 16 to 20 bases (Kim and Oriel, 2000), except in the case of *P. thermophilia* where the ORF overlap by 4 bases (Yamaki *et al.,* 1997). *Geobacillus pallidus* RAPc8, *B. smithii*, *P. thermophilia*, *R. rhodochrous* J1 and *Bacillus* sp. BR449 have their α-subunit ORF downstream the β-subunit ORF. (Cameron *et al.,* 2005; Takashima *et al.,* 1996; Yamaki *et al.,* 1997; Kobayashi *et al.,* 1997; Komeda *et al.,* 1996b; Kim and Oriel, 2000). Some NHase operons also have the α-subunit ORF upstream the β-subunit ORF. The genes coding for the α and β subunits do not differ much in size and ranges from 609 bp to 660 bp and 636 bp to 706 bp respectively.

NHases have been categorized into low-molecular (L-NHases) and high-molecular weight (H-NHases) due to variations in native molecular weights. L-NHases and H-NHases have native molecular weights which range from 30 kDa (Gilligan *et al.,* 1993) and 500-530 kDa (Kobayashi *et al.,* 1991) respectively. All NHases have been shown to be about 100 bp downstream the amidase gene with the exception of *R. rhodochrous* J1. *R. rhodochrous* J1 has α-subunit gene of the L-NHase upstream of the amidase by 1.9kb but the H-NHase gene does not have amidase close to it (Kobayashi *et al.,* 1992). There is an insertion sequence (IS1164) upstream the α and β subunits of H-NHase instead of an amidase gene (Kobayashi *et al.,* 1997; Komeda *et al.,* 1996b). Apart from the arrangement of the α and β subunit genes of

University of Cape Town

thermophilic NHase-producing organisms such as *B. smithii* (Takashima *et al.*, 1996) and *P. thermophilia* (Yamaki *et al.*, 1997), not much has been reported on the layout of their gene operons.

1.3.2 NHase operon of *Geobacillus pallidus* RAPc8

There are 8 putative ORFs in the gene operon of *Geobacillus pallidus* RAPc8 and a ninth (cbiO) is truncated at 3' *Bam*HI site of the pNHB4 insert (Fig. 1.3). The ORFs are arranged in this order: amidase; NHase β; NHase α, P14K accessory protein; a gene which is a homologue of 2Fe-2S class of ferredoxins; and 4 homologues of cobalt incorporating proteins in the order cibM, cbiN, cibQ and cbiO (Cameron *et al.*, 2005). There are no transcriptional terminators between the α and β-subunits of the above gene operons.

University of Cape Town

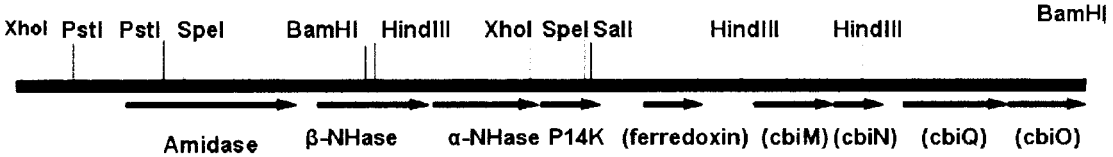


Fig. 1.3. Gene cluster of *Geobacillus pallidus* RAPc8 (Cameron *et al.*, 2005). The ORFs are represented by the closed arrows.

1.3.3 Activator Protein P14K

Cameron *et al.,* 2005 reported significant similarities between the sequences of P14K of *Geobacillus pallidus* RAPc8 and the β-subunit homologues of *Bacillus* BR449; NhhG; Nh1E; and both the low and high molecular weight *R. rhodochrous* J1 NHase operon. The P14K of *Geobacillus pallidus* RAPc8 has also been shown to have sequence similarities to P14K of *P. putida* 5B (Cameron *et al.,* 2005). Although the exact function of P14K has not been established, it is important for heterologous expression of *R. rhodochrous* J1 and *P. putida* 5B NHases in *E. coli* (Komeda, *et al.,* 1996; Kobayashi, *et al.,* 1991; Wu *et al.,* 1997; Komeda *et al.,* 1997).

P14K has been postulated to act as a catalyst or chaperone, since *P. Putida* P14K was expressed at a very low level during expression studies (Wu *et al.,* 1997). P14K may also aid in the incorporation of the cobalt into the active site (Vajda *et al.,* 2004). It has been postulated that the P14K acts specifically as a sub-unit catalyst and 4 possible mechanisms (Fig 1.4 A-D) have been proposed (Cameron *et al.,* 2005). The D mechanism was suggested as most probable because P14K had distinct homology with the β-subunit and the inability of P14K to activate NHase to form apoenzyme in the absence of cobalt.

University of Cape Town

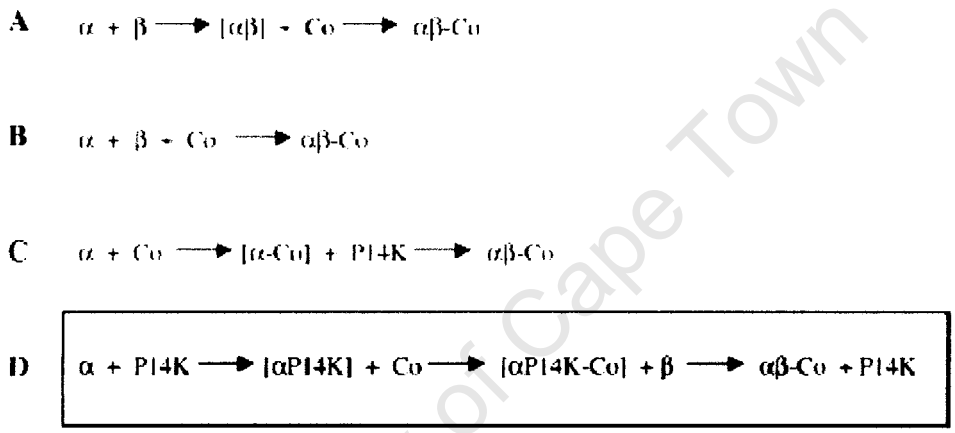


Fig. 1.4. Proposed mechanisms for NHase holoenzyme formation: (A) Formation of αβ apoenzyme followed by Co co-factor binding; (B) Uniform binding of α, β and Co co-factor; (C) binding of Co co-factor to folded alpha subunit, followed by β-subunit binding to form holoenzyme (D) P14K stabilizes α-subunit to aid Co binding, β-subunit then replaces the P14K (Cameron *et al.*, 2005).

University of Cape Town

1.4 Structure of NHases

The α and β monomeric subunits range from 22 to 28 kDa in size and the α-subunit is conventionally assigned as the smallest in size. NHase is a metalloenzyme which either contains cobalt or an iron. The Fe- types are made up of a single low-spin nonheme Fe^{+3} per αβ dimer, whilst the Co-types contain a single non-corrin Co^{+3} per αβ dimer (Kobayashi and Shimuzu, 1999).

Crystal structures of both Fe-types and Co-types have been solved. In all the reported structures, they are each made up of one αβ dimer per asymmetric unit but the crystal structures are heterotetramers made up of a αβ dimer which is bound to another αβ dimer across a crystallographic two-fold axis of symmetry. The αβ dimer is stabilized by the wrapping around of the N-terminal loops of the two monomers in opposite direction (Nakasako *et al.,* 1999). The interaction of the N-terminal loops suggests that the dimerisation is dynamic rather than simple docking (Nakasako *et al.,* 1999). The electrostatic charges around the interface of each monomer are different and the β monomer surface is negatively charged. About 50 waters of hydration are densely distributed over the β monomer. These water molecules may help in stabilizing the dimer by moderating the charges and decreasing the free energy of heterodimerization (Nakasako *et al.,* 1999).

University of Cape Town

1.4.1 Comparison between the Fe-types and Co-types NHases

One of the major differences between the Co-types and Fe-type NHases is the region β95 to β138. This region in Co-types contains an insertion of about 12 to 20 residues. This region has a α-helix (β111-β125) that interacts with another helix from the α-subunit (Miyanaga *et al.,* 2001). These α-helical interactions have been proposed to enhance the thermostability of cobalt containing NHases.

The active site residue βW72 (Tryptophan residue 72 of the β-subunit) in *P. thermophilia* is oriented in different direction to βY76 (Tyrosine residue 76 of the β-subunit) in *Rhodococcus* sp. N-771. These differences in orientations tend to make the active site channel slightly wider in Co-types than the Fe-type NHases. Residues βW72 and βY76 are conserved in all Co-types and Fe-type NHases respectively (Nagashima *et al.,* 1998) and may contribute to the differences in the substrate specificities between the types as suggested earlier on (Miyanaga *et al.,* 2001).

University of Cape Town

1.4.2 Active site of NHases

The co-factors are located at the central crevice at the interface between the α and β subunits. The ligands to the co-factors are provided primarily by the α-subunit. Both cobalt and iron atoms are octahedrally coordinated and are liganded to cysteines in the sequence V-C- (T/S)-L-C-S-C-. The cobalt and iron containing NHases differ in position 3 which is occupied by threonine and serine respectively. From the crystal structures of *Rhodococcus* sp. R312 (Huang *et al.,* 1997) and *Rhodococcus* sp. N771 (Nagashima *et al.,* 1998) NHases, the Fe^{+3} co-factor is octahedrally coordinated. The

vertices of the octahedron are occupied by the side chains of 3 cysteine residues and 2 main chain amide nitrogen atoms. The sixth vertex is occupied by a solvent exchange group (hydroxide or water) or nitric oxide (Scarrow *et al.,* 1996). In the structure of *Rhodococcus* sp. R312 NHase, the 3 side-chains of αC110 (Cysteine residue 110 of α-subunit), αC113 (Cysteine residue 113 of α-subunit), αC115 (Cysteine residue 115 of α-subunit) and the main chain amide nitrogen of αSer114 (Serine residue 114 of α-subunit), and αC115 formed the 5 vertices of the octahedron. And in the NHase structure of *Rhodococcus* sp. N-771 (Fe-type), αC112 (Cysteine residue 112 of α-subunit) and αC114 (Cysteine residue 114 of α-subunit) were postranslationally oxidised to cysteine sulphinic acid and cysteine sulphenic acid respectively. The metal binding motif V-C- (T/S)-L-C-S-C- is thus referred to as the claw-setting motif (Fig. 1.5). The post-translationally modified cysteine residues and the oxygen atom of serine in the metal binding motif induce a claw-setting of oxygen atoms which capture nitric oxide molecule.

Conflicting roles of the NHase co-factors have been reported, but the cobalt has been shown to be needed for protein activity rather than for correct folding or expression (Wu *et al.,* 1997).

University of Cape Town

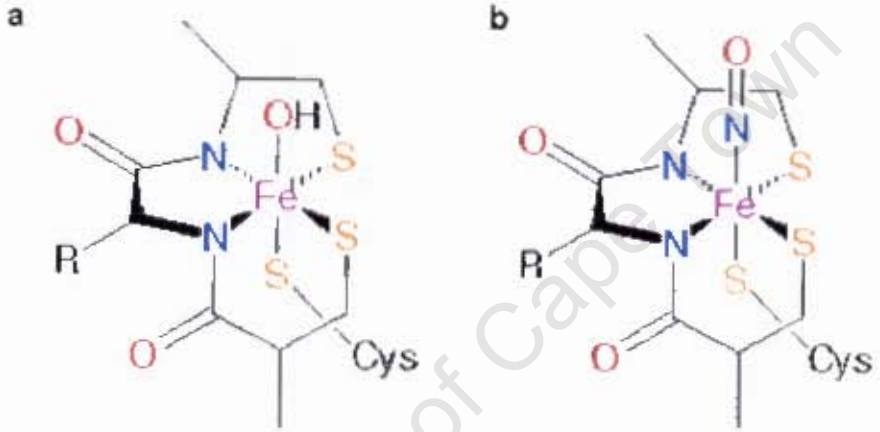


Fig. 1.5. The Claw-setting motif of (a) active hydoxylated and (b) inactive nitrosylated NHase from *Rhodococcus* sp. N771 (Diagram taken from http://metallo.scripps.edu/PROMISE/NITASE.html).

University of Cape Town

1.4.3 Nitric oxide (NO) activation

Electron-nuclear double resonance (ENDOR) analysis of *Rhodococcus* sp. R312 NHase indicated that a hydroxy or NO group could occupy the sixth coordination position of the Fe-cofactor (Mascharak, 2002). The NO is involved in the photoactivation of Fe-type NHases and was thought to be stabilized by the oxygen atoms of the cysteine residues involved in the claw-setting (Nagashima *et al.,* 1998). The photoactive NHases are inactive in the dark but upon light radiation they become active and may be due to light breaking the Fe-N (NO) bond. So far no Co-type NHase has been shown to exhibit photoactivity.

1.5 Industrial importance of NHases

The application of biocatalysts in chemoenzymatic processes has led to efficient production of important commodity materials such as acrylamide and nicotinamide. This is attributed to the increased recovery rate, drastic reduction in production cost and deleterious waste (Thomas *et al.,* 2002).

1.5.1 Application of NHases in biotransformation

Previously copper salts were used as catalyst in the industrial conversion of acrylonitrile to acrylamide but this proved to be costly and chemically ineffective (Kobayashi *et al.,* 1992). This deficiency led Mitsubishi Rayon of Japan and others to employ NHase of *R. rhodochrous* J1 for the enzymatic conversion of acrylonitrile to acrylamide. The same approach has been used by Lonza Guangzhou Fine chemicals (China) to chemoenzymatically synthesize nicotinamide from initial raw materials of

University of Cape Town

3-methyl-1, 5-diaminopentane, using NHase of *R. rhodochrous J1* as one of the enzymes in the stages of production (Thomas *et al.,* 2002).

Immobilized bed of *P. chlororaphis* B23 cells has also been used to successfully convert adiponitrile (ADN) to 5-cyanovaleramide (5-CVAM) for the production of azafenidin (Thomas *et al.,* 2002). This enzymatic process had a huge advantage over chemical catalysis, because the yield increased from 25% recovery rate to 93% (Hann *et al.,* 1999).

Kobayashi and Shimuzu in 1998 reported on the potential usage of NHases that exhibit regioselectivity, regiospecificity and substrate specificity in pharmaceutical production. The amide products resulting from the hydration of nitrile containing compounds could be used as raw materials in pharmaceutical production lines. To fully exploit the potentials of NHases in producing optically pure substances such as keto acids, amino acids and hydroxy acids, there is the need to use the NHase as pure enzymatic systems, immobilized cells or together with amidases (Nagasawa and Yamada, 1989). This concept has been employed in the hydrolysis of lactonitrile to DL-lactic acid using cells of *Rhodococcus* sp. R312 (Thompson *et al.,* 1988) and the stereospecific conversion of aminoproprionitrile to L-alanine by cells of *Acinebacter* sp (Macadam and Knowles, 1985).

University of Cape Town

1.5.2 Application of NHases in Bioremediation and biodegradation

There is growing concern about the accumulation of synthetic nitriles in water bodies and non-biodegradable synthetic polymers in the environment. Because of the high cost involved in treating these industrial wastes, cheaper bioremediation techniques are being exploited to drastically decrease the effects of these environmental hazards (Kobayashi and Shimuzu, 1998).

Two important methods have been shown recently to be effective in the bioremediation of acrylonitrile in industrial effluent (Cowan *et al.*, 2003). The first involve the use of NHases or Nitrilase (NTases) as components of the aqueous emulsion in the manufacture of raw rubbers and plastics. The NHases or NTases would then degrade enzymatically the acrylonitrile (Battistel *et al.*, 1997). The second involve the use of stable activated sludge made of lots of microbial cells including NHase-producing isolates to degrade acrylonitrile in large-scale industrial waste (Thompson *et al.*, 1988).

University of Cape Town

Cowan and others in 2003 reported the apparent lack of published work and under utilization of thermophilic enzymes in bioremediation and biotransformation. This deficiency could be attributed to 3 main reasons. (1) Successes chalked by the use of *Rhodococcus* biocatalysts have made it extremely difficult for their replacement. (2) The fact that thermostability alone cannot account for efficient industrial production, but also important factors such as yield and substrate specificity. And (3) thermophilic enzymes are relatively new in the scientific arena (Cramp *et al.*, 1997).

1.6 Thermostability of NHases

Most NHases characterised so far are thermolabile. Thermostability has been shown to correlate to the type of co-factor present in the NHase (Miyanaga *et al.,* 2004). Different methodologies were employed in determining the stability amongst NHases (table 1.1); therefore a general comparison cannot be made directly. Co-containing NHases are more thermostable than Fe-containing NHases. Amongst the five moderate thermophiles characterized so far, four are *Bacillus* isolates and the fifth is a *P. Thermophila* strain. N-butyric acid has been shown to be a potential stabilizing agent for some NHases by preventing the oxidation of modified cysteine sulphinic acid to cysteine sulphenic acid (Nagasawa and Yamada, 1987; Odaka *et al.,* 2001). It has been shown that thermophilic Co-containing NHases have similar structures and functions that are conserved (Cowan *et al.,* 1998). The stability of Co-type NHases may be attributed to the two α-helices that interact with each other (*Ps. thermophilia* NHase), but these helices are absent in the Fe-type NHases (Huang *et al.,* 1997; Nakasako *et al.,* 2001). The claw-setting motif at the active site of NHases is also thought to contribute to stability.

University of Cape Town

Microorganism	Characteristics	Stabilising agent	Temp (°C)	Incubation time	% Activity[b]	Reference
Bacillus sp. BR449	Thermophile	-	60	2 h	100	Padmakumar and Oriel (1999)
Bacillus sp. RAPc8	Thermophile	-	50	2.5 h	50	Pereira *et al.* (1999)
			60	16 min	50	
B. pallidus DAC521	Thermophile	-	50	51 min	50	Cramp and Cowan (1999)
			60	7 min	50	
Bacillus smithii	Thermophile	-	55	1.5 h	50	Takashima *et al.* (1998)
Ps. Thermophilia	Thermophile	34 mM n-butyric acid	50	2 h	100	Yamaki *et al.* (1998)
			60	2 h	90	
P. chlororaphis		22 mM n-butyric acid	20	10 min	100	Nagasawa and Yamada (1987)
			35	10 min	53	
P. putida (whole- cell assay)		-	50	10 min	0	Fallon *et al.* (1997)
P. putida		40 mM n-butyric acid	50	20 min	60	Payne *et al.* (1997)
Rhodococcus sp. N-774	Mesophile	-	30	30 min	100	Nagasawa and Yamada (1995)
R. rhodochrous J1 H-NHase	Mesophile	44 mM n-butyric acid	50	30 min	100	Nagasawa *et al.*, (1991)
			60	1 h	50	Kobayashi *et al.* (1992)
R. rhodochrous J1 L-NHase	Mesophile	-	30	30 min	100	Wieser *et al.* (1998)

Table 1.1. Thermal stability of NHases[a] (Cowan *et al.*, 2003). [a]Stability is defined by remaining activity following incubation for a specific time period at a specific temperature. [b]Remaining activity is 100% and there was reported drop of activity levels above the specified temperatures.

University of Cape Town

1.8 Substrate specificity

There is a general notion that the Fe-type NHases hydrate aliphatic nitriles while the Co-type NHases are specific for aromatic nitriles (Kobayashi and Shimuzu, 1998; Miyanaga *et al.,* 2004), but this trend is inconsistent with Table 1.1 (Cowan *et al.,* 2003). Fe-type NHase of *R. erythropolis* demonstrated broad substrate specificity, by hydrating aliphatic and heterocyclic nitriles (Duran *et al.,* 1993). It has also been reported that Co-type NHases of *Bacillus smithii*, *Geobacillus pallidus* DAC521, *Geobacillus pallidus* RAPc8 and *P. putida* showed no appreciable activity on (homo) aromatic nitriles (Payne *et al.,* 1997; Pereira *et al,* 1998; Takashima *et al.,* 1998; Cramp and Cowan, 1999).

The entrance to the active site channel of Fe-type NHases is narrow and may allow only aliphatic nitriles that are small to have access to the active site and this may explain the lack of aromatic specificity in Fe-containing NHases (Nakasako *et al.,* 1998; Miyanaga *et al.,* 2004). The spatial arrangement of residues around the catalytic cavity, and the displacement of phenylalanine and leucine in βW72 of *P. thermophilia* NHase (when compared to βY76 of *Rhodococcus* sp. N-771) have made the catalytic volume very larger. This could be one of the reasons why Fe-type and Co-type NHases have different substrate preferences.

A study of *P. thermophilia* NHase in relation to the size of substrates (Miyanaga *et al.,* 2004), showed similarities in kinetic parameters with low-molecular-mass Co-type *R. rhodochrous* NHase (Wiese *et al.,* 1998). Putative residues βL48 (Leucine residue 48 of the β-subunit), βF51 (Phenylalanine residue 51 of the β-subunit) and βW72 (Tryptophan residue 72 of the β-subunit) that are responsible for substrate

University of Cape Town

specificity in *P. thermophilia* NHase, were also found to be conserved in low-molecular-mass Co-type *R. rhodochrous* NHase (Miyanaga *et al.,* 2004). The presence of these residues may explain the similarities in kinetic parameters.

The kinetic parameters for the hydration of small aliphatic nitriles by high-molecular-mass Co-type *R. rhodochrous* NHase (Nagasawa *et al.,* 1991) were found to be similar to that of *P. thermophilia* NHase (Miyanaga *et al.,* 2004). The only exception was the K_m value which was 1000 times higher than that of high molecular mass Co-type *R. rhodochrous* NHase. It could be seen that the equivalent residues for βL48, βF51 and βW72 are tryptophan and serine residues respectively in high-molecular-mass Co-type *R. rhodochrous* NHase (Miyanaga *et al.,* 2004). The Fe-containing *Pseudomonas chlororaphis* B23 has been shown to minimally hydrate aromatic nitrile substrates (Nagasawa *et al.,* 1987). The equivalent residues for βL48, βF51 and βW72 in *Pseudomonas chlororaphis* B23 are valine, valine and tyrosine respectively. These residues are known to cause narrow substrate-binding pocket and are conserved in all known Fe-type NHases (Miyanaga *et al.,* 2001). The above findings tend to support the fact that different forms and sizes of hydrophobic pocket seem to influence substrate-binding preferences in NHases (Miyanaga *et al.,* 2004).

The hydroxyl moiety of βY68 (Tyrosine residue 68 of the β-subunit) in *P. thermophilia* NHase has been found to play a role in catalysis and substrate recognition. This role was attributed to the increased K_m value and a significant decrease in K_{cat} of the βY68F mutant compared to the wild type NHase (Miyanaga *et al.,* 2004). The above shows how mutagenesis may be used to understand substrate-binding preferences and also to identify which residues may be important in catalysis. N-butyric acid (organic acid) could probably inhibit NHases as imidate analogue

University of Cape Town

intermediate. βY68 may be playing a role in stabilizing this intermediate and also in claw-setting as it forms H-bond with αS112 (Miyanaga *et al.*, 2004).

Geobacillus pallidus RAPc8 NHase has no activity on benzonitrile and benzyl cyanide, which are homoaromatic nitriles but heteroaromatic nitriles such as 3-cyanopyridine can be considered as substrates (Pereira *et al.*, 1998). Benzonitrile has been proposed as an uncompetitive inhibitor of both the wild type and βW76G mutant (Tsekoa, 2005). Benzonitrile inhibition was 3-fold lower in the presence of βW76G mutation compared to the wild type. The decrease in inhibition was speculated to be the result of increase in active site pocket dimensions. The pocket volume and surface area of the wild type increased by 11% and 20% respectively when compared to the mutant.

The absence of activity on aromatic substrates may be explained by the aromatic interactions between aromatic substrates and the aromatic residues around the active site channel of the enzyme. One of the possible means of aromatic interactions reduction may be mutating these bulky aromatic residues into smaller and non-aromatic ones. This could give insight into aromatic specificity and guide future engineering of mutant NHases.

University of Cape Town

Substrate	Relative Activity[a] for							
	G. pallidus RAPc8	*G. pallidus* DAC521[b]	B. smithii	*R. rhodochrous* J1 H-NHase	*R. rhodochrous* J1 L-NHase	*Rhodococcus* sp. YH3-3	*R. erythropolis*	*P. chlororaphis*
Acetonitrile	100	28	540	605			5	2
Chloroacetonitrile	43	-		592	100		20	81
Acrylonitrile	67	37	390	476	15	156	45	100
Propionitrile	32	19	100	432	22	278	15	15
Methacrylonitrile	64	37	-	86	21	322	-	77
Butyronitrile	69	26	290	350	54	449	100	0.1
Isobutyronitrile	59	18	23	1	2	56.3	-	3
Valeronitrile	108	23	240	3	-	654	11	0
Isobutyronitrile	56	0		0	-		-	0
cis, trans-Crotonitrile	56	38	23	78	5	81.4	-	
	32	56						
Benzyl cyanide	0	-	-	13	-		-	0
Benzonitrile	0	0	1	68	17	226	46	0
3-Cyanopyridine	53	-	4	100	11	100	55	-
Malononitrile	7	0	-	-	-		-	-
Glutaronitrile	32	0	-	-	-		-	-
Adiponitrile	29	0	7	-	-		-	-

Table 1.2. Comparison of substrate specificities of various NHases (Cowan *et al.*, 2003): [a] Data taken from Pereira *et al.*, (1998), Cramp and Cowan (1999), Takashima *et al.*, (1998), Nagasawa *et al.*, (1991), Wieser *et al.*, (1998), Kato *et al.*, (1999), Duran *et al.*, (1993), and Nagasawa and Yamada (1987). [b]specificity of *B. pallidus* DAC521 NHase expressed as percentage substrate (50mM) initial concentration utilized over 30-min hydrolysis period.

University of Cape Town

1.9 AIMS

The aims of this work were:

1. To engineer the βF55LF52G mutant NHase.

2. Solve the crystal structure of the mutant.

3. Use the crystal structure to gain insight into the substrate specificity of NHase on aromatic substrates.

University of Cape Town

Chapter 2: Materials and Methods

2.1 Preparation of competent *E. coli* cells

2.1.1 Electrocompetent cells

Electrocompetent cells of *E. coli* DH5α were prepared according to the method described in the molecular cloning manual (Sambrook and Russell, 2001). These *E. coli* cells were used for transforming plasmid DNA during site-directed mutagenesis (Section 2.3.2).

2.1.2 Chemically competent cells

Chemically competent cells of *E. coli* BL21 (DE3) were prepared following the protocol in the Molecular Cloning Manual (Sambrook and Russell, 2001) as described by Inoue and others in 1990. These *E. coli* cells were used for transforming plasmid DNA during protein expression (Section 2.4.1).

University of Cape Town

2.2 Analytical Procedures

2.2.1 Agarose gel Electrophoresis

Fragments of DNA were separated according to size by Agarose gel electrophoresis, stained with ethidium bromide and visualized under UV light at wavelength of 365nm (Sambrook and Russell, 2001).

2.2.2 SDS-polyacrylamide gel electrophoresis

Proteins were separated using 15% SDS PAGE (Laemmli, 1970). Gels were stained in Coomassie Blue solution (0.2% Coomassie Blue R-250, 40% methanol and 10% acetic acid) for 3h and destained overnight in SDS PAGE destaining solution (5% methanol and 7.5% acetic acid).

2.3 Site-Directed mutagenesis

2.3.1 Mutagenesis

βF55L mutant Plasmid acquired from ARCAM (Tsekoa, 2005) and designed primers were used for in vitro site-directed mutagenesis. The sense primer (βF55F52GS) is (gattggggatgaaggccggtgatgaactcaggatcg) and the antisense primer (βF55F52GA) is (cgatcctgagttcatcaccggccttcatccccaatc). The antisense primer is the reverse complement of the sense primer. The cycling parameters for the PCR are indicated in Table 2.1. Stratagene's QuikChange Site-directed Mutagenesis kit was used according to the manufacturer's instruction.

University of Cape Town

Segment	Cycles	Temperature (°C)	Time (seconds)
1	1	95	30
2	12	95	30
		55	60
		68	90

Table 2.1. Cycling parameters for the site-directed mutagenesis PCR.

2.3.2 Digestion with restriction endonuclease and Transformation

Dpn I endonuclease was used to digest the parental DNA template and the desired mutated plasmid was selected. The Dpn I endonuclease is specific for the digestion of hemimethylated and methylated plasmid DNA. This enables the mutated DNA to be selected since the parent DNA would be digested. The desired mutated plasmid was then transformed into electrocompetent *E. coli cells* of DH5α

2.3.3 Isolation of plasmid DNA

5 ml LB culture was inoculated overnight with *E. coli* cultures containing the parent and desired mutant plasmid DNA, the appropriate antibiotic was added to the culture media. Plasmid DNA was extracted from these cultures using the Plasmix Mini Prep kit, following the manufacturer's instruction.

University of Cape Town

2.3.4 Mutant screening

Colony PCR of the double mutant from the *E. coli* culture of DH5α was done with NHOP1 (GCTCATATGAACGGTATTCATGATGTTGG) and NHOP3 (GCAGCGGCCGCCTAACCTACCGTAACTTTAGG) primers acquired from ARCAM to amplify the α and β genes of NHase using Taq polymerase. The cycling parameters for the PCR are indicated in Table 2.2. The PCR product was run on 1% agarose gel and the gel band with the desired size under UV light was cut and purified according to the Easy Fast and Flexible DNA purification protocol. Restriction digest with HAE III was done with the purified plasmid DNA and parental DNA extract. The digested products were run on 2.5% agarose gel to verify the sizes of the digested

DNA fragment. The βF55LF52G Mutant was sequenced by Inqaba Biotech Company to verify the success of the mutation.

No. of cycles	Temperature (°C)	Time
1	95	5 minutes
25	95	30 seconds
	55	30 seconds
	68	1.5 minutes
1	72	7 minutes

Table 2.2. Cycling parameters for the colony PCR.

University of Cape Town

2.4.1 Protein expression

Geobacillus pallidus RAPc8 NHase was expressed recombinantly in *E. coli* BL21 (DE3) and transformation of cells was carried out with pNH14K expression vector. The pNH14K is a derivative of pET21a (+) which carry the genes encoding the βF55LF52G-NHase, α-NHase and P14K proteins. The gene operons coding for the βF55LF52G -NHase, α-NHase and P14K proteins are separate, adjacent and do not overlap. The gene expression is regulated by IPTG inducible T7lac promoter of pET21a (Cameron *et al.*, 2005).

800 ml culture of LB containing 70 μg/ml ampicillin was grown at 37°C and 220 rpm until an optical density (OD) of 0.4 measured at 600 nm was achieved. At this OD, expression was induced with 0.4 mM IPTG whilst 0.1 mM cobalt chloride was added 20 min before induction. 5h after induction, cells were harvested and washed in 0.05 M potassium phosphate buffer at pH 7.2.

2.4.2 Cell free extracts preparation

Recovered cell pellets were resuspended in 25 ml 0.05 M potassium phosphate buffer at pH 7.2 and were subjected to a cycle of freezing at -80°C and thawed at 37°C for 10 min. The cells were then kept on ice for 30 min, after which they were sonicated briefly. The cell lysate was centrifuged at 5000 x g for 20 min and the supernatant crude protein was collected. The crude protein was saturated with 20% solid ammonium sulphate and left on ice overnight at 4 °C. The saturated supernatant was centrifuged at 7000 x g for 30 min at 4 °C to remove the precipitated proteins.

2.4.3 Protein purification

2.4.3.1 Hydrophobic Interaction chromatography (HIC)

The supernatant obtained was then loaded onto a HighLoad 16/10 Phenyl-Sepharose column (Amersham Biosciences) which has been equilibrated with Buffer A (1.0 M ammonium sulphate, 50 mM potassium phosphate; pH 7.2). Linear gradient of decreasing ammonium sulphate concentration which was generated with 50 mM potassium phosphate buffer; pH 7.2 (5 column-volumes, 1.0 M–0 M ammonium sulphate) was used to elute bound proteins (Tsekoa *et al.*, 2004).

University of Cape Town

2.4.3.2 Anionic exchange chromatography (IEC)

Pooled fractions from HIC were buffer exchanged with Buffer B (25 mM potassium phosphate; pH 7.2) using Vivaspin 15R 5000 mwco protein concentrators and were loaded onto a HighLoad 26/10 Q-Sepharose FF column (Amersham Biosciences) which had been equilibrated with Buffer B. Linear gradient of increasing NaCl concentration generated with Buffer C (500 mM sodium chloride, 25 mM potassium

phosphate; pH 7.2) (5 column volumes, 0 M–500 mM sodium chloride) was used to elute bound proteins.

2.4.3.3 Concentration of pure protein

Pooled fractions from anionic exchange chromatography were dialysed against 20 mM Tris (pH 7.2), filtered through 0.22-μm filter and concentrated to 30 mg/ml using Vivaspin 15R 5000 mwco. The concentration of the sample was measured spectrophotometrically at a wavelength of 280 nm using the Beer-Lambert's law (A = ECL). The A, E, C and L correspond to the absorbance, molar extinction coefficient ($M^{-1}cm^{-1}$), concentration (M) and sample path length (cm) respectively. The molar extinction coefficient of NHase was determined as 78330 M^{-1} cm^{-1} using the protein sequence via the ExPASy Server (Gasteiger *et al.*, 2005). A cuvette with 1 cm path length was used in the spectrophotometric assay. Samples of protein from various stages of purification and expression were analyzed for NHase using 15% SDS PAGE.

University of Cape Town

2.5.1 Protein crystallization

Geobacillus pallidus RAPc8 NHases have been shown to crystallize under PEG400 and sodium citrate conditions (Tsekoa *et al.*, 2004). These conditions were optimized for the crystallization trials. Crystallization trials were setup using the hanging drop vapour diffusion method under room temperature. Separate conditions of PEG400 and sodium citrate were used for the trials. 30% PEG400, 100 mM magnesium chloride and 100 mM MES ((2 [*N*morpholino] ethanesulphonic acid); pH 6.5) were used for

PEG400 crystallization trials. 0.8 M, 1.1 M and 1.3 M Sodium citrate concentrations in 0.1 M Tris (pH 7.2) were used for sodium citrate trials. Protein concentrations of 15mg/ml, 20mg/ml and 30 mg/ml were used for each trial. 2 μl protein and 2 μl well solutions were mounted on siliconized glass cover slips.

2.6 X-ray diffraction data collection and processing

2.6.1 X-ray diffraction data collection

X-ray diffraction data was collected on the X-ray machine at the University of the Western Cape [consisting of Rigaku RUH3R copper rotating-anode X-ray source operated at 40 kV, 22 mA; a Rigaku R-axis IV+ image plate camera; an X-stream 2000 low-temperature system; and an AXCO PX50 glass capillary optic with a 0.1 mm focus]. A crystal grown from sodium citrate conditions (1.1 M sodium citrate, 0.1 M Tris, pH 7.2, 15 mg/ml protein) was mounted on a cryoloop under cryoconditions at a temperature of 100K with a crystal-to-detector distance of 160 mm. 320 oscillation frame images were collected at a starting angle of 15^{o} at 0.5^{o} range for every 10 min.

University of Cape Town

2.6.2. X-ray data processing

2.6.2.1 Auto-indexing, integration, scaling and merging

Auto-indexing, integration, scaling, merging and space group determination were done with d*TREK (Plugrath, 1999).

2.6.2.2 Molecular replacement and refinement

Molecular replacement was done with Molrep (CCP4 suite, 1997) using the crystal structure of wild type *Geobacillus pallidus* RAPc8 NHase (Tsekoa, 2005) as the starting model. A series of both restrained and rigid refinement were done using refmac5 (Murshudov *et al.,* 1997). Molecular viewing and manual model building were done using O (Jones *et al.,* 1991). Water molecules were added to the model automatically using ARP/wARP (Zwart *et al.,* 2002) and manually using O (Jones *et al.,* 1991). Water molecules were also tidied and rationalised using WATERTIDY and DISTANG (CCP4 suite, 1997).

Structure validation was done using the Rampage server (Lovell *et al.,* 2002). Model quality was also assessed stereochemically using PROCHECK (Laskowski *et al.,* 1993). The quality of the fit between the model and the experimental data was assessed using SFCHECK (Vaguine *et al.,* 1999). The model structure was uploaded onto the Protein - Protein Interaction Server (Jones and Thornton, 1996) to determine which outlier residues in the Ramachandran plot are in the model structure interface.

University of Cape Town

2.7 Model assessment

Structural alignments were carried out using ALIGN (Cohen, 1997) and PyMOL (Delano). Unless otherwise stated all macromolecular structures were visualised using PyMOL (Delano). Normal mode analysis (NMA) of all structures was done using the Web-server for Normal Mode Analysis of proteins - v2.0 (Hollup *et al.,* 2005). NMA was used to determine the vibrational modes with the least frequencies: such modes describe the largest domain movements and show the intrinsic flexibility within the

molecules. Pockets in the model structures were identified and analysed using CASTp Server (Liang et al., 1998) and PyMOL (Delano).

The University of Houston Brownian Dynamics (UHBD) program suite was used to compute electrostatic energies of the mutant and wild type NHase structures. UHBD was also used to compute energies due to dimerisation. The grid, ionic strength, protein and solvent dielectric constants were set at 0.5 Å, 50 mM, 12.0 and 78.4 respectively. Prior to the calculation of the energies, protons were built into the NHase structures using the WHAT IF server (http://swift.cmbi.kun.nl/WIWWWI/).

University of Cape Town

Chapter 3: Protein engineering and purification

3.1 Protein Engineering

3.1.1 Site directed mutagenesis

Site-directed mutagenesis is used to gain insight into protein structure and functions, gene expression, and for vector modification. The aim of mutation F52G was to replace the aromatic phenylalanine to smaller and uncharged glycine residue, in such a way that the protein fold, stability and active site conformation may not be disrupted (Figure 3.1). From the electron density map of the F55L mutant (Tsekoa, 2005), the F52 residue was flexible and as such could block the entry of bulky aromatic substrates gaining access to the active site. And also the enzyme has no activity on homoaromatic substrates which may be due to aromatic interaction between the aromatic residues surrounding the entrance of the active site cavity and the bulky aromatic substrates.

University of Cape Town

Site-directed mutagenesis (Fig. 3.2) was carried out using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) and the methodology followed the manufacturer's instructions. The mutagenic primers (Section 2.3.1) used for the PCR introduced new restriction sites into the engineered double mutant DNA so that the mutagenic products could be effectively screened through restriction analysis. The mutated genes were amplified (Section 2.3.4) using NHOP1 and NHOP3 primers and digested with HAE III to screen for the presence of the desired mutated genes. Figures 3.3 and 3.4 demonstrate the success of the mutation. The sizes of the bands in Figure 3.3 were computed with AlphaEaseFC software

(http://www.labtrade.com/alpha/alpha ease fc.htm) based on that of Lambda DNA/Pst I maker. The sizes of the fragments were approximately 288 bp, 454 bp, 670 bp and 1400 bp. The addition of fragments 288 bp, 454 bp and 670 bp is 1412 bp, which is closer to the largest fragment of about 1400 bp, so it could be concluded that the largest restriction fragment may be as a result of incomplete digestion. The 2.5% agarose gel made it difficult to resolve fragments above 1000 bp, although AlphaEaseFC software was used, the size of the largest fragment may not be accurate. The sense mutagenic primer aligned with the sequenced gene (Fig. 3.4) to show the success of the introduction of the mutation into the gene. The mutation was also verified in the sequence of the gene (Data not shown). The electron density of the model protein structure (Figures 4.4 and 4.5) also confirmed the success of the mutation (Section 4.3.1).

Fig. 3.1 Alignment of β-subunits of both wild type and βF55LF52G mutant NHase sequences using BioEdit (Hall, 1999). Identical amino acid positions are denoted by dots. The mutations are indicated in amino acid positions F52G (New) and F55L (existing).

University of Cape Town

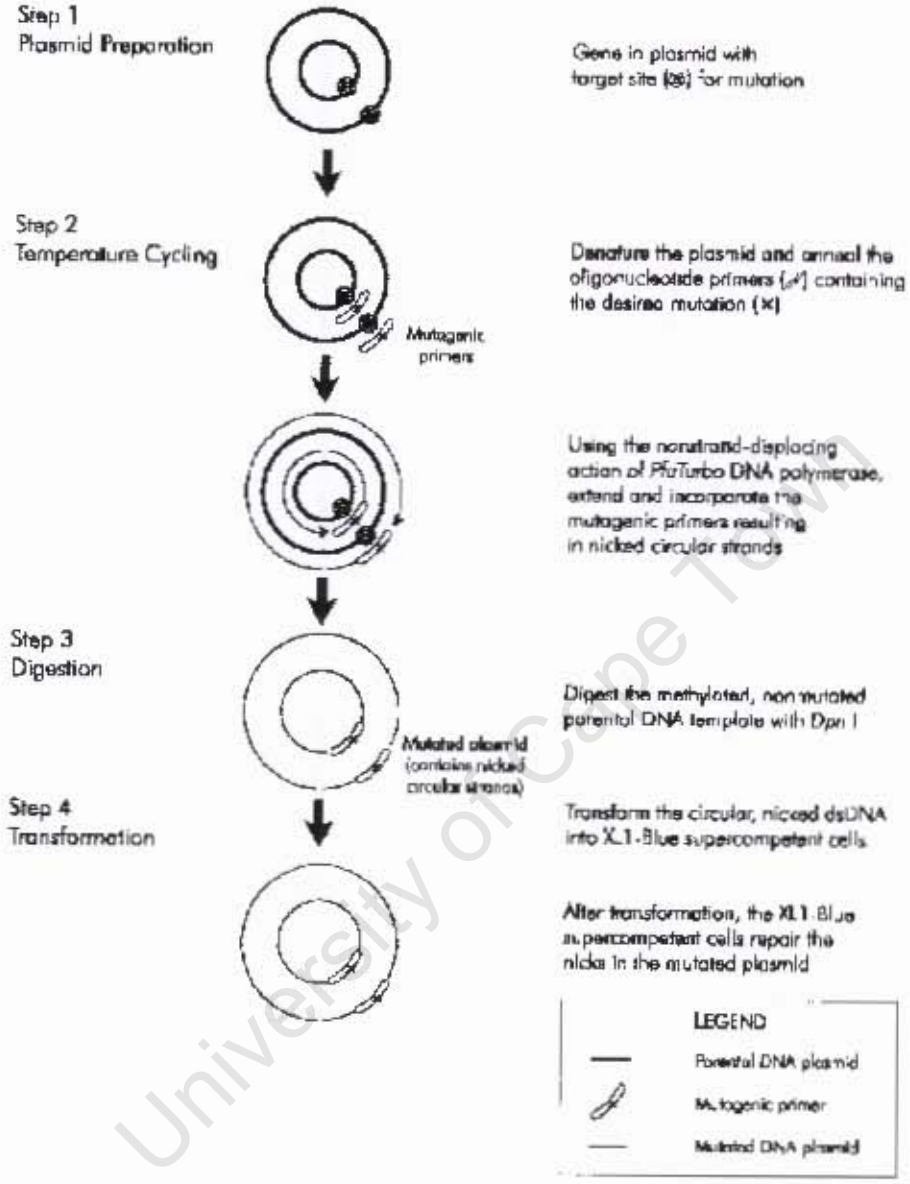


Fig.3.2: A general diagram of Quik-Change site-directed mutagenesis method (Stratagene QuikChange manual)

University of Cape Town

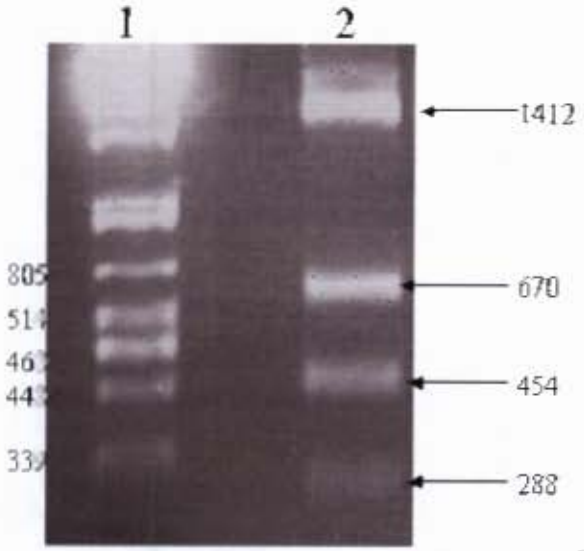


Fig. 3.3. Restriction digest of βF55LF52G with HAEIII. Lane 1: Lambda Pst I maker; 2- restriction fragments corresponding to 1412 bp, 670 bp, 454 bp and 288 bp.

University of Cape Town

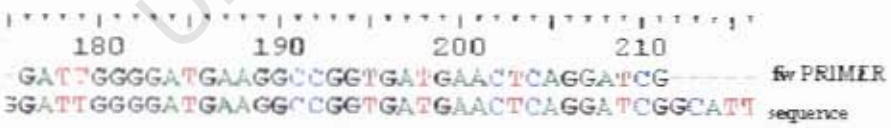


Fig. 3.4 Alignment of sense primer (fw PRIMER) and sequence of mutant gene (sequence) using BioEdit (Hall, 1999) to confirm successful mutation.

3.2.1 Protein expression and purification

Proteins must be purified to homogeneity before successful crystallization can be achieved. The essence of protein purification is to achieve conformational (purity > 90%) and molecular homogeneity to attain proper folding.

The pNH14K expression vector containing the gene coding for the α and β F55LF52G was isolated (Section 2.3.3) and transformed into the *E. coli* BL21 (DE3) expression host (Section 2.4.1). Since the mutant being expressed had two aromatic residues replaced with glycine and leucine residues respectively, the thermal stability may have been compromised based on UHBD program analysis (Section 2.7). A predicted decrease in stability by 2.9 kcal/mol in the mutant NHase (Sections 2.7) was computed using the University of Houston Brownian Dynamics (UHBD) program suite.

Efficient protein precipitation (Section 2.4.2) was achieved by leaving the ammonium sulphate saturated crude protein overnight on ice. SDS PAGE analysis of the crude protein (Section 2.4.2 and Figure 3.4) confirmed the over expression of the mutant against a large background of contaminant proteins which necessitated the need to further purify by hydrophobic interaction chromatography (2.4.3.1). SDS PAGE analysis and chromatograph (Figures 3.5, 3.6 and 3.7) from Pooled HIC fractions showed the removal of the majority of contaminant protein. The dialysed fractions from HIC were purified using the anionic exchange chromatography (2.4.3.2). The pooled fractions from (AEC) gave a single sharp peak (Fig. 3.8) corresponding to the molecular weight of native protein (51kDa) and confirming the chemical homogeneity; and two clear protein bands (Fig. 3.5) corresponding to

University of Cape Town

electrophoretically pure α and β subunits. Thus there was no need to further purify by gel filtration step before crystallization.

University of Cape Town

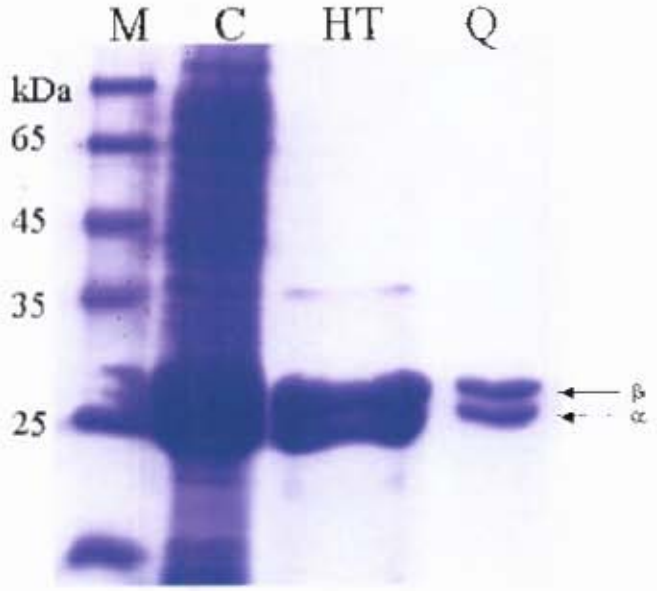


Fig. 3.5. SDS-PAGE gel showing Lane M: molecular weight markers; lane C: crude extract; lane HT: pooled fractions from Phenyl-Sepharose hydrophobic interaction chromatography; lane Q: pooled fractions from Q Sepharose ion exchange chromatography.

University of Cape Town

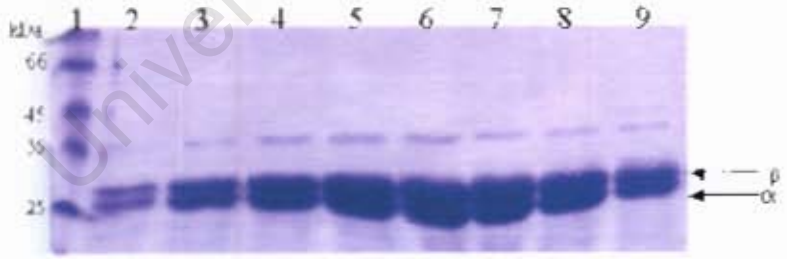


Fig. 3.6. SDS-PAGE gel of HIC fractions. Lane 1: molecular weight markers; Lanes 2-9: HIC fractions.

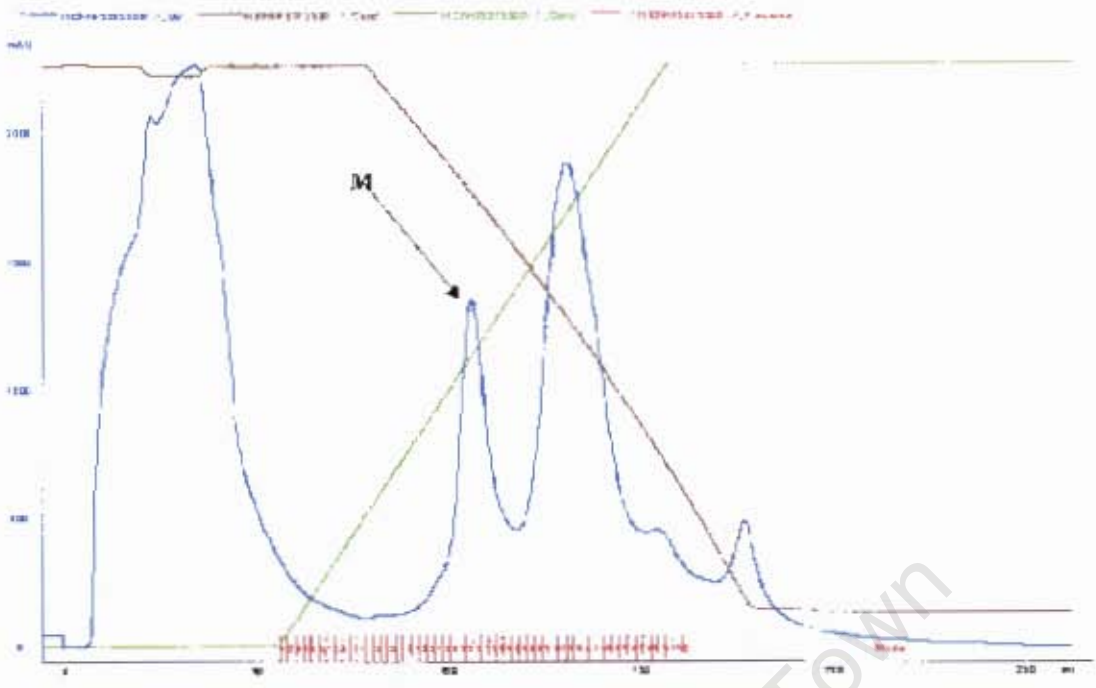


Fig. 3.7. Chromatograph of Phenyl Sepharose hydrophobic interaction chromatography. M: peak corresponding to NHase. The SDS PAGE of the pooled fractions from peak M is indicated in Figures 3.4 (lane HT) and 3.5.

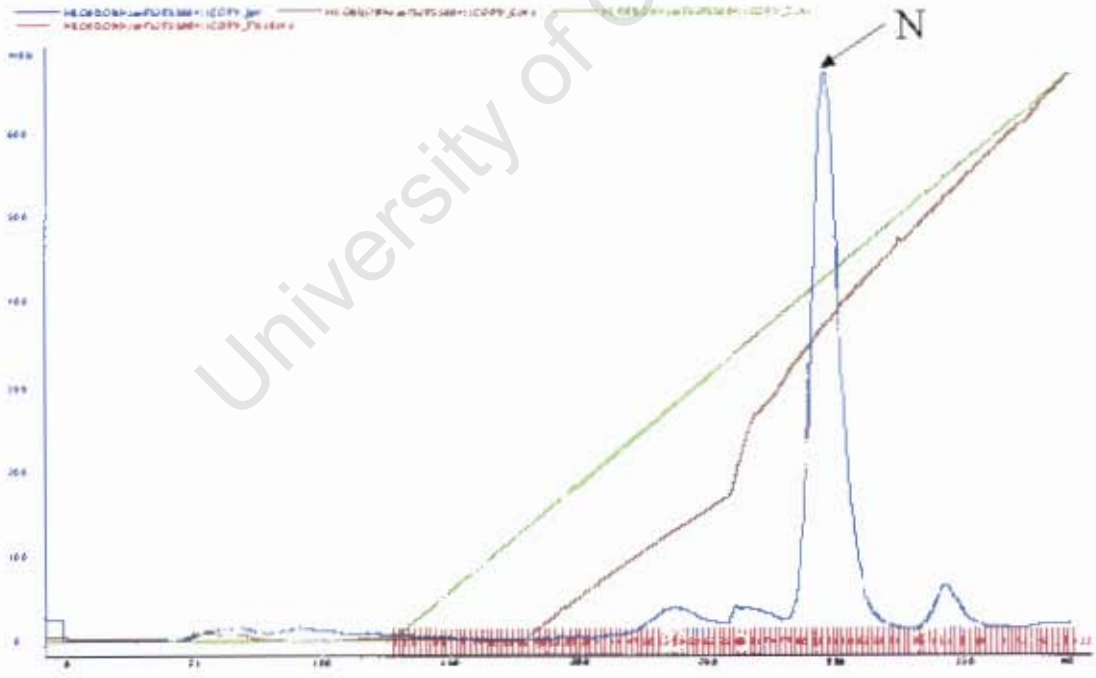


Fig. 3.8. Chromatograph of Q-Sepharose ion exchange chromatography. N: peak corresponding to electrophoretically pure NHase. The SDS PAGE of the pooled fractions from peak N is indicated in Figure 3.4 (lane Q).

University of Cape Town

Chapter 4: Crystallization, X-ray data collection and processing

4.1.0 Protein crystallization

The objective of protein crystallization is to obtain well-ordered crystals, but this task is difficult since protein molecules usually have irregular surfaces and are difficult to pack, leaving holes or channels which may be filled by disorientated solvent molecules.

In the hanging drop vapour diffusion method, the concentration of the precipitant solution and protein in the drop are slowly increased to a state of supersaturation by losing water to the precipitant in the well. Once a state of supersaturation state is reached, nuclei are formed which grow into larger crystals.

Concentration of protein, precipitants, buffer pH and temperature were varied since crystallization is a multi-parameter technique. Amorphous protein precipitates were obtained after several weeks in the PEG400 crystallisation trials. The PEG400 conditions did not produce any protein crystal. The sodium citrate crystallisation trials (Section 2.5.1) gave rise to about 15 crystals per single trial (overnight). The (1.1 M sodium citrate, 0.1 M Tris, pH 7.2, 15 mg/ml protein) condition had a single and large crystal (after 4 weeks) with morphology suitable for diffraction (Fig. 4.1).

University of Cape Town

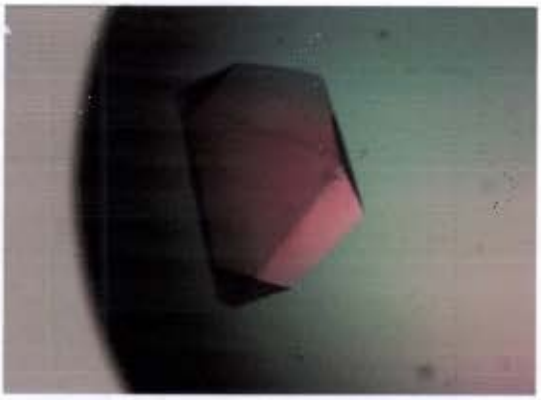


Fig. 4.1. Mutant NHase crystal (grown from 1.1 M sodium citrate, 0.1 M Tris, pH 7.2, 15 mg/ml protein) with approximate dimensions (0.3 x 0.2 x 0.2 mm) used for X-ray diffraction data collection (Section 2.6.1).

4.2.1 Data collection and processing

The 320 oscillation frames (Figure 4.2) collected on the Rigaku R-axis IV image plate detector (Section 2.6.1) were auto-indexed and integrated using d*TREK (Plugrath, 1999).

Auto-indexing assigns Miller indices to all the reflections in the diffraction pattern via Fourier-space whilst during integration the refined geometry values from auto-indexing are used to predict the reflections in all the 320 frames sequentially. The integrated data were scaled and merged with d*TREK (Plugrath, 1999). The merged data contains averaged intensities specific for each unique reflection in the asymmetric unit of the chosen point group. The output data from d*TREK (Plugrath, 1999) are indicated in Table 4.1. The space group and bravais lattice were $P4_12_12$ and primitive tetragonal respectively. The unit cell dimensions of the mutant (Table 4.1) were almost the same as the wild type (a= b= 106.61 Å, c= 83.23 Å; α= β= γ= 90°). The unit cell is described as the parallelepiped repeating unit existing in a protein crystal (Otwinowski *et al.*, 1997).

University of Cape Town

Mosaicity is the smallest angle that would enable reflections to be observed, when a crystal is rotated around an axis or combination of axis. It indicates orderliness of a crystal or the quality of the diffraction data collected. The lower the mosaicity, the higher the quality of the diffraction data and the crystal is better ordered and vice versa. Mosaicity ranges from 0.1 to several degrees (Otwinowski *et al.,* 1997). In this work, a high mosaicity of 1.89° was estimated which resulted in reduced diffraction data quality

The Rmerge, Intensity/Sigma (I/sigI) and % completeness value are statistical parameters which were used for evaluating the quality of the integrated data. Rmerge is an indicator of the quality of reflection data and is a measure of the agreement between multiple measurements of the same reflection during intensity data collection. The completeness of a diffraction data set is the ratio of number of unique reflections measured to total number of unique reflections and indicates the % of the possible intensities which have been actually measured at least once. <Intensity/SigmaI> is the average measured intensity divided by the associated sigma value for the intensity data after averaging equivalent reflections (Plugrath, 1999).

University of Cape Town

Data was processed to a resolution limit cutoff of 2.5 Å based on R-merge, completeness and I/sigma values in the last resolution shell although there are no generally agreed values for these parameters. A suggested resolution cutoff in d*TREK is the shell where the Rmerge is greater than 0.30 and the un-averaged I/sigI is less than 2. Overall Rmerge could be as low as 0.05 but values beyond 0.20 can be attributed to weak data or inaccurate processing. Completeness of more than 95 % has

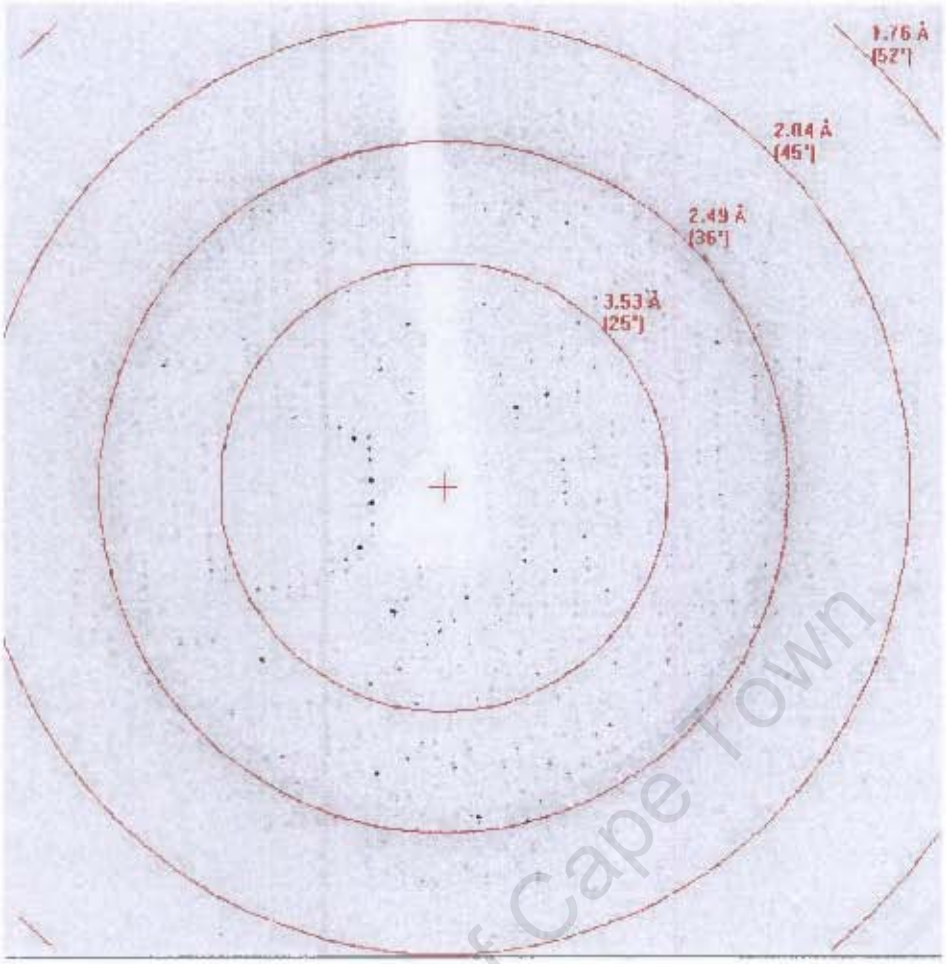
been suggested (Blow, 2002; Plugrath, 1999; Otwinowski *et al.,* 1997). In this work, completeness of 99.2 % and overall Rmerge of 0.29 were obtained (Table 4.1). The high Rmerge of 0.29 suggest that the quality of the data is poor. Rmerge ranges from about 0.219 – 0.44 for all the resolution shells. The completeness of 99.2 % was within the generally acceptable limits. The mutant dataset statistics (Table 4.1) are inferior to that obtained from wild type NHase with mosaicity less than 1.0° and Rmerge of 0.091 respectively (Tsekoa *et al.,* 2004).

Table 4.1. Summary of X-ray data collection statistics

Wavelength (Å)	1.5418
Space group	$P4_12_12$
Mosaicity	1.89°
Unit cell dimensions (Å)	a = b = 108.5, c = 81.15
Cell angles	$\alpha = \beta = \gamma = 90.00^{o}$
Resolution range (Å)	33.05 – 2.5
Number of unique reflections	16331
Completeness (%)	99.2 (99.1)
Rmerge	0.29 (0.44)
Output <I/sigI>	6.8 (4.1)

Note: Values in () are for the last resolution shell.

University of Cape Town



University of Cape Town

Fig. 4.2. A typical X-ray oscillation frame from mutant NHase crystal. The data was processed at a resolution limit cutoff of 2.5 Å.

4.2.2 Molecular replacement

Molecular replacement (Section 2.6.2.2) was carried out with Molrep (CCP4 suite, 1997) at 2.5 Å resolution using the crystal structure of wild type *Geobacillus pallidus* RAPc8 NHase as the starting model. The phase calculation via Molrep (CCP4 suite,

1997) gave a correlation coefficient of 63% and R-factor of 0.373. Prior to molecular replacement, the averaged and scaled intensities were truncated to structure factors. As expected, the mutant structure was made up of one αβ dimer per asymmetric unit.

4.3.1 Refinement and model building

The model structure from molecular replacement (Section 2.6.2.2) was refined automatically using Refmac5 (Murshudov *et al.,* 1997) and manually using O (Jones *et al.,* 1991). The model structure from molecular replacement underwent 5 separate cycles of rigid and least-squares restrained refinement resulting in R-factor of 0.324 and R_{free} of 0.385.

During refinement the atomic model parameters are adjusted to improve the agreement with diffraction data. The parameters are the restraints and the constraints that are applied to the model. Some of the typical restraints and constraints are bond lengths, bond angles, van der Waals distances, chirality and atomic B-factors. The R-factor is used to measure the agreement between the model and the diffraction data intensities, but this statistical indicator is prone to errors. These errors may be due to overfitting of the diffraction data with undesired parameters or by deliberately eliminating weak data from the refinement (Tickle *et al.,* 1998). Since in crystallography the electron density map is computed by combining amplitudes from the diffraction data and phases from the model, the final crystal structure could be model biased (CCP4 suite, 1997). The R_{free} is used to check refinement errors independently of the R-factor (Brünger, 1992) and is based on the standard modelling technique of jack-knifing or cross-validation residuals (McCullagh and Nelder, 1983). The R_{free} is conventionally the same as the R-factor but is calculated based on a test

University of Cape Town

set. The test set exclude about 5- 10% of reflections from structural refinement whilst only 90- 95% are used as the working set (CCP4 suite, 1997). The test set reflections are selected randomly and errors in them are not correlated to the working set (Tickle *et al.,* 1998).

In this work, 5% of all the reflections (h, k, l) were set aside to calculate the R_{free}, whilst 95% were used for the R-factor. The 5% test set used for computing the R_{free} of the mutant structure is independent of the test set used for the refinement of the wild type model. During the refinement and optimisation of the mutant structure, the working set reflections (95%) were from the diffraction data of the mutant and not the wild type. In order to prevent biased mutant model, the processing of the mutant structure was done independently of the wild type restraints and constraints.

The $2|F_{obs}| - |F_{calc}|$ map (Fig.4.3) of the wild type model showed that the side chains of F52 and F55 residues were clearly out of the electron density thereby confirming the success of the mutagenesis. The presence of a negative density at the position of the F52 residue in the $|F_{obs}| - |F_{calc}|$ map (Fig. 4.4), also supported the previously made assertion.

In-silico mutations and fitting of the F55LF52G model in the electron density were carried out in O (Jones *et al.,* 1991). The side chains of the F55LF52G model fitted exactly into the electron density of the $2|F_{obs}| - |F_{calc}|$ map (Fig.4.3). Although series of both manual and automatic (retrained and rigid body) refinements were carried out, there were no significant improvement in the R-factor and R_{free}. After R_{free} of 0.381 and R-factor of 0.312, water molecules were added to the structure. Water molecules were added manually using O (Jones *et al.,* 1991) and automatically using

University of Cape Town

ARP/wARP (Zwart *et al.*, 2002). Added water molecules were first inspected to be within hydrogen bonding length (2.4 –3.4 Å) of neighbouring residues or other water molecules. After subsequent restrained refinements, water molecules were tidied and rationalised using WATERTIDY and DISTANG (CCP4 suite, 1997). WATERTIDY indicated the specific residue that each water molecule was H-bonded to and when necessary moved the H_2O to the nearest symmetry related position. DISTANG computed the various bond lengths within the coordinates. R_{free} of 0.343 and R-factor of 0.236 were obtained after accurately adding 7 water molecules. During the initial stages of refinements, R_{free}- R-factor was within acceptable range but the difference increased after subsequent refinements, even though loose constraints were not deliberately applied to the process.

After exhaustive refinements and due to time constraints, preliminary assessment of the structure was carried out to gain initial insights into the engineering of homoaromatic specificity in the βF55LF52G mutant. The R_{free} is a global indicator and as such insensitive to small changes in a model. Building of small number of water molecules into a model would not affect it but the addition of about 50 may affect the R_{free} noticeably (Kleywegt and Brünger; 1996). In this work the R_{free} is high, further accurate addition of about 100 water molecules may bring both the R_{free} and R-factor down. The ease of refinement is associated with the quality and completeness of the diffraction data. The high mosaicity and generally poor dataset may have influenced the quality of the refinement data, since data quality has been shown to contribute to discrepancy between R-factor and R_{free} (Kleywegt and Brünger; 1996). The fact that every refinement process is unique and different, it is difficult to define guidelines for acceptable R_{free}. Models with R_{free} values above 40%

University of Cape Town

have been shown to be grossly incorrect (Kleywegt and Brünger; 1996). In this work the R_{free} was obtained as 34.3%.

Kleywegt and Brünger in 1996 analysed about 357 protein structures deposited in the protein databank. The average R_{free}- R-factor and range were determined as 7% and 3-17.6% respectively. The R_{free}- R-factor of the mutant NHase structure is 10.7% (Table 4.2). From a plot of R_{free}- R-factor as a function of resolution (Kleywegt and Brünger; 1996), 10.7% was within the range of the reported R_{free} values at the resolution of 2.5 Å.

At the current R-factor and R_{free}, the model structure is adequate to be analysed. It is evident that further refinement must be carried out in the future for complete structural analysis.

The stereochemical folding of the main chain was checked by assessing the dihedral angles from the Ramachandran plot using the Rampage server (Lovell *et al.,* 2002). The Ramachandran plot (Lovell *et al.,* 2002) of the final F55LF52G model (Fig. 4.5) showed 16 residues in the outlier regions. The Protein Interaction Server (Jones and Thornton, 1996) indicated that none of the residues at the active site or those reported to be involved in substrate recognition are in the outlier region. Model quality assessment from PROCHECK (Laskowski *et al.,* 1993) and SFCHECK (Vaguine *et al.,* 1999) indicated that the structure is adequately enough to be analysed (Table 4.2).

University of Cape Town

Table 4.2. Summary of Refinement statistics

Final R-factor (%)	23.6
R_{free} (%)	34.3
R_{free}- R-factor (%)	10.7
R.m.s. deviation from ideality (target values)	
Bond length (Å)	0.045 (0.022)
Bond angles (°)	3.830 (1.969)
<B> (for atomic model)	23.0
Sigma (B)	9.69
B_overall (by Patterson)	34.1
No. of atoms (protein)	3486 atoms
No. of water molecules	7
Number of chains	2
No. of cobalt atoms	1
No. of residues in chain A	211
No. of residues in chain B	277
Volume not occupied by model	38.2 %
Corresponding solvent (%)	49.39
Quality of Ramachandran plot	
Residues in favoured region (%)	85.5
Residues in allowed regions (%)	10.7
Residues in disallowed regions (%)	3.8

University of Cape Town

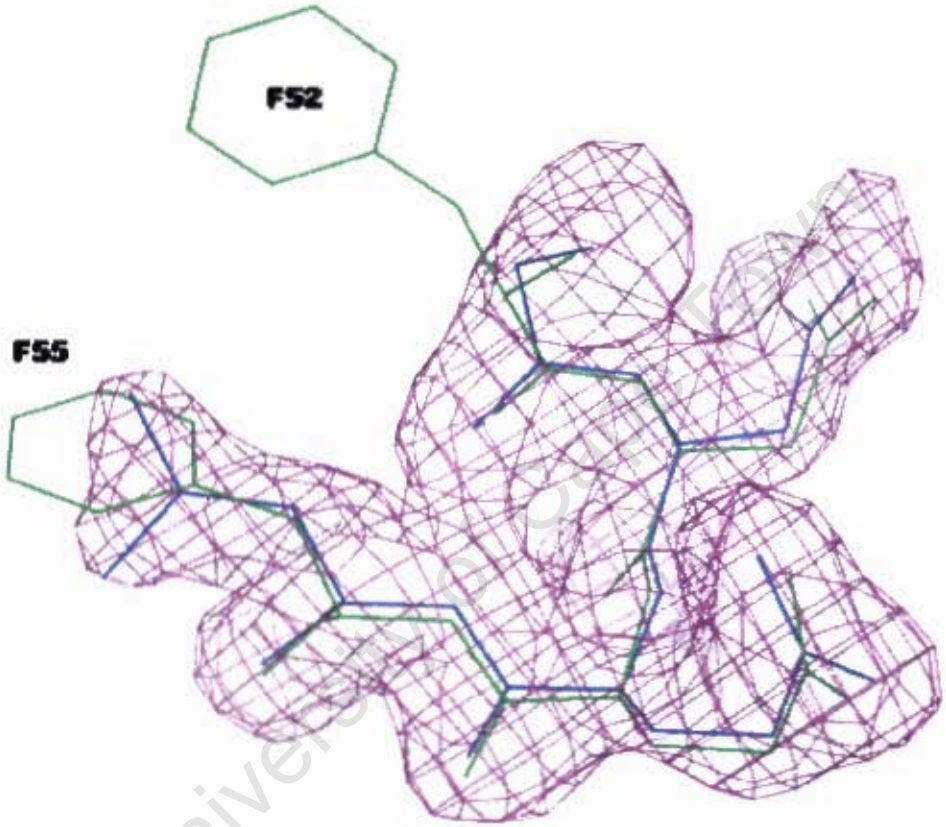


Fig. 4.3. 2| F_{obs}| - |F_{calc} | map showing the electron density associated with the wild type NHase model after molecular replacement generated at 1.0 contour level. Green sticks: Wild type NHase structure with side chains of F52 and F55 clearly out of the electron density; Blue sticks: Mutant NHase with F52 and F55 side chains fitting the density.

University of Pune Library

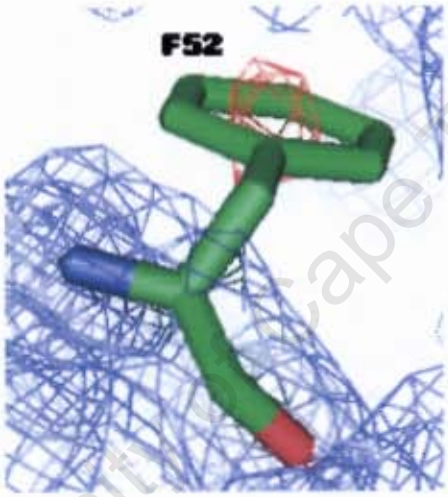


Fig. 4.4. 2| F_{obs}| - |F_{calc} | (contour level 1) and | F_{obs}| - |F_{calc} | (contour level –3.0) maps showing negative electron density at the F52 position in the wild type NHase model.

University of Cape Town

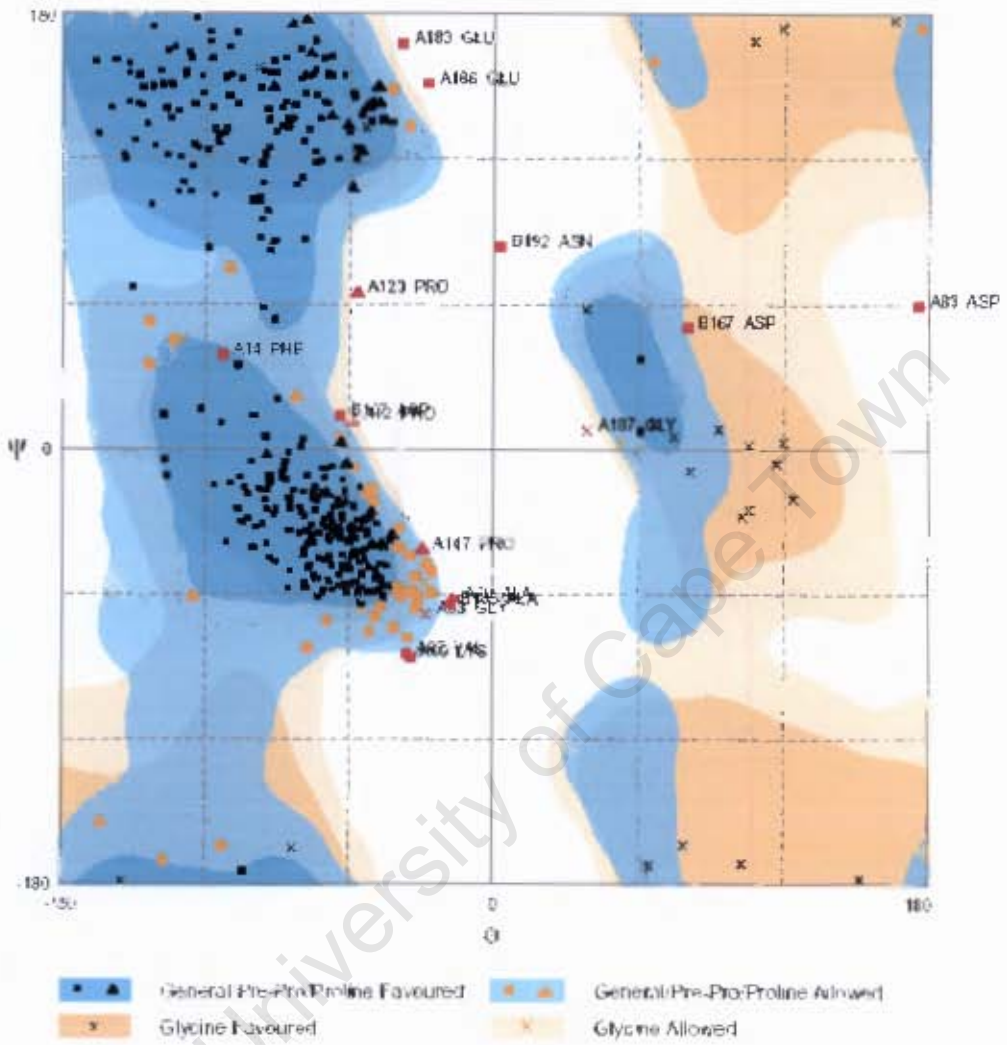


Fig. 4.5. Ramachandran plot (Lovell *et al.*, 2002) of mutant NHase model. Number of residues in favoured region: 359 (85.5%); Number of residues in allowed region: 45 (10.7%); Number of residues in outlier region: 16 (3.8%).

University of Cape Town

Chapter 5: Mutant NHase structure and analysis

5.0 Mutant NHase structure

Alignment of the Cα atoms of the mutant and wild type NHase (Fig. 5.1a) structures yielded a RMSD of 0.364 Å. There was no major deviation between the mutant and the wild type. Alignment of the cobalt binding residues of the wild type and mutant structures also showed no deviations (Fig. 5.2). There were no unusual side chain modifications when the aromatic residues around the active site pockets of both the wild type and mutant structures were aligned (Fig. 5.3). The first 9 residues at the N-terminal of the α subunit were absent as in the case of the wild type.

The single non-corrinoid Co^{III} co-factor is located at the internal crevice at the interface between the α and β monomers. The ligands to the cobalt co-factor are provided by the α-subunit. The cobalt atom is octahedrally coordinated and is liganded to 5 vertices consisting of three sulfur atoms of αC116, αC119, αC121 and two main chain amide nitrogen atoms of αS120 and αC121. The sixth vertex may be occupied by a solvent exchangeable group such as water or hydroxide (Fig. 5.4). αC119 and αC121 are post-translationally oxidised to cysteine sulphinic acid and cysteine sulphenic acid respectively. These modifications and the rest of the metal binding residues form the claw setting motif. The mutant NHase consist of non-homologous α and β monomers (Fig. 5.5) which are bound together to form αβ dimer. The mutant crystal structure is a heterotetramer made up of αβ dimer bound to another αβ dimer in two-fold axis symmetry (Figures 5.1b and 5.6).

University of Cape Town

The N-terminal loops of both the α and β monomers wrap around each other in opposite directions and this may stabilize the αβ dimer. UHBD predicted very low electrostatic energy contribution to dimerisation by both the wild type and mutant NHase (Section 2.7). This could be an indication of dynamic dimerisation rather a static one.

The region β95 toβ138 in Co-type NHases contains an insertion of about 12-20 residues. This region in the mutant NHase has an α-helix that interacts with another from the α-subunit. This inserted α-helix is absent in Fe-type NHases and may contribute to stability as proposed in Co-type *Ps. thermophilia* (Miyanaga *et al.*, 2001).

University of Cape Town

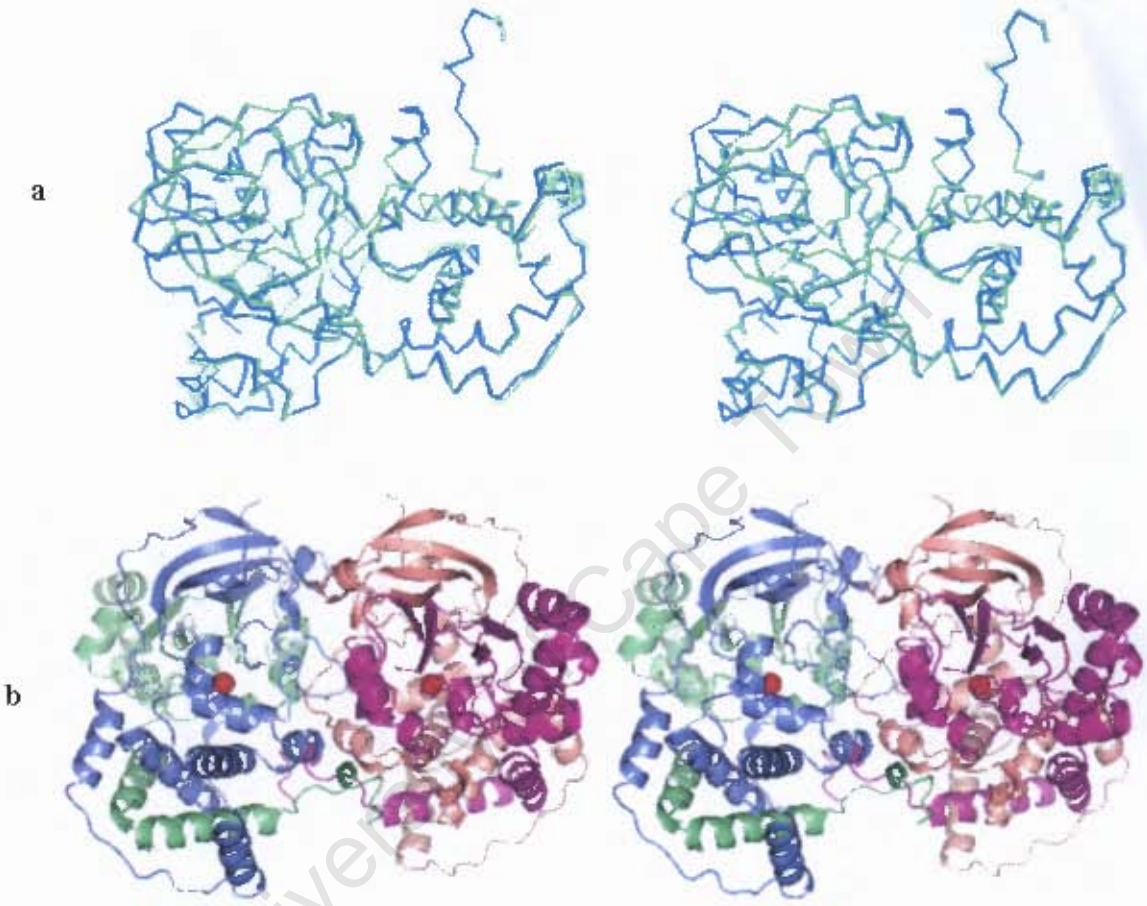


Fig. 5.1. (a) Stereo representation of Cα Alignment of mutant and wild type NHase heterodimers. Blue ribbon: mutant NHase; green ribbon: wild type NHase (b) Stereo representation of mutant NHase heterotetramer. Magenta and green cartoons: alpha subunit; salmon and blue cartoons: beta subunit; Red spheres: cobalt metal.

University of Cape Town

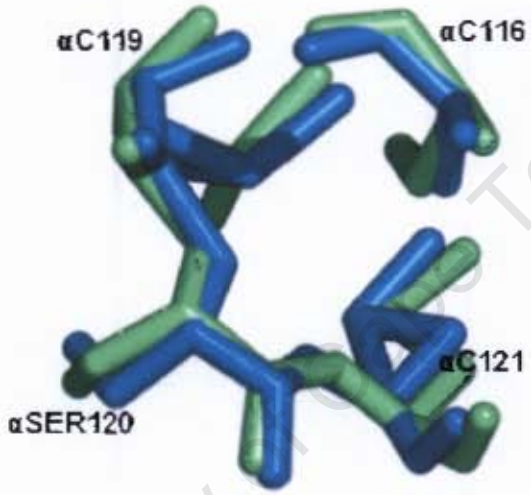


Fig. 5.2. Alignment of the cobalt binding residues. Blue ribbons: mutant NHase; green ribbons: wild type NHase.

University of Cape Town

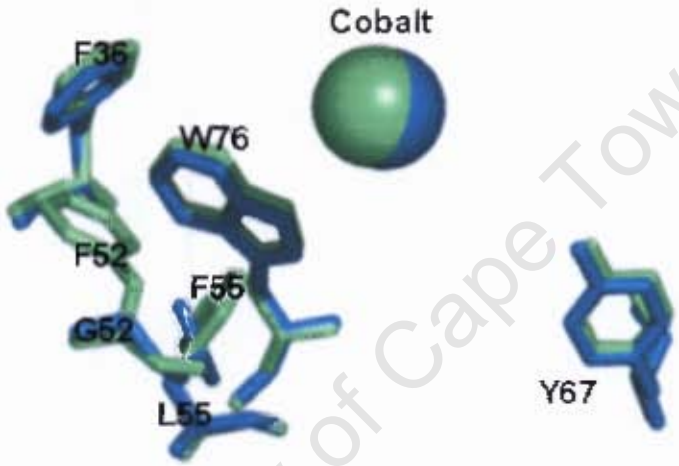


Fig. 5.3. Alignment of aromatic residues around the active site cavities of mutant and wild type NHase heterodimers. Blue ribbon: mutant NHase; green ribbon: wild type. Green sphere: mutant cobalt; blue sphere: wild type cobalt co-factor. βG52 of mutant aligned with βF52 of wild type NHase. βL55 of mutant aligned with the βF55 of wild type NHase.

University of Cape Town

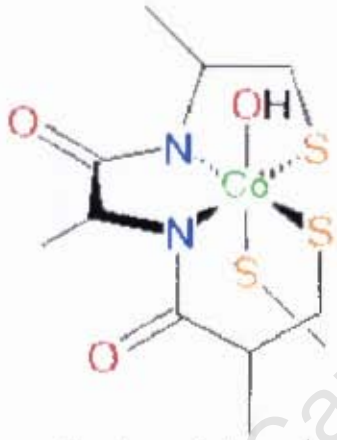


Fig. 5.4. The octahedral coordination of the active site of the mutant NHase. The Cobalt co-factor is liganded to three sulpher (S) atoms of cysteine and two main chain amide nitrogen (N) atoms of serine and cysteine. The sixth position is occupied by either hydroxide (OH) or water.

University of Cape Town

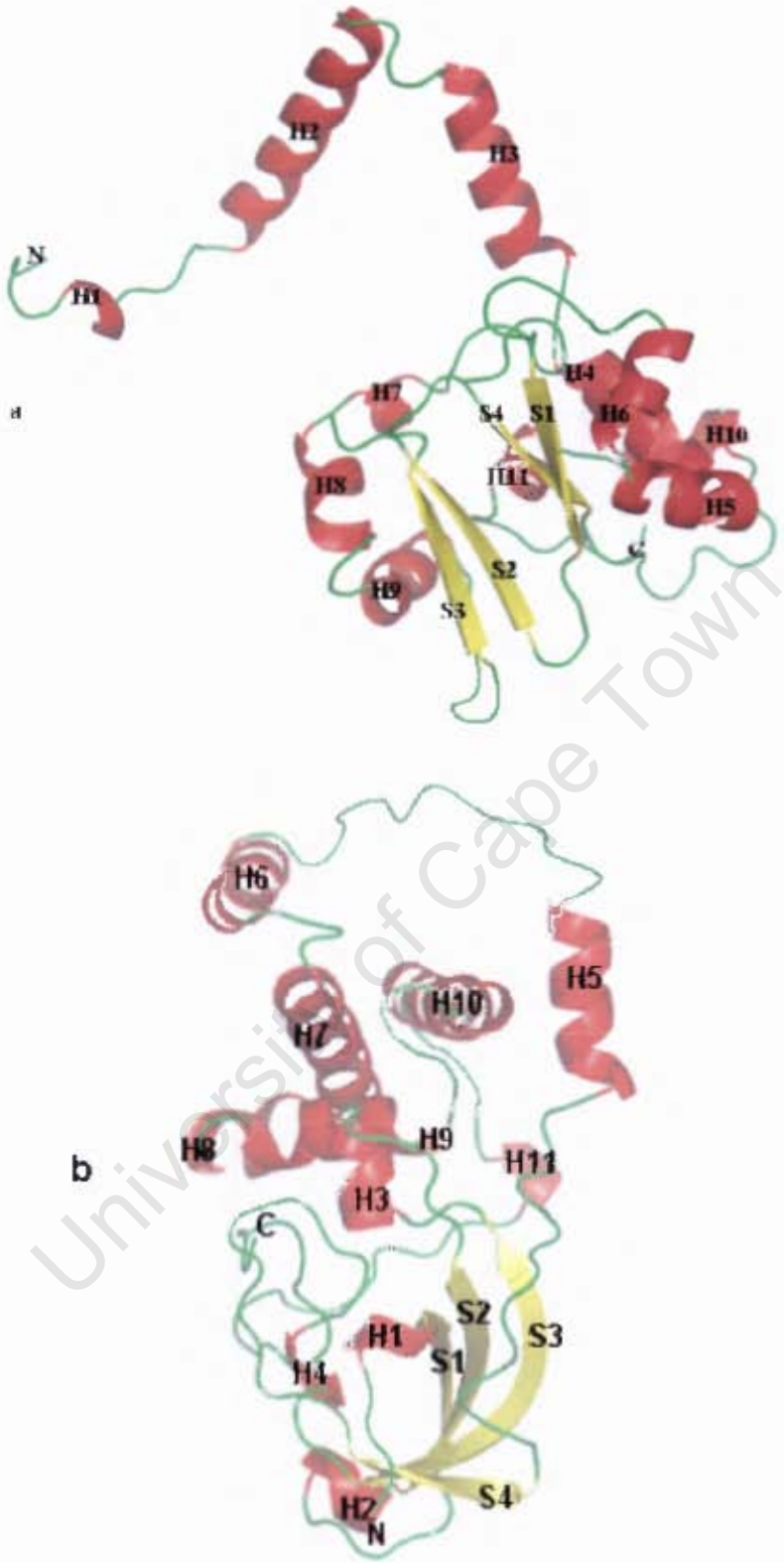


Fig. 5.5. Cartoon representation of mutant NHase monomers. (a) Alpha subunit, (b) beta subunit. N and C represent N- and C- terminals respectively. Alpha helices and beta sheets are labeled H and S respectively.

University of Cape Town

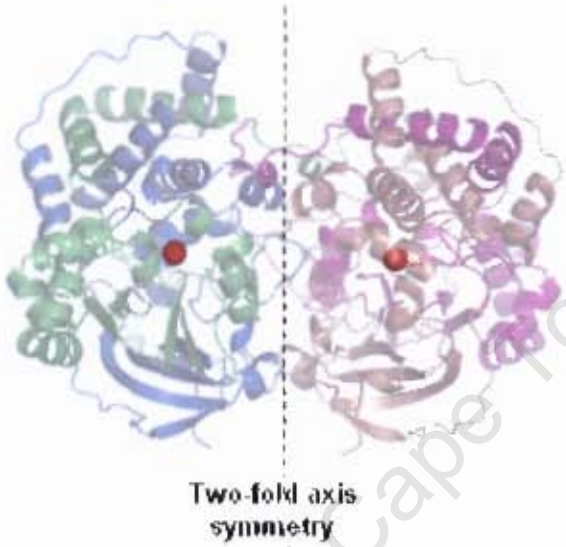


Fig. 5.6. Heterotetramer of mutant NHase showing two-fold axis crystallographic symmetry. Magenta and green cartoons: alpha subunit, salmon and blue cartoons: beta subunit; red spheres: cobalt metal.

University of Cape Town

5.1 Normal mode analysis

Normal mode analysis (NMA) of all structures was done using the Web-server for Normal Mode Analysis of proteins - v2.0 (Hollup *et al.*, 2005). Normal mode analysis demonstrates the intrinsic flexibility within protein structures. This concept is based on vibrational normal modes with the least frequencies. These low frequency modes give insight into functionally relevant domain movement within a protein molecule (Hollup *et al.*, 2005). The program computes the deformation energies of about 200 lowest frequency modes. Deformation energy measures collective movement within a mode and is calculated from all the atoms in the protein. Low deformation energy indicates rigidity in domain movement. High values indicate higher degree of intrinsic flexibility (Hollup *et al.*, 2005).

The deformation energies of the NHase structures with the least frequency modes computed from normal mode analysis (Hollup *et al.*, 2005) are indicated in Table 5.1. The deformation energy of the mutant heterodimer was higher than that of the wild type, indicating a high degree of flexibility within the mutant heterodimer. The deformation energy of each heterotetramer was about 700 units lower than their respective heterodimer. This is an indication that the heterodimers are more flexible than the heterotetramers.

The deformation energy difference between the mutant and wild type tetramer is about 10 units. Deformation energies computed using the Web-server for Normal Mode Analysis of proteins - v2.0, have no units and standards against which they could be compared. The difference of 10 units may indicate almost equal magnitude in flexibility. This difference could also be attributed to the possible existence of some

University of Cape Town

few extra amplitudinal movements in the mutant, which may be probably absent in the wild type. As indicated by the movements of the modes (Fig. 5.7), there is a high degree of intrinsic flexibility within both NHases. Intrinsic flexibility has been reported to induce functionally relevant conformational changes in proteins (Hollup *et al.*, 2005).

The existence of these intrinsic movements in NHase may induce functionally relevant aromatic substrate recognition conformation in the mutant NHase. These conformational changes and the possible reduction in the aromatic interactions, may probably favour the binding of the homoaromatic substrate to the active site. After possible entry of homoaromatic substrate into the catalytic channel of the mutant, the few extra amplitudinal movements probably present in the mutant may move the bulky homoaromatic substrate towards the active site.

University of Cape Town

NHase structure	Deformation energy (mode 7)
Mutant heterodimer	1299.1
Wild type heterodimer	1258.1
Mutant heterotetramer	565.9
Wild type heterotetramer	575.7

Table 5.1. The lowest deformation energies for the least frequency modes of NHase structures.

University of Cape Town

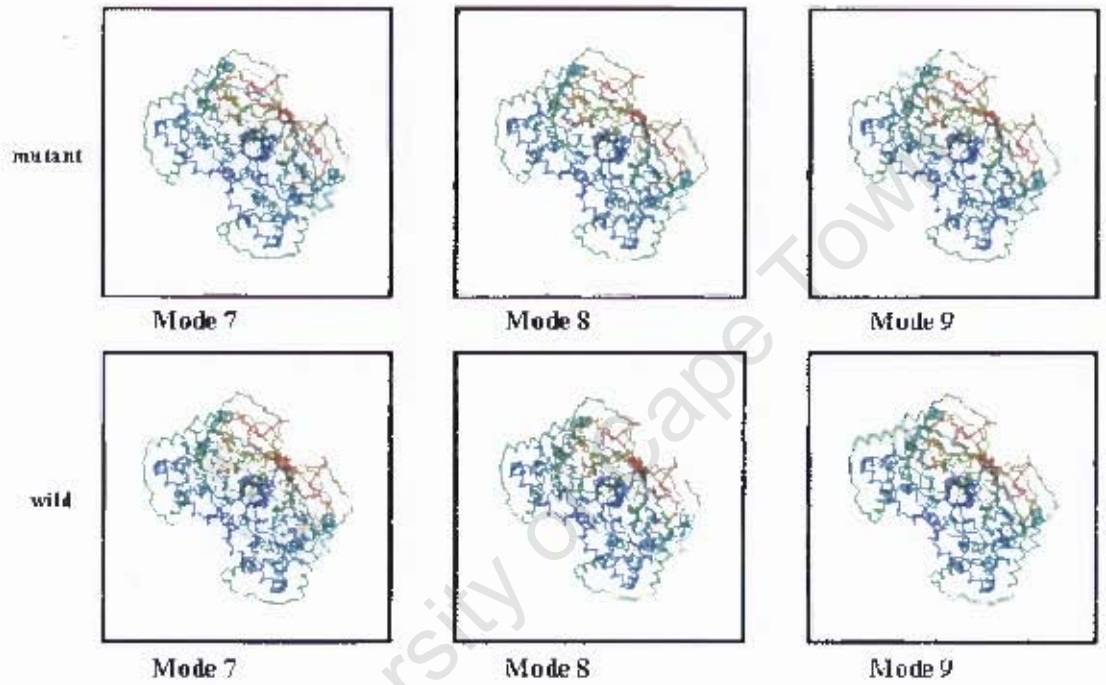


Fig. 5.7. Various modes of the NHase structures (Hollup *et al.*, 2005). Mutant: mutant NHase modes; wild: wild type NHase modes. Red represent most rigid regions (low deformation energies); green represent regions of deformation threshold, and blue represent most flexible regions (above deformation threshold).

University of Cape Town

5.2 Cavity calculations

Active and binding sites in proteins are mostly located in cavities and pockets. Knowledge of the dimensions and location of these pockets may give insights into substrate recognition in proteins (Liang *et al.,* 1998). Pockets in NHase structures were identified and analysed using CASTp Server (Liang *et al.,* 1998) and PyMOL (Delano). CASTp Server employs the alpha complex and weighted Delaunay triangulation to compute shape measurements (Liang *et al.,* 1998).

As indicated in Table 5.2 and Figure 5.8, each dimer of mutant NHase has 1 main channel running across it through their respective internal crevice. Each channel has 2 opposing openings. Each pocket is also lined separately with all the cobalt binding residues (αC119α, αS120 and αC121) except αC118. Residues βG52, βL55, βF36, βY67 and βW76 which are around the active site cavity, also line these 2 individual pockets separately. These 2 pockets running across the tetramer have equal dimensions.

The wild type tetramer NHase has 3 main pockets (Data not shown). 2 of these pockets are fused together to form a single cavity. This cavity is lined with the cobalt binding residues except αCys118. The cavity is also lined with the aromatic residues (βG52, βL55, βF36, βY67 and βW76) around the active site.

Electron density, cavity calculations of the mutant NHase demonstrate that the βF52 may be acting as a lid by regulating the entry of bulky aromatic substrate. There are 2 independent pockets running through the F55LF52G mutant NHase with no internal blockage. From the alignment of the aromatic residues around the active site (Fig.

University of Cape Town

5.3), the substitution of the side chain of βF52 with H in βG52 of the F55LF52G mutant may have contributed extra volume previously occupied by the phenylalanine side chain. The increase in pocket dimensions and the existence of one single channel per dimer in F55LF52G may indicate the possible removal of a F52 lid. Since the wild type did not have a single channel passing through the entire length of any of the dimers, the presence of F52 might have contributed to the absence of a single main channel across the dimers. The removal of this lid and 4 possible entrances for the homoaromatic substrates in the mutant, may contribute to homoaromatic substrate recognition in the F55LF52G mutant.

University of Cape Town

Table 5.2. Dimensions of pockets identified in heterotetramer NHase structures.

NHase	Surface Area (Å^2)	Volume (Å^3)
Mutant	2374	3042
	2374	3042
Wild type	1755	2166

University of Cape Town

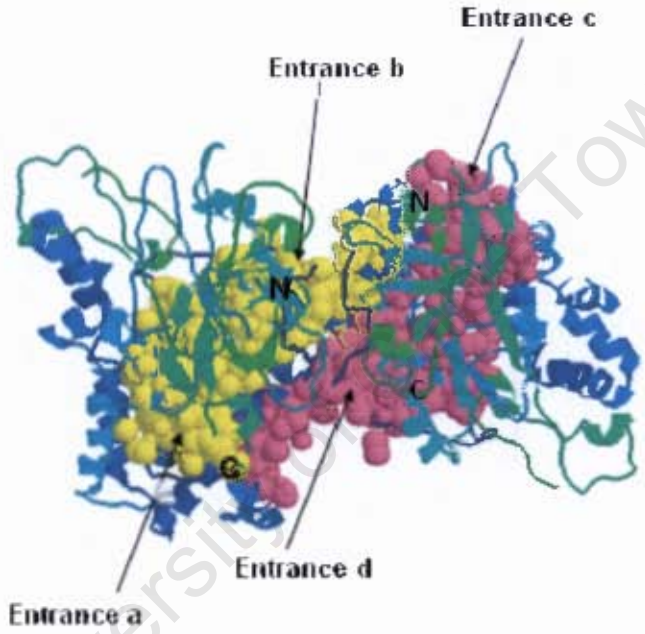


Fig. 5.8. Cavities and pockets within the mutant tetramer. Yellow spheres: single cavity through one dimer with entrances a and b; Hot pink spheres: single cavity through another dimer with entrances c and d; N and C: represent the N and C-terminals respectively.

University of Cape Town

5.3 Possible substrate entry mechanisms in the F55LF52G mutant

The intrinsic flexibility in the NHases induces conformational changes. These changes create movements that may be described as shear motions in the structures of NHases analysed so far. Shear motions are parallel to the interface between the closely packed segments of the protein. This motion cause changes in the main and side chain torsion angles of the segments in the protein interface. (Perutz, 1989; Gerstein *et al.,* 1994; Janin and Wodak, 1983). These motions shift and rotate the various segments relative to each other, and may give rise to flip-flop substrate entry mechanism. This flip-flop entry mechanism in NHases was previously identified as one of the possible substrate entry mechanisms, but was not attributed to the possible existence of shear motions (Bishop and Sewell, 2006)

From Figures 5.8, the flip-flop substrate entry mechanism may be demonstrated in the mutant NHase as follows: shear motions cause segments around the C- terminals of the beta subunits to come closer to each other, with the simultaneous moving apart of segments around the N- terminals of the alpha subunits; and vice versa. A substrate may enter through entrance b, when the N- terminal segments of the alpha subunits move apart; or enter through channel d, when the C- terminal segments of the beta subunits move apart. Though entry through channels a and c may be due to shear motion, they may not be described as flip- flop mechanism. Flip-flop entry mechanisms in this instance may not be related to catalysis, it is merely describing one of the possible entry mechanisms of substrate into the active site channel. The possible existence of flip-flop entry mechanism and shear motions in the F55LF52G may contribute to homoaromatic substrate recognition in the mutant.

University of Cape Town

5.4 Kinetic studies of *Geobacillus pallidus* RAPc8 Nitrile Hydratase

Geobacillus pallidus RAPc8 NHase is known to possess broad substrate specificity by converting branched, linear, cyclic, dinitriles and heteroaromatic (Acrylonitrile and 3-cyanopyridine) substrates to their corresponding amides. Homoaromatic substrates such as benzyl cyanide and benzonitrile are not catalysed by the enzyme (Pereira *et al.,* 1998). The inability of *Rhodococcus* sp. N-771 NHase to catalyse homoaromatic substrates was suggested to be as a result of smaller substrate cavity volume (Nagashima *et al.,* 1998). Since both homoaromatic and heteroaromatic substrates are equally bulky in size, absence of activity may not be attributed to smaller cavity volume. The absence of activity may be attributed to aromatic- aromatic interactions between the aromatic substrate and the aromatic residues around the active site cavity. These aromatic interactions formed the basis of this research.

Extensive Kinetic research on the double mutant engineered in this work has been undertaken at the University of Bath by Parik Kowlessur. The K_m for acrylonitrile was obtained as 2.58 mM. Acrylonitrile K_m of about 11 mM was obtained from previous studies (Cameron, 2003; Tsekoa, 2005) on the wild type NHase. This indicated a 3-fold decrease in K_m by the mutant with respect to the wild type. From the K_m values, the βF55LF52G NHase has a higher affinity for acrylonitrile. The βF55LF52G NHase was active on benzonitrile with K_m of 136 μm (unpublished). This is the first time that benzonitrile activity has been shown in *Geobacillus pallidus* RAPc8 NHase. It must be noted that further kinetic work on the mutant is currently on going.

University of Cape Town

Since alignment of the F55LF52G and the Wild type NHases showed no major deviation, homoaromatic activity in the mutant could be investigated through normal mode analysis and substrate cavity volumes. CASTp server was used to compute pocket dimensions of the wild type and mutant, hence direct comparison could be made. The surface area and volume of the substrate pocket of the βF55LF52G mutant increased by 26% and 28% respectively when compared to the wild type. Benzonitrile activity in βF55LF52G mutant could be due to the increase in substrate cavity volume and surface area. Upon entry of the benzonitrile into the substrate cavity, intrinsic flexibility within the mutant may induce homoaromatic functionally relevant conformation within the structure. This may expose the active site for the benzonitrile to bind.

The substitution of aromatic F55 and F52 with leucine and glycine respectively, may have decreased the aromatic interactions between the aromatic benzonitrile and the aromatic residues around the active site cavity.

University of Cape Town

5.5 Proposals for future work

5.5.1 Reprocessing of Refinement data

The intensity dataset must be integrated, averaged and scaled again to improve the data quality statistics. Since only 7 water molecules have been built into the mutant NHase structure, adequate water molecules must be added to bring the R-factor and R_{free} within acceptable limits. The completely refined structure can then be re-analysed to confirm the earlier conclusions and inferences made. Additional crystallization trials should also be set up in the future with the aim of obtaining better crystals for data collection.

5.5.2 Engineering of future of mutants

Future mutants may be engineered to have homoaromatic substrate specificity by increasing the catalytic cavity volumes and surface areas. Reasonable mutation of aromatic residues around the catalytic volume interface could be carried out to decrease the aromatic- aromatic interactions. Kinetic data must also be obtained in future work to support every speculation.

5.5.3 Brownian dynamics simulations of NHase substrate complexes

Reasonable models of the NHase mutant and aromatic substrate complexes could be obtained. This can be done by docking (Pearson and Gross, 1998; Ouporov *et al.*, 1999) aromatic substrates computationally into the active sites of the crystal structures of NHases. Homology modelling (Sali and Blundell, 1993) based on the wild type crystal structure would be used to generate other mutants. The modelled complexes

University of Cape Town

could also be improved by molecular mechanics structural refinement (Pearson and Gross, 1998; Ouporov *et al.,* 1999). This would provide the opportunity for other of mutants to be modelled. This concept can be extended to all structures of NHases in the protein data bank (Bernstein *et al.,* 1977).

Brownian dynamics simulations of the modelled NHase substrate complexes would be carried out (Pearson and Gross, 1998; Ouporov *et al.,* 1999). The NHase in the modelled complex with the highest docking probability and the most favourable electrostatic interactions (energy) may be the NHase that would have activity on aromatic substrates. This could aid in identifying the critical residues involved in the interaction between aromatic substrates and NHases. The identified NHase would then be engineered and kinetic studies carried out to confirm the computational work. The crystal structure can also be determined to further enhance the understanding of substrate preference through structural analysis.

University of Cape Town

ABBREVIATIONS

Å	Angstrom
Co-type NHase	Cobalt containing nitrile hydratase
Da	Dalton
EC	Enzyme commission
ENDOR	Electron-nuclear double resonance
Fe- type NHase	Iron containing nitrile hydratase
g	gram
h	hour
IPTG	Isopropyl β-D-1-thiogalactopyranoside
k	kilo
kDa	kilodalton
K_m	Michaelis-Menten constant
L	Litre
M	milli
Molar	M
min	minute
NHase	Nitrile hydratase
NO	Nitric oxide
NMA	Normal mode analysis
NTase	Nitrilase
ORFs	Open reading frames
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction

University of Cape Town

PDB	Protein Data bank
R.M.S.D.	Root mean square deviation
R-factor	Crystallographic R-factor = $\sum\|\|F_{obs}-\|F_{calc}\|\|/ \sum\|F_{obs}\|$
$\mathbf{R_{free}}$	Crystallographic R-factor for the test set (5% of reflections)
$\mathbf{R_{merge}}$	Merging R-factor = $\sum (\|(I-<I>)\|)/ \sum (I)$
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
s	Seconds
sp.	Specie
$\mathbf{2\|F_{obs}\| - \|F_{calc}\|}$	Electron density map associated with model atoms
$\mathbf{\|F_{obs}\| - \|F\sum_{calc}\|}$	Difference electron density map
µ	micro

University of Cape Town

References

Battistel E., Bernardi A., Maestro P. (1997). Enzymatic decontamination of aqueous polymer emulsions containing acrylonitrile. Biotechnol. Lett. **19**, 319- 25.

Bernstein F. C., Koetzle T. F., Williams G. J., Meyer E. E., Jr., Brice M. D., Rodgers J. R., Kennard O., Shimanouchi T. and Tasumi M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. **112,** 535-542.

Bishop A. O. T. and Sewell T. (2006). A new approach to possible binding mechanisms for nitrile hydratase. Biochem. Biophys. Res. Commun. **343**, 319–325.

Blow D. (2002). Outline of crystallography for biologists. Oxford University Press. UK. Pp. 111-119.

Brünger A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature. **355**, 472-75.

Cameron R. (2003). Nitrile degrading enzymes from extreme environments. Ph.D. Thesis University of London.

Cameron R. A., Muhammed S., Cowan D. A. (2005). Molecular analysis of the nitrile catabolism operon of the thermophile *Bacillus pallidus* RAPc8. Biochim. Biophys. Acta. **1725**, 35 – 46.

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

Collaborative Computational Project, Number 4. (1994). The CCP4 Suite: Programs for protein Crystallography. Acta Cryst. **D50**, 760–763.

Cowan D. A., Cameron R. A., Tsekoa T. L. (2003). Comparative biology of mesophilic and thermophilic nitrile hydratases. Adv. Appl. Microbiol. **52**, 123–158.

Cowan D., Cramp R., Pereira R., Graham D., Almatawah Q. (1998). Biochemistry and biotechnology of mesophilic and thermophilic nitrile metabolizing enzymes. Extremophiles. **2,** 207-216.

Cramp R. *et al.* (1997). Novel nitrile degrading enzymes in thermophilic bacteria. Ph.D. Thesis. University of London.

Cramp R. A., and Cowan D. A. (1999). Molecular characterisation of a novel thermophilic nitrile hydratase. Biochim. Biophys. Acta. **1431,** 249-260.

DeLano W. L. "The PyMOL Molecular Graphics System." DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org.

University of Cape Town

Duran R, Nishiyama M., Horinouchi S., Beppu T. (1993) Characterization of nitrile hydratase genes cloned by DNA screening from *Rhodococcus erythropolis*. Biosci. Biotechnol. Biochem. **57**, 1323–1328.

Fallon R. D., Stieglitz B., Turner I. (1997). A *Pseudomonas putida* capable of stereoselective hydrolysis of nitriles. Appl. Microbiol. Biotechnol. **47**, 156-161

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. Pp. 571-607.

Gerstein M., Lesk A. M., Chothia C. (1994). Structural Mechanisms for Domain Movements in Proteins. Biochemistry. **33**, 6739-6749.

Gilligan T., Yamada H., Nagasawa T. (1993). Production of S-(+)-2-phenylpropionic acid from (R,S)-2-phenylpropionitrile by the combination of nitrile hydratase and stereoselective amidase in *Rhodococcus* equi TG328. Appl. Microbiol. and Biotechnol. **39**, 720-727.

Hall T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. [http://www.mbio.ncsu.edu/BioEdit/bioedit.html]. Nucleic Acids Symp. Ser. **41**, 95-98.

Hann E. C., Eisenberg A., Fager S. K., Perkins N. E., Gallagher F. G., Cooper S. M., Gavagan J. E., Stieglitz B., Hennessey S. M., DiCosimo R. (1999). 5-Cyanovaleramide production using immobilized *Pseudomonas chlororaphis* B23. Bioorg. Med. Chem. **7**, 2239–2245.

Heald S. C., Brandao P. F., Hardicre R., Bull A. T. (2001). Physiology, biochemistry and taxonomy of deep-sea nitrile metabolising *Rhodococcus* strains. Ant. V. Lee. **80**, 169-183

Heveling J., Armbruster E., Utiger L., Rohner M., Dettwiler H., Chuck R. (1998). Process for preparing nicotinamide. US. Patent 5,719,045.

Hollup S. M., Sælensminde G, Reuter N. (2005). WEBnm@: a web application for normal mode analysis of proteins. BMC Bioinf. **6**, 52.

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.

Inoue H., Nojima H., Okayama H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. Gene. **96**, 23-28.

Janin J. and Wodak S. J. (1983). Structural domains in proteins and their role in the dynamics of protein function. Prog. Biophys. Molec. Biol. **42**, 21-78.

University of Cape Town

Jones S. and Thornton J. M. (1996). Principles of Protein-Protein Interactions Derived From Structural Studies. PNAS. **93**, 13-20.

Jones T. A., Zou J-Y. , Cowan S. W. Kjeldgaard M. (1991). Improved methods for the building of protein models in electron density maps and the location of errors in these models. Acta Cryst. A. **47**, 110-119.

Kato Y., Tsuda T., Asano Y. (1999). Nitrile hydratase involved in aldoxime metabolism from *Rhodococcus* sp. strain YH3-3 purification and characterization. Eur. J. Biochem. **263**, 662-70.

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

Kim C. Y., Chandra P. P., Jain A., Christianson D. W. (2001). Fluoroaromatic-fluoroaromatic interactions between inhibitors bound in the crystal lattice of human carbonic anhydrase II. J. Am. Chem. Soc. **123**, 9620–9627.

Kim S. and Oriel P. (2000). Cloning and expression of the nitrile hydratase and amidase genes from *Bacillus* sp. BR449 into *Escherichia coli*. Enzyme Microb. Technol. **91**, 492-501.

Kim S. H., Padmakumar R., Oriel P. (2001). Cobalt activation of *Bacillus* BR449 thermostable nitrile hydratase expressed in *Escherichia coli*. Appl. Biochem. Biotechnol. Spring. **91**, 597- 603.

Kleywegt G. J. and Brünger A. T. (1996). Checking your imagination: applications of the free R value. Structure. **4**, 897- 904.

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

Kobayashi M. and Shimizu S. (1999). Cobalt proteins. Eur. J. Biochem. **261**, 1-9.

Kobayashi M., Komeda H., Shimuzu S., Yamada H., Beppu T. (1997). Characterization and distribution of IS1164 that exists in the high molecular mass NHase gene cluster of the industrial microbe *Rhodococcus rhodochrous* J1. Proc. Jpn. Acad. Series B-Phys. Biol. Sci. **73**, 104-108.

Kobayashi M., Nishiyama M., Nagasawa T., Horinouchi S., Beppu T., Yamada H. (1991). Cloning, nucleotide sequence and expression in *Escherichia coli* of two cobalt-containing nitrile hydratase genes from *Rhodococcus rhodochrous* J1, Biochim. Biophys. Acta. **1129**, 23– 33.

Kobayashi M., Nagasawa T., Yamada H. (1992). Enzymatic synthesis of acrylamide: A success story not yet over. Trends Biotechnol. **10**, 402-408.

Komeda H., Kobayashi M., Shimizu S. (1996b). Characterization of the gene cluster of high-molecular-mass nitrile hydratase (H-NHase) induced by its reaction-product In *Rhodococcus-rhodochrous* J1. PNAS. **93**, 4267–4272.

University of Cape Town

Komeda H., Kobayashi M., Shimizu S. (1996a). A novel gene cluster inducing the *Rhodococcus rhodochrous* J1 *nhlBA* genes encoding a low molecular mass nitrile hydratase (L-NHase) induced by its reaction product. J. Biol. Chem. **271**, 15796-15802.

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

Laskowski R.A., MacArthur M. W., Moss D. S., Thornton J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. App. Cryst. **26,** 283

Liang J., Edelsbrunner H., Woodward C. (1998). Anatomy of Protein Pockets and Cavities: Measurement of Binding Site Geometry and Implications for Ligand Design. Prot. Sci. **7**, 1884-1897.

Lovell S.C., Davis I.W., Arendall III W.B., Bakker P.I.W. de, Word J.M., Prisant, J.S. Richardson M.G., Richardson D.C. (2002). Structure validation by Ca geometry: f/y and Cb deviation. Prot.: Struc., Func. & Gen. **50**, 437-450

Macadam M. and Knowles C. (1985). The stereospecific bioconversion of alpha-aminoproprionitrile to L-alanine by an immobilized bacterium isolated from soil. Biotech. Lett. **7**, 865-870.

Mascharak P. K. (2002). Structural and Functional Models of Nitrile Hydratase. Coord. Chem. Rev. **225**, 201-214.

McCullagh p. and Nelder J. A. (1983). Generalised Linear Models. pub. Chapman and Hall London. UK. Pp. 209-221.

Miyanaga A., Fushinobu S., Ito K., Shoun A., Wakagi T. (2004). Mutational and structural analysis of cobalt-containing nitrile hydratase on substrate and metal binding. Eur. J. Biochem. **271**, 429–438

Miyanaga A., Fushinobu S., Ito K., Wakagi T. (2001). Crystal structure of cobalt-containing nitrile hydratase. Biochem. Biophys. Res. Commun. **288**, 1169–1174.
Murshudov G. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst. Section D. **53**, 240-255.

Nagasawa T. and Yamada H. (1987). Nitrile hydratase is a quinoprotein. A possible new function of pyrroloquinoline quinone: activation of H_2O in an enzymatic hydration reaction. Biochem. Biophys. Res. Commun. **147**, 701–709.

Nagasawa T., and Yamada H. (1989). Microbial transformation of. nitriles. Trends Biotechnol. **7**, 153-158.

Nagasawa T., Nanba H., Ryuno K., Takeuchi K. and Yamada H. (1987). Nitrile hydratase of *Pseudomonas chlororaphis* B23. Eur. J. Biochem. **162**, 691–698.

University of Cape Town

Nagasawa T., Takeuchi K., Yamada H. (1991). Characterization of a new cobalt-containing nitrile hydratase purified from urea-induced cells of *Rhodococcus rhodochrous* J1. Eur. J. Biochem. **196**, 581–589.

Nagashima S., Nakasako M., Dhomae N., Tsujimura M., Takio K., Odaka M., Yohda M., Kamiya N., Endo I. (1998) Novel non-heme iron center of nitrile hydratase with a claw setting of oxygen atoms. Nat. Struct. Biol. **5**, 347-351.

Nakasako M., Odaka M., Yohda M., Dohmae N., Takio K., Kamiya N. Endo I. (1999). Tertiary and quaternary structures of photoreactive Fe-type nitrile hydratase from *Rhodococcus* sp. N-771: Roles of hydration water molecules in structures and the structural origin of the substrate specificity of the enzyme. Biochemistry. **38**, 9887-9898.

Odaka M., Tsujimura M., Endo I. (2001). Post-translation modifications in nitrile hydratase family. RIKEN Rev. **41**, 58-60.

Otwinowski Z. and Minor W. (19970. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymol. **276**, A307-326.
Ouporov I. V., Knull H. R., Thomasson K. A. (1999). Brownian dynamics simulations of interactions between aldolase and G- or F-actin. Biophys. J. **76**,17-27

Padmakumar R and Oriel P. (1999). Bioconversion of acrylonitrile to acrylamide using thermostable nitrile hydratase. P. Appl. Biochem. Biotechnol. **77-9**, 671-679

Papworth C., Bauer J. C., Braman J., Wright, D. A. (1996). Mutagenesis in one day with >80% efficiency. Strategies. **9**, 3–4.

Payne M. S., Wu S., Fallon R. D., Tudor G., Stieglitz B., Turner I. M. Jr & Nelson M. J. (1997). A stereoselective cobalt-containing nitrile hydratase. Biochem. **36**, 5447-5454.

Pearson D. C. and Gross E. L. (1998). Brownian dynamics study of the interaction between plastocyanin and cytochrome f. Biophys. J. **75**, 2698-2711

Pereira R. A., Graham D., Rainey F. A., Cowan D. A. (1998). A novel thermostable nitrile hydratase, Extremophiles. **2**, 347–357.

Perutz M. F. (1989). Mechanisms of cooperativity and allosteric regulation in proteins. Q. Rev. of Biophys. **22**, 139-236.

Plugrath J. W. (1999). The finer things in X-ray diffraction data collection. Acta Cryst. Section D. **55**, 1718–1725.

Precigou S., Goulas P., Duran R. (2001). Rapid and specific identification of nitrile hydratase (NHase)-encoding genes in soil samples by polymerase chain reaction. FEMS Microbiol. Lett. **204,** 155-161.

University of Cape Town

Sali A. and Blundell T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. **234**, 779-815.

Sambrook J. and Russell D.W. (2001). Molecular cloning: a laboratory manual. 3rd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Scarrow R.C., Brennan B.A., Cummings J.G., Jin H., Duong D.J., Kindt J.T., Nelson, M.J. (1996) X-ray spectroscopy of nitrile hydratase at pH 7 and 9. Biochemistry. **35**, 10078-10088.

Serrano L., Bycroft M., Fersht A. (1991). Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. J. Mol. Biol. **218,** 465-75.

Shimuzu S. (2001). Special substances: Vitamins and related compounds: microbial production. In " Biotechnology" (H. Rehm, Ed.), Vol.10. Pp. 322-326. Wiley-VCH, Weinheim, Germany.

Takashima Y., Mukumoto F., Mitsuda S. (1996). Process for the production of amide compounds using microorganism. U.S. patent 5,563, 053.

Takashima Y., Yamaga Y., Mitsuda S. (1998). Nitrile hydratase from a thermophilic *Bacillus smithii*. J. Ind. Microbiol. Biotechnol. **20**, 220- 226.

Thomas S., Dicosimo R., Nagarajan V. (2002). Biocatalysis: applications and potentials for the chemical industry. Trends Biotechnol. **20**, 238–242.

Thompson L., Knowles C., Linton E., J., W. (1988). Microbial biotransformation of nitriles. Chem. Br. **24**, 900.

Tickle I. J., Laskowski R. A., David M. S. (1998). R_{free} and the R_{free} ratio. Part 1: derivation of expected values of cross- validation residuals used in macromolecular least- squares refinement. Acta Cryst. **D54**, 547- 557.

Tsekoa T. (2005). Structure, function and Enzymology of *Bacillus* sp. RAPc8 nitrile hydratase. Ph.D. thesis, University of the Western Cape.

Tsekoa T. L, Sayed M. F., Cameron R. A., Sewell B. T., Cowan D. A. (2004). Purification, crystallization and preliminary X-ray diffraction analysis of thermostable nitrile hydratase. S. Afr. J. Sci. **100**, 488-490.

Vaguine A. A., Richelle J., Wodak S. J. (1999). SFCHECK: a unified set of procedure for evaluating the quality of macromolecular structure-factor data and agreement with atomic model. Acta Cryst. **D55**, 191-205.

Vajda S. And Camacho C.J. (2004). Protein–protein docking: is the glass half-full or half-empty? Trends Biotechnol. **22**, 110– 116.

Wieser M., Takauchi T., Wada Y., Yamada H., Nagasawa T. (1998) Low-molecular-mass nitrile hydratase from *Rhodococcus rhodochorous* J1: purification, substrate

University of Cape Town

specificity and comparison with analogues high-molecular-mass-mass enzyme. FEMS Microb. Lett. **169**, 17- 22.

Wu S., Fallon R., Payne M. (1997). Over-production of stereoselective nitrile hydratase from Pseudomonas putida 5B in Escherichia coli: activity requires a novel downstream protein. Appl. Microbiol. Biotechnol. **48**, 704–708.

Yamaki T., Oikawa T., Ito K., Nakamur T. (1997). Cloning and sequencing of a nitrile hydratase gene from *Pseudonocardia thermophila* JCM3095. J. Ferment. Bioeng. **83**, 474– 477.

Zwart P.H., Langer G.G., Lamzin, V.S. (2004). Modelling bound ligands in protein crystal structures. Acta Cryst. Sec. D. **60**, 2230-2239.

University of Cape Town