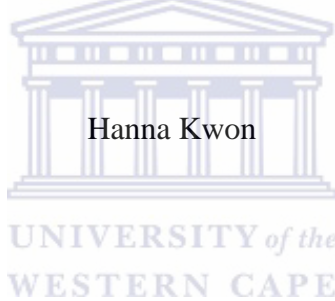


**Cloning, Expression and Characterisation of
Amidase Genes from a Psychrotolerant
Nesterenkonia isolate**

by



A thesis submitted to the Department of Biotechnology
in partial fulfillment of the requirements for the degree of Magister Scientae

Nov, 2009

Abstract

A nitrile and amide hydrolysing *Nesterenkonia* sp. was isolated from Antarctic soil and was characterised as a psychrotolerant, halotolerant and alkaliphilic extremophile. Amidases are widely distributed in both prokaryotic and eukaryotic organisms. These enzymes hydrolyze C-N bonds other than peptide bonds and are particularly interesting for their potential industrial application. This study aimed to identify and characterize amidase genes from this novel psychrotolerant microorganism.

Using BLAST analysis, two ORFs with conserved amidase sequences were identified from the complete genome sequence of the organism. Two ORFs, AmiF and AmiS, were assigned to two different gene families, the aceta/formamidase family and amidase signature family, respectively. On the genome, the spatial orientation and intergenic distance (1bp overlap) of the ORF's suggested that *amiF* and *amiS* could possibly be co-transcribed which was confirmed by reverse transcription PCR. A third ORF with a conserved amidase sequence was found ± 500 bps downstream from *amiS*, suggesting the possible presence of a multi amidase operon. The two genes were cloned and expressed as N-terminal 6x His-Tag fusion proteins. AmiS and Ami F were partially purified using Ni-chelation chromatography. Although both proteins were subjected to activity assay, their activities are yet to be established.

Homology modeling of the AmiF and AmiS translated sequences showed that the proteins had the significant similarities to the members of their families. Although the sequence identities between the AmiF and AmiS and their templates were very low (24 % and 25% respectively), the evaluation of the models showed that the quality of the models were good.

This study reports the genetic and functional characterisation of amidase genes from the cold adapted microorganisms.



Acknowledgements

I would like to express my most sincere appreciation to the following people:

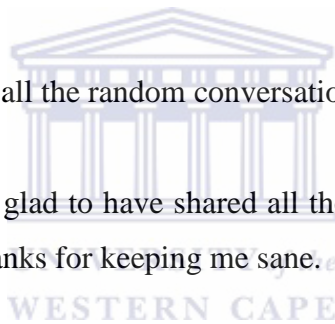
Professor Don A. Cowan for his support, encouragement and patience throughout this study. I'm grateful to have worked under his supervision.

Dr. Marla Tuffin and Dr. Heide Goodman for their day to day running of the lab as well as their support and assistance.

I also would like extend my appreciation to Professor. Cristoph Syldatk for his support and more importantly the kindness he showed to us during my visit to Germany.

Mr. Lonnie van Zyl, thanks for all the random conversations, discussions and support.

To Peter, my gtalk buddy, I'm glad to have shared all those muffins with you despite of the weight we have gained. Thanks for keeping me sane.



To William, Thanks for all your support over gtalk. I envy your wisdom.

To Heidi Pasques, I really enjoy and appreciate our inconsistent correspondence. You have been a good teacher and more than that a good friend to me.

To Jay who stood by me and supported me through everything. Thanks for explaining to people what I do when I fail to.

To my friends who are back at home, I really appreciate your continued support, love and encouragement. I'm proud to have you guys as my friends.

To my family...

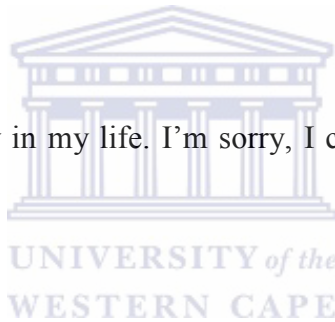
Thanks for all your prayers and the unconditional love and support.

Dad, thanks for all your support even when you were going through tough time. Without your continued support, I would not have made it.

Mom you have inspired me and I respect you for having courage to study and getting into college to achieve your goals and dreams while building your own career.

My brother and sister-in-law, I appreciate all your support and love but most of all thank you so much for giving me a niece. She made me realize the meaning of the unconditional love.

To my niece, you are little joy in my life. I'm sorry, I could not spend more time with you.



Declaration

I hereby certify that all of the work described within this thesis is the original work of the author. Any published (or unpublished) ideas and/or techniques from the work of others are fully acknowledged in accordance with the standard referencing practices.

Hanna Kwon

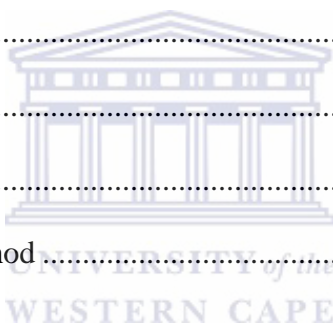
Nov, 2009



Table of Contents

Abstract	ii
Acknowledgements	iv
Declaration	vi
Chapter 1 Literature Review	1
1.1 Introduction	1
1.2 Classification of Amidases	2
1.2.1 Nitrilase related Amidases	2
1.2.1.1 Aliphatic Amidases	7
1.2.2 Signature Amidases	9
1.3 Nitriles and Amides Metabolism in Bacteria	14
1.4 Applications of Nitrile Metabolising Enzymes	18
1.5 Cold Adaptation	19
1.6 Protein Structure Modeling	21
1.6.1 Fold Assignment	23
1.6.2 Target-Templates Alignment	23
1.6.3 Model Building and Side Chain Optimization	25
1.6.4 Model Evaluation	26
1.7 Reliability of Comparative Modeling Methods	26
1.8 Aim	28
Chapter 2 General Materials and Methods	29

2.1 Chemicals and Reagents.....	29
2.2 Bacterial Strains and Plasmids	30
2.3 Media.....	31
2.4 DNA extractions.....	34
2.5 DNA Purification	35
2.6 Amplification of Amidase Genes	35
2.7 Preparation of Electro-Competent Cells	36
2.8 Preparation of Chemically Competent Cells.....	38
2.9 Cloning	38
2.10 Electroporation	39
2.11 Heat Shock	39
2.12 Minipreps	40
2.12.1 Alkaline Lysis Method	40
2.12.2 Miniprep Kits.....	41
2.13 Screening	41
2.13.1 Restriction Fragment Analysis	41
2.14 Sequencing	42
2.15 Electrophoresis	42
2.15.1 Agarose Gel Electrophoresis	42
2.15.2 Denaturing SDS-polyacrylamide Gel Electrophoresis	42
2.15.3 Native Polyacrylamide Gel Electrophoresis.....	43



Chapter 3 Sequence Analysis and Homology Modeling	45
3.1 Introduction	45
3.2 Sequence Analysis.....	46
3.3 Phylogenetic Analysis.....	52
3.4 Homology Modeling	54
3.4.1 Search for Templates	55
3.4.2 Pairwise Sequence Alignment	55
3.4.3 Model Building and Side-Chain Optimisation	57
3.4.4 Model Evaluation	60
3.4.5 Description of AmiF and AmiS homology models	63
3.5 Discussion	66
Chapter 4 Expression and Characterisation of Amidases	70
4.1 Introduction	70
4.2 Materials and Methods	72
4.2.1 Expression of AmiF.....	72
4.2.2 Expression of AmiS.....	72
4.2.4.1 Urea Solubilisation.....	74
4.2.4.2 Two Step Refolding	74
4.2.4.3 Chloramphenicol Method	75
4.2.4.4 β -mercaptoethanol Method	75
4.2.4.5 Sodium Deoxycholate Method	76

4.2.5 His-Tag Affinity Chromatography	76
4.2.6 RNA Extraction	77
4.2.7 Reverse Transcription PCR (RT-PCR).....	78
4.2.8 Activity Assay	80
4.3 Results and Discussion.....	81
4.3.1 Cloning of AmiF and AmiS.....	81
4.3.2 Expression and Purification of Ami F	81
4.3.3 Expression of Ami S.....	83
4.3.4 Solubilisation of Ami S	86
4.3.4.1 Urea Solubilisation Method	86
4.3.4.2 Two Step Refolding Method (Urea based)	91
4.3.4.3 Chloramphenicol Method	92
4.3.4.4 β -mercaptoethanol Method	94
4.3.4.5 Purification of Inclusion Bodies using Na. Deoxycholate.....	96
4.3.4.6 Effect of His-Tags on Solubility of AmiS	98
4.3.4.7 Co-factors.....	98
4.3.5 Reverse Transcription PCR	99
4.3.6 Ammonia Detection Assays	104
Chapter 5 General Discussion.....	107
References.....	112

List of Figures

Figure 1-1 The nitrilase superfamily catalytic triad motifs.	4
Figure 1-2 Modeled structural fold of an amidase from <i>Pseudomonas aeruginosa</i> (PDB ID; 1k17).....	4
Figure 1-3 Multiple alignment of the sequences of microbial nitrilases with those of the eleven members in the nitrilase superfamily for which structural information is available.	6
Figure 1-4 Model of transcriptional regulation of <i>P. aeruginosa</i> aliphatic amidase operon.	8
Figure 1-5 Alignment over the amidase signature sequence region of all members of the GGSS motif containing amidases.....	11
Figure 1-6 Sequence alignments and secondary structure assignment.....	13
Figure 1-7 Crystal structure of the MAE2 monomer.....	14
Figure 1-8 Different pathways of nitrile metabolism	15
Figure 1-9 Catalytic mechanisms of nitrile hydrolysing enzymes.....	16
Figure 1-10 Reactions catalyzed by amidases	17
Figure 1-11 Possible catalytic mechanism for the nitrilase reaction.	18
Figure 1-12 Steps in comparative protein structure modeling.....	22
Figure 3-1 Genetic map of two amidases and neighboring ORFs identified on <i>Nesterenkonia</i> sp.	47
Figure 3-2 Nucleotide and deduced amino acid sequences of AmiF.....	49

Figure 3-3 Alignment of AmiF, Formamidase from <i>Methylophilus methylotrophus</i> and Acetamidase from <i>Mycobacterium smegmatis</i> showing conserved sequences.....	50
Figure 3-4 Nucleotide and deduced amino acid sequences of AmiS.....	52
Figure 3-5 Neighbour-joining phylogenetic dendrograms of AmiF (A) and AmiS (B)...	53
Figure 3-6 Alignment between AmiF and 2ii1	56
Figure 3-7 Alignment between AmiS and 2f2a	56
Figure 3-8 Cartoon representation of AmiF before (A) and after (B) side-chain optimisation (c) Superimposed (A) and (B).....	58
Figure 3-9 Cartoon representation of AmiS before (A) and after (B) side-chain optimisation (c) Superimposed (A) and (B).....	59
Figure 3-10 Ramachandran plots for the template (A, PDB ID: 2ii1) and the AmiF model (B).	62
Figure 3-11 Ramachandran plots for the template (A, PDB ID: 2f2a) and the AmiS model (B).	62
Figure 3-12 Superimposed model and template structures.....	63
Figure 3-13 AmiS monomer fold.....	64
Figure 3-14 Superimposed catalytic triad of the AmiS model (pink) and the template (blue, PDB ID: 2f2a).....	65
Figure 3-15 Superimposition of 1OCK (cyan) and AmiS model (pink) (A) and the residues involved in catalysis (B)	66
Figure 4-1 Expression of Ami F in <i>E.coli</i> Rosetta2 (DE3) pLysS with 0.4mM IPTG.....	82

Figure 4-2 His-Tag Purification of AmiF	83
Figure 4-3 Expression of AmiS in <i>E.coli</i> Rosetta2 (DE3) pLysS with 0.4mM IPTG induced at various temperature	84
Figure 4-4 Optimisation of IPTG concentration for AmiS	85
Figure 4-5 SDS-PAGE of Solubilised of AmiS by Urea	87
Figure 4-6 His-Tag purification of AmiS solubilised by 6M urea.....	89
Figure 4-7 Native-PAGE of refolded AmiS	90
Figure 4-8 His-Tag steps of two step refolding of AmiS solubilised by 6M urea.....	91
Figure 4-9 Native-PAGE gel of AmiS refolded by two step refolding method	92
Figure 4-10 Solubilization of AmiS by Chloramphenicol Method	93
Figure 4-11 His-Tag Purification of the solubilised protein by beta-mercaptoethanol	94
Figure 4-12 SDS-PAGE analysis of inclusion body purification and solubilisation of AmiS using DOS.....	96
Figure 4-13 Analysis of RNA extraction on an agarose gel	100
Figure 4-14 PCR on RNA before DNase treatment.....	101
Figure 4-15 PCR reactions on RNA after DNase treatment	102
Figure 4-16 Reverse transcription PCR	103

List of Tables

Table 2-1 Bacterial Strains Used in This Study.....	30
Table 2-2 Plasmids Used in This Study.....	30
Table 2-3 Primers used and PCR condition.....	37
Table 4-1 Reverse Transcription PCR Reaction Set Up.....	79
Table 4-2 Amidase substrates used for ammonia detection assay	106



Chapter 1

Literature Review

1.1 Introduction

The formation and hydrolysis of the carbon-nitrogen (C-N) linkages are essential processes in both eukaryotic and prokaryotic organisms. Although peptide bonds are the most well-studied class of carbon-nitrogen bonds, there are other non-peptide carbon-nitrogen linkages whose metabolism is less well studied (Brenner, 2002).

Compounds containing carbon-nitrogen bonds are widespread in nature. Many nitrile hydrolysing bacterial species have been isolated from diverse ecosystems including deep sea trenches, nitrile contaminated soil and thermal lake sediments. Nitriles and amides metabolising bacteria are typically isolated using trophic selection strategies, where substrates are used as the sole carbon and/or nitrogen sources to support the growth of these organisms (Cowan *et al.*, 2003).

C-N bond containing compounds include organic cyanides or nitriles ($R-C\equiv N$), inorganic cyanides ($H-C\equiv N$), acid amides [$R-C(=O)-NH_2$], secondary amides [$R-C(=O)NH-R'$] and N-carbamyl amides [$R-NH-C(=O)-NH_2$]. The hydrolysis of these compounds is mainly performed by nitrile hydrolysing enzymes, namely Nitrilases (Nases), Nitrile

Hydratases (NHases) and amidases from the nitrilase superfamily (Pace & Brenner, 2001). These enzymes attack either the cyano carbon of a linear nitrile or planar carbon of amides using a conserved cysteine residue (Bork & Koonin, 1994). Signature amidases and amido-transferases also act on C-N bonds even though their sequences and structures are un-related to the nitrilase superfamily enzymes (Brenner, 2002).

1.2 Classification of Amidases

Although several previous attempts have been made to classify amidases, the classification of the proteins is not definitively formulated (Pertsovich, 2005). However, based on their sequence and structure similarities, the amidases are broadly divided into two groups. Group 1 consists of amidases that are structurally related to the nitrilase superfamily members while amidases in group 2 belong to the signature amidase family and are structurally unrelated to nitrilases (Farnaud *et al.*, 1999; Kimani *et al.*, 2007; Pertsovich, 2005).

1.2.1 Nitrilase related Amidases

Nitrilase related amidases belong to the nitrilase superfamily. Initially, members of the nitrile-hydrolysing superfamily included nitrilases, cyanide hydratases, aliphatic amidases, β -ureidopropionases, β -alanine synthetases and N-carbamyl-D-amino acid amidohydrolases (Bork & Koonin, 1994). However, on the basis of sequence similarity

and domain-fusion characteristics, the nitrilase superfamily was re-classified into 4 major groups (nitrilases, amidases, carbamylases and N-acyltransferases) (Pace & Brenner, 2001) distributed in 13 different branches. Only one branch of the nitrilase superfamily has the members that catalyse nitrilase substrates, whereas the other 12 branches consist of amidases with distinct specificities (Brenner, 2002). In the nitrilase superfamily, the protein fusions have been common occurrence, and assist in functional classification (Brenner, 2002).

Branch 1 enzymes have nitrilase activity and consist of nitrilases, cyanide dihydratases and cyanide hydratases (Pace & Brenner, 2001). Branch 2 through branch 9 consists of amidases with varying substrate specificities. Branch 10 is represented by a fusion protein, NitFhit, where a nitrilase domain is fused to a nucleotide binding Fhit protein (Pace & Brenner, 2001). Branch 11 to 13 members are grouped according to their sequence relation as their functions are not clear (Pace & Brenner, 2001). In seven of these, nitrilase superfamily branch is fused to another domain (Pace & Brenner, 2001).

The nitrilase superfamily members share conserved catalytic triad residues (EKC), four glycine residues as well as an extra glutamine residue (Kimani *et al.*, 2007)

Figure 1-1). The extra glutamate has been implicated in the nitrilase reaction mechanism in *G. pallidus* aliphatic amidase (Kimani *et al.*, 2007). All members of the nitrilase superfamily have a signature sequence that flanks the catalytic triad.

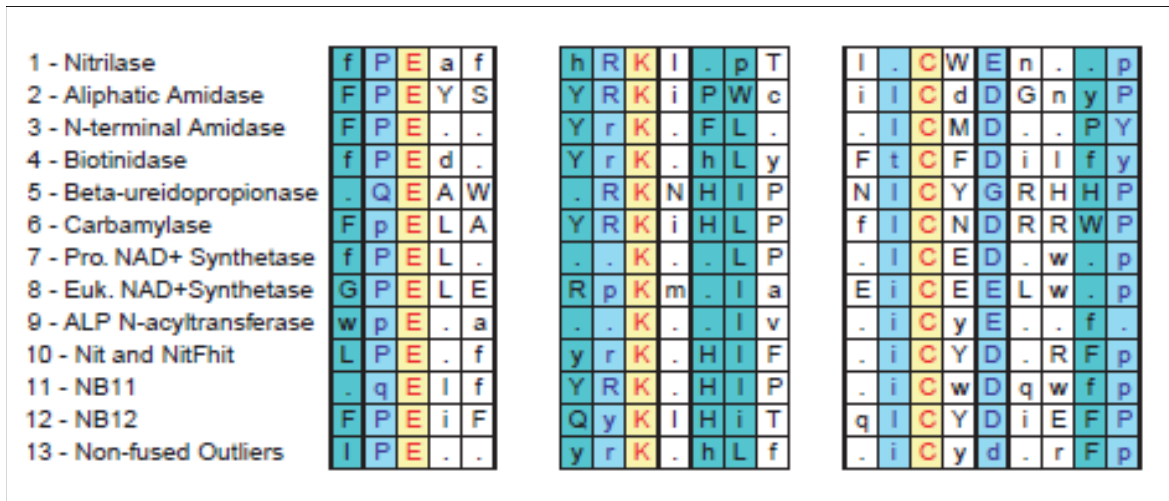


Figure 1-1 The nitrilase superfamily catalytic triad motifs.

Red letters in a yellow background indicate the same residue is conserved for all the branches. Dark blue letters on light blue background indicate the residue is conserved in nine or more branches. Upper case letters indicate 90% or greater consensus levels within a branch, whereas lower case are 50% or greater. Figure taken from Pace & Brenner, 2001.

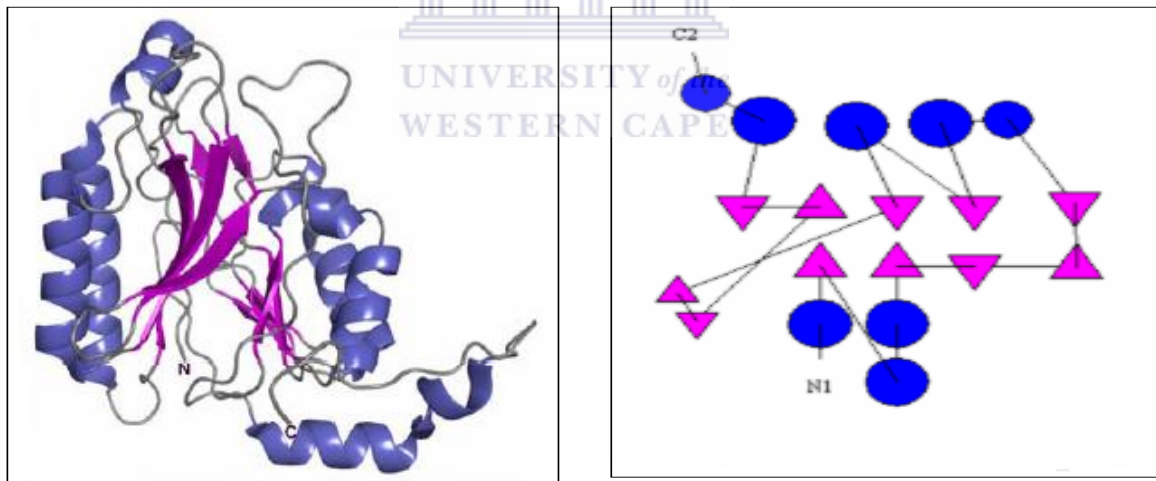


Figure 1-2 Modeled structural fold of an amidase from *Pseudomonas aeruginosa* (PDB ID; 1k17).

Cartoon diagram of PamE (Novo *et al.*, 2002) (left), Secondary structure topology diagram of the same protein (right). Helices are shown in blue, β -sheets are shown in magenta and the loops in grey. Topology diagram shows conserved $\alpha\beta\alpha$ sandwich fold of the core region. Figure taken from Novo *et al.*, 2001.

Although the sequence alignment between the nitrilase superfamily members with solved structures show low sequence conservation (Figure 1-3), the three dimensional (3D) structures of this family are highly similar. All structures shared a characteristic monomer fold of $\alpha\beta\beta\alpha$ (Thuku *et al.*, 2009). Most of the characterised nitrilase related amidases form homotetrameric and homohexameric complexes in solution (Pertsovich, 2005). These enzymes show preference towards the short chain aliphatic amides as their substrates (Kimani *et al.*, 2007; Pertsovich *et al.*, 2005). There are few crystal structures of nitrilase related amidases (PDB ID; 2plq, 3HKS) available in addition to a homology model of *Pseudomonas aeruginosa* amidase (PDB ID; 1k17, Figure 1-2). These structures show a conserved $\alpha\beta\beta\alpha$ sandwich fold that resembles the conserved structural fold of the nitrilase superfamily. All of the structures also contain the catalytic triad residues (Glu, Lys and Cys) which support the classification of these amidases in the nitrilase superfamily.

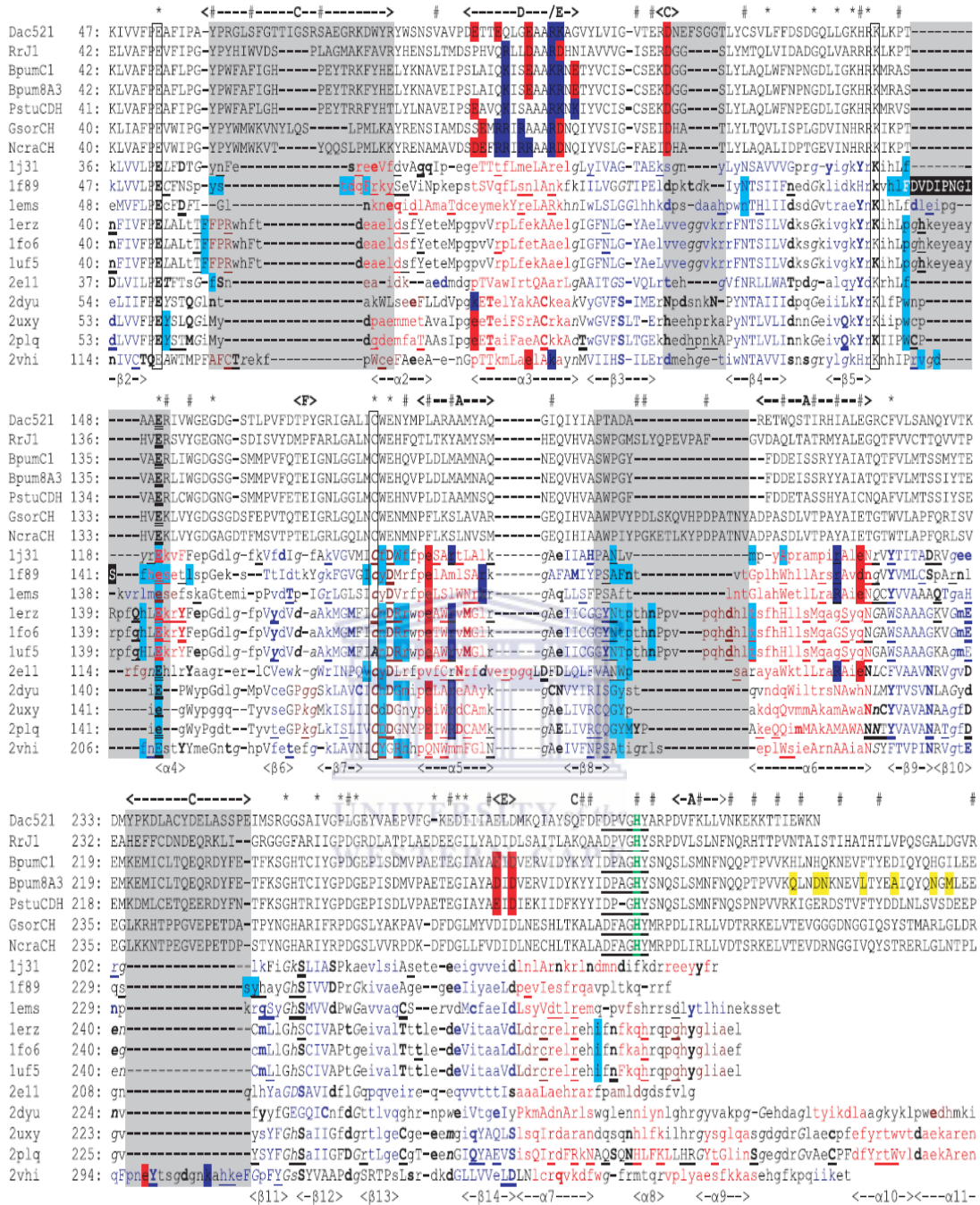


Figure 1-3 Multiple alignment of the sequences of microbial nitrilases with those of the eleven members in the nitrilase superfamily for which structural information is available.

The protein sequences of Nases of *G. pallidus* DAC521(Dac521, accession no ABH04285) and *R. rhodochrous* J1 (RrJ1, Accession no BAA01994), and the 10 members of the nitrilase superfamily for which structural information is available. The secondary structural elements identified for 2plq (Kimani et al. 2007) are indicated in the bottom line. The conserved catalytic residues (CEK) are shown in a box outline. The additional glutamate catalytic residue is in bold and double underlined. # indicates the locations of the catalytic residues and four glycine residues that are conserved in all members of the nitrilase superfamily. The conserved sequence motif 'DP/FXGHY' in the tail region of the spiral forming Nases is in bold and underlined. The approximate regions of the interacting surfaces of Nases, namely 'A', 'C', 'D' and 'E' are indicated on the top line. Figure taken from Thuku et al. (2009).

1.2.1.1 Aliphatic Amidases

Amidases belong to the hydrolase family which is a subclass of acrylamide amidohydrolases. This subclass consists of enzymes which hydrolyse linear amides (EC 3.5.1) and cyclic amides (EC 3.5.2).

The aliphatic amidases have been identified within NHases operons in *Bacillus* sp. BR449 and *G. pallidus* (Cameron *et al.*, 2005). Due to their proximity in the operon, co-expression of the genes has been suggested (Cameron *et al.*, 2005).

The gene regulation of an amidase operon from *Pseudomonas aeruginosa* is one of the most well characterised amidase operon. The total size of the operon is about 5kb and the amidase is activated when short chain amides are present in the growth media. When the genes are not induced, there is constitutive expression of the operon from the pE promoter which is regulated by catabolite repression that terminates at transcription terminator sequence T1 and T2 (Figure 1-4) (Wilson & Drew, 1995). When the genes are induced, the substrate binds to the AmiC - AmiR complex, resulting in a conformation

change, releasing AmiR from the inhibitor AmiC (Wilson & Drew, 1995). Free AmiR then induces the expression of the other genes in the operon through an anti-termination reaction at T1 (Wilson & Drew, 1995).

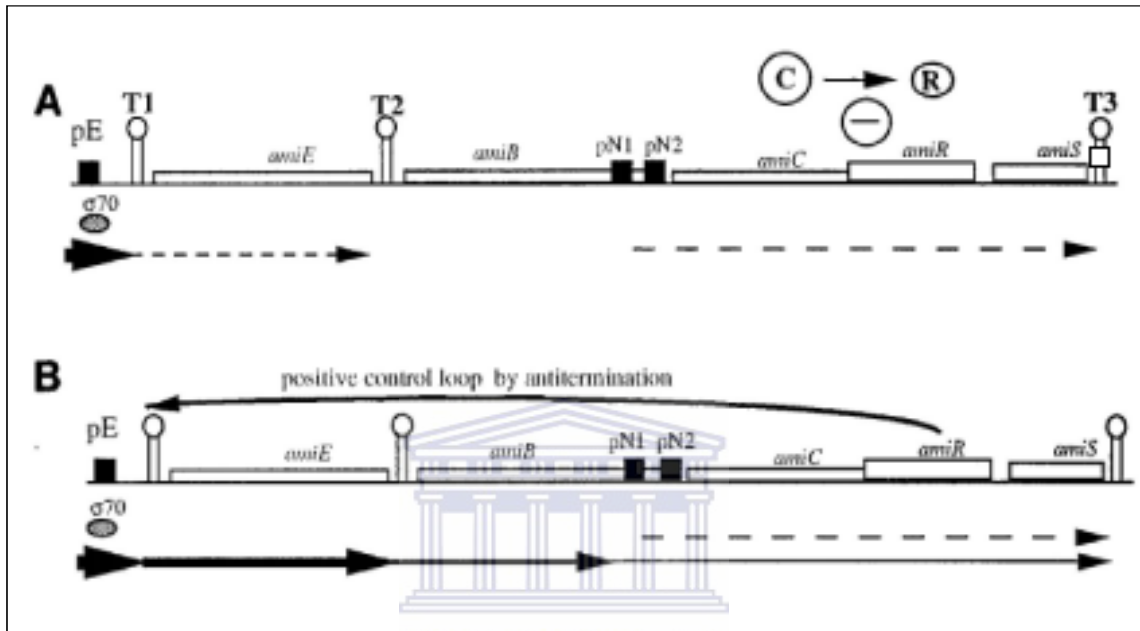


Figure 1-4 Model of transcriptional regulation of *P. aeruginosa* aliphatic amidase operon.

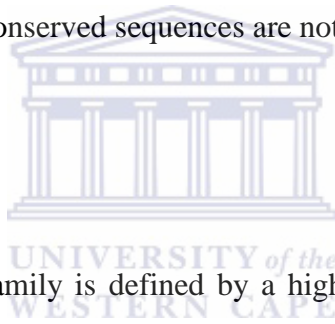
A: under non-inducing condition, B: under inducing condition. Arrow thickness represents the level of transcription for growth. Promoter sequences are shown ■ and labeled. Transcription termination sequences are shown as stem loops and labeled. AmiR antitermination is RNA sequence dependent and occurs only at the T1 terminator. Figure taken from Wilson & Drew, (1995).

Although most common activity of amidases is an amide hydrolysis reaction, some amidases are known to exhibit acyl transfer activity in the presence of hydroxylamine. Other reactions that have been observed are acid transfer, ester hydrolysis and ester transfer (Fournand *et al.*, 1997; Fournand *et al.*, 1998a; Fournand *et al.*, 1998b).

However, these reactions are much slower than amide hydrolysis (Fournand & Arnaud, 2001). Many studies have revealed aliphatic amidases exhibit broad substrate specificities as well as stereo- and enantio-selectivity.

Although a number of studies of amidase classification (Chebrou *et al.*, 1996; Fournand & Arnaud, 2001) have shown that amidases fall into two categories, there are few aliphatic amidases that lack both conserved sequences, a novel catalytic triad EKC and GGSS signature sequence. These amidases belong to aceta/formamidase family most of its members contain EKC residues hence also belong to nitrilase superfamily. However, other members that lack both conserved sequences are not well studied.

1.2.2 Signature Amidases



The amidase signature (AS) family is defined by a highly conserved ~130 amino acid region called the amidase signature sequence (Shin *et al.*, 2002). These enzymes are typically homodimers of approximately 110kDa. The amidase signature family consists of more than 200 proteins derived from bacteria to humans (Labahn *et al.*, 2002). Among those that have been characterised, most have shown hydrolase activity while a few act as amide transferases in complex with other proteins (Labahn *et al.*, 2002).

Signature amidases are characterised by the presence of an invariant GGSS (Gly-Gly-Ser-Ser) signature motif in their primary sequence (Fournand & Arnaud, 2001;

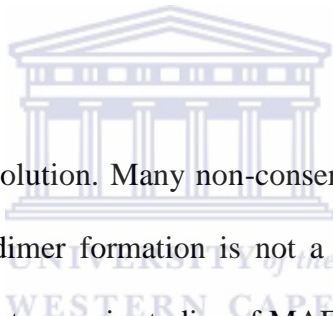
Kobayashi *et al.*, 1997) (Figure 1-5). They also have aspartate and serine residues in the active site, replacing the cysteine residue that is found in the nitrilase superfamily (Novo *et al.*, 1995). Signature amidases exhibit wide substrate specificity including aliphatic and aromatic amides as well as α -substituted carboxylic acids. Signature amidases also exhibit stereo-selectivity (Kobayashi *et al.*, 1997). In addition to the C-N bond metabolism, signature amidases are also reported to transfer NH_3 from glutamine, resulting in the generation of properly charged tRNA^{Gln} (Curnow *et al.*, 1997). This is mostly carried out by heterotrimeric Glu- tRNA^{Gln} whose subunit A belongs to AS family. This function of the family protein is very important as it provides the precursor for protein biosynthesis (Shin *et al.*, 2002). Signature amidases that have an additional CX_3C motif are capable of nitrile hydrolysis (Kobayashi *et al.*, 1999). In mammals, proteins belonging to AS families are also involved in degradation of neuromodulatory fatty acids. Other studies have shown that the amidases from this family represent a large class of serine-lysine catalytic dyad hydrolases that resemble serine hydrolases (Boger *et al.*, 2000). However, the lack of aspartic acid and histidine in the sequences of AS family, indicates that the AS family enzymes are non-classical serine hydrolases (Shin *et al.*, 2002).

R312	170	GGSSGG	SAALVANG	DVDFAL	GGDQ	GGSIRI	-PAAF	CGVVG	HKPTFG	214
Rho	147	GGSSGG	TAAALAAG	LIFAGM	GTDT	GGSIRI	-PAAV	CGTVGL	KPTYG	191
J1-L	169	GGSSSG	SGALVAG	SGQVDM	AVGGD	QGGSIRI	-PAAF	CGIVG	HKPTFG	213
N-774	170	GGSSGG	SAALVANG	DVDFAL	GGDQ	GGSIRI	-PAAF	CGVVG	HKPTFG	214
B23	171	GGSSSG	SAALVAG	SGEVDI	AVGGD	QGGSIRI	-PSAF	CGTYG	MKPTFG	214
Asp-f2	12	LD	SQYFGKIYL	GTPPQ	EFTVLE	DTGSS	DFWV	PSIY	CKSNACK	NHQR
Asp-f1	10	LDTEY	FGTIGIG	TPAQD	TFVIF	DTGSS	NLWV	PSVY	CSSLAC	SDHNQ
Asp-11	196	SITMD	GETIAC	-SGGCQ	-AIV	-DTGT	SLLTG	PTS	SAI	AINIQSDIGA
Asp-12	197	SVTIS	GVVAC	-EGGCQ	-AIL	-DTGT	SKLVG	PSSDI	-LNI	QQAIGA
VDHAP	222	GGSSGG	EGALIAC	GGSL	LIGIGS	DVAGSIRL	-PSSF	CGLC	GLKPTGF	266
FAAH	215	GGSSGG	EGALIG	SGGSP	LGLGTDI	GGSIRF	-PSAF	CGIC	GLKPTGN	259
Yeast	207	GGSSGG	EGSLIG	CAHGS	LLGLGTDI	GGSIRI	-PSSY	QGLF	GLKPTFG	251
EI	148	GGSSGG	SGAAV	AAALSP	VAHGN	DAAGSVRI	-PASV	CCVVG	LKPTRG	192
IND-P	144	GGSSGG	SAAAVAS	GI	VPLSVG	DTGG	SIRI	-PAAF	CGITG	FRPTIG
IND-A	146	GGSSGG	VAAAVAS	RMLGG	IGTD	GASVRL	-PAAL	CGVVG	FRPTIG	190
IND-B	147	GGSSGG	AGSAVA	AGIGTIA	HGNDI	GGSIRW	-PAHC	NGVATIK	PTOG	191
ACE-O	203	GGSSGG	EGAMI	AMRGA	IGTIGTDI	GGSIRV	-PAAF	N	SLYGI	RPSHD
ACE-N	202	GGSSGG	EGALIVG	IRGGVIG	VGTDI	GGSIRV	-PAAF	N	LYGLR	PSHG
Com	153	GASSSG	SGVATA	AGLCYAS	IGTD	GGSIRF	-PAAAN	GLIGI	KPTWG	197
Urea	151	GGSSSG	SGSVV	ARGIAC	LALTTDT	AGSIRV	-PAAL	NNLIS	IKPSVG	195

Figure 1-5 Alignment over the amidase signature sequence region of all members of the GGSS motif containing amidases.

From top to bottom: R312: amidase from *Rhodococcus* sp. R312; Rho: amidase from *Rhodococcus* sp.; J1-L: amidase from *Rhodococcus rhodochrous* J1; N-774: amidase from *Rhodococcus* sp. N-774; B23: amidase from *Pseudomonas chlororaphis* B23; Asp-f2: the N-terminal region of calf chymosin; Asp-f1: the N-terminal region of porcine pepsin; Asp-11: the C-terminal region of porcine pepsin; Asp-12: the C-terminal region of calf chymosin; VDHAP: a vitamin D3 hydroxylase-associated protein from cockerel; FAAH: oleamide hydrolase from rat; Yeast: putative amidase from *Saccharomyces cerevisiae*; EI: EI enzyme from *Flavobacterium*; IND-P: indoleacetamide hydrolase from *Pseudomonas savastanoi*; IND-A: indoleacetamide hydrolase from *Agrobacterium tumefaciens*; IND-B: indoleacetamide hydrolase from *Bradyrhizobium japonicum*; ACE-O: acetamidase from *Aspergillus oryzae*; ACE-N: acetamidase from *Aspergillus nidulans*; Com: amidase from *Comamonas acidovorans*; Urea: urea amidolase from *Candida utilis*. Arrows indicate the amino-acid residues that correspond to Asp191, Ser195 and Cys203 of the *Rhodococcus rhodochrous* J1 amidase examined in this sequence similarity (Chebrou et al. 1996). Residues highlighted in reverse type are conserved across both amidases and aspartic proteinases in at least four of 20 sequences. Figure was taken from Fournand *et al.*, (2001).

Most of the structurally characterised signature amidases form homodimeric and homooctameric quaternary structure complexes in solution (Kobayashi *et al.*, 1997). Several crystal structures of signature amidases have been solved. The first structure solved in the AS family was malonamide E2 (MAE2; PDB ID: 1OCK). The structure has a mixed α/β fold which consists of 12 α -helices and one central and highly irregular β -sheet composed of 11 β -strands (Figure 1-6). The α/β sandwich fold with a central β -sheet core surrounded by α -helices is conserved across the members of the amidase signature family (Figure 1-7). The core of the molecule consists of the conserved AS sequence and interacts with a substantial part of remainder of the sequence (Shin *et al.*, 2002) (Figure 1-5).



The MAE2 forms a dimer in solution. Many non-conserved residues were found at the interface. This indicated that dimer formation is not a common feature of AS family enzymes (Shin *et al.*, 2002). Mutagenesis studies of MAE2 have revealed that the protein has a catalytic triad rather than the suggested catalytic dyad. The catalytic triad consists of Ser 155-Ser131-Lys 62 with unusual *cis* conformation of Ser 131(Shin *et al.*, 2002). As it was the first report of the catalytic triad found in hydrolase family, the structure of MAE2 presented a new class of serine hydrolases that has a novel catalytic triad in a unique protein fold (Shin *et al.*, 2002).

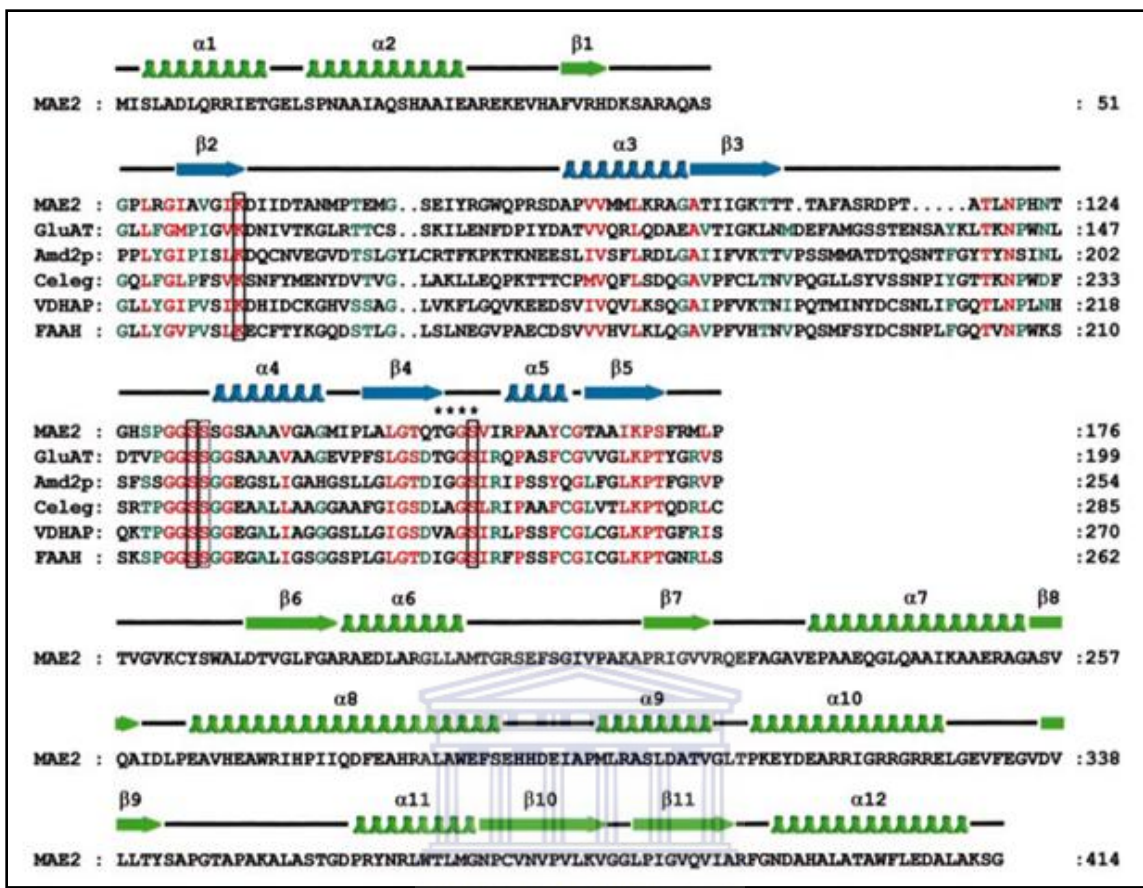


Figure 1-6 Sequence alignments and secondary structure assignment.

The AS sequences of selected AS family enzymes from low to high organisms are aligned: GluAT, Glu-tRNA^{Gln} amidotransferase, *Bacillus subtilis* (gi: 2589195); Yeast Amd2p, a putative amidase, *Saccharomyces cerevisiae* (gi: 6320448); Celeg, a putative amidase, *Caenorhabditis elegans* (gi: 6425411); VDHAP, vitamin D3 hydroxylase-associated protein, *Gallus domesticus* (gi: 1079452); FAAH, fatty acid amide hydrolase, *Homo sapiens* (gi: 4557575). The secondary structure assignment is shown at the top of the sequence. The solid-line boxes indicate the three invariant residues of the catalytic triad and the dotted-line box indicates the highly conserved residue that provides the augmenting catalytic role. Asterisks indicate the residues comprising the oxyanion hole. Figure was taken from Shin *et al.*(2002).

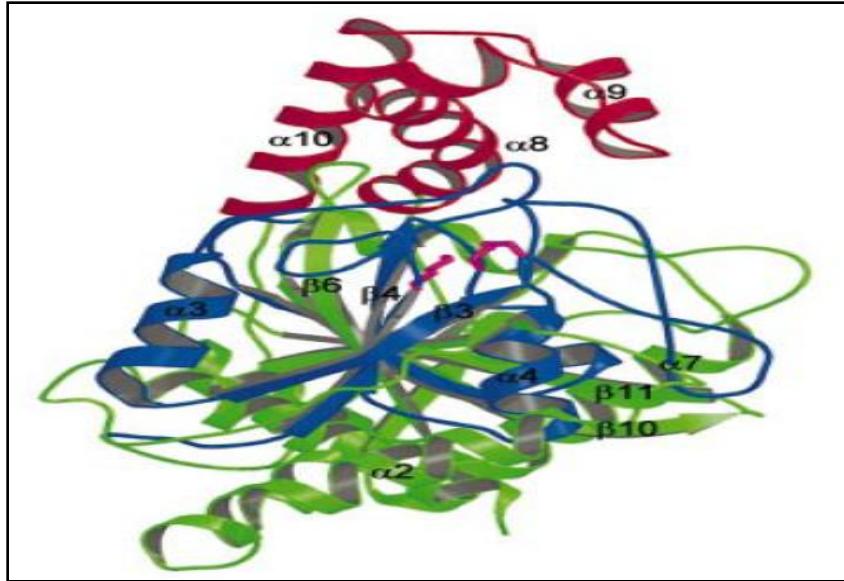
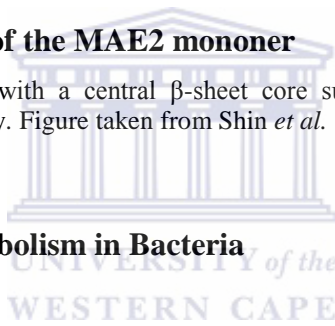


Figure 1-7 Crystal structure of the MAE2 monomer

MAE2 has the α/β sandwich fold with a central β -sheet core surrounded by α -helices. This fold is conserved in signature amidase family. Figure taken from Shin *et al.* (2002).



1.3 Nitriles and Amides Metabolism in Bacteria

The enzymatic hydrolysis of nitriles and amides is catalyzed by nitrile metabolising enzymes: Nases, NHases and amidases. These enzymes are found in plants and fungal families and most frequently in bacteria. Although the physiological role of nitrile degrading enzymes in micro-organisms is not clear, many bacteria e.g. *Acinetobacter*, *Corynebacterium*, *Arthrobacter*, *Pseudomonas* are known to metabolise nitriles as the sole source of carbon and nitrogen (Kato *et al.*, 1998; Kato *et al.*, 2000; Kobayashi *et al.*, 1998a). In plants, nitrile metabolising activities are implicated in nutrient metabolism as well as in the synthesis of indole acetic acid (Banerjee *et al.*, 2002). It has also been

suggested that, in plants, nitrile degrading enzymes are components of complex pathways controlling both production and degradation of cyanogenic glycosides and related compounds where aldoximes are the key intermediates (Banerjee *et al.*, 2002). Aldoxime dehydratases were found to be responsible for the formation of nitriles from aldoximes (Kato *et al.*, 2000). Nitriles then undergo hydrolysis, oxidation and reduction by various enzymes (Figure 1-8) (Banerjee *et al.*, 2002).

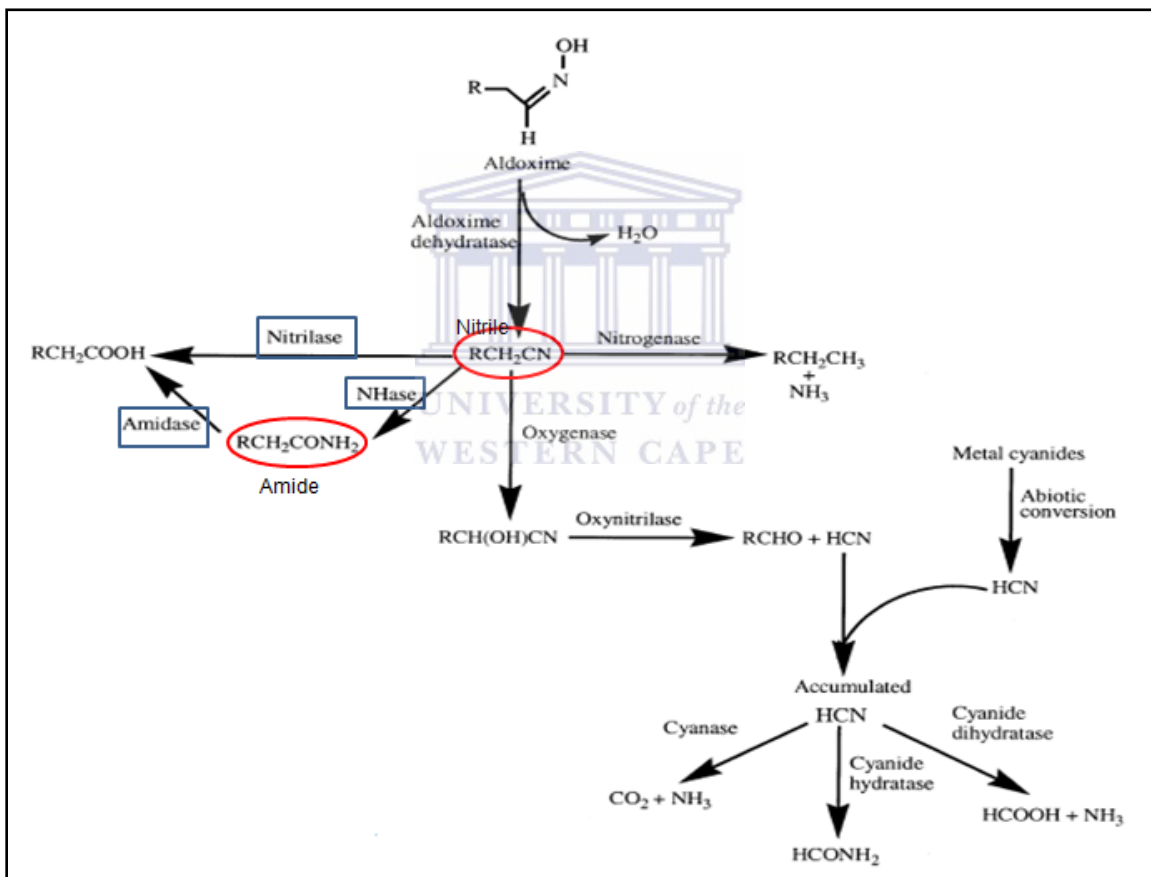


Figure 1-8 Different pathways of nitrile metabolism

Figure taken from Banjeree *et al.*, (2002).

Nitrile hydrolysis follows two enzymatic pathways. (i) A single enzymatic pathway catalysed by Nases involving conversion of organic nitriles to the corresponding acids and ammonia (Figure 1-9). (ii) A bi-enzymatic pathway that involves the hydration of nitriles to their corresponding amides by NHases, followed by the conversion of amides to their corresponding organic acids and ammonia by amidases (Figure 1-9). However, some amidases exhibit activities other than amide hydrolysis (Figure 1-10).

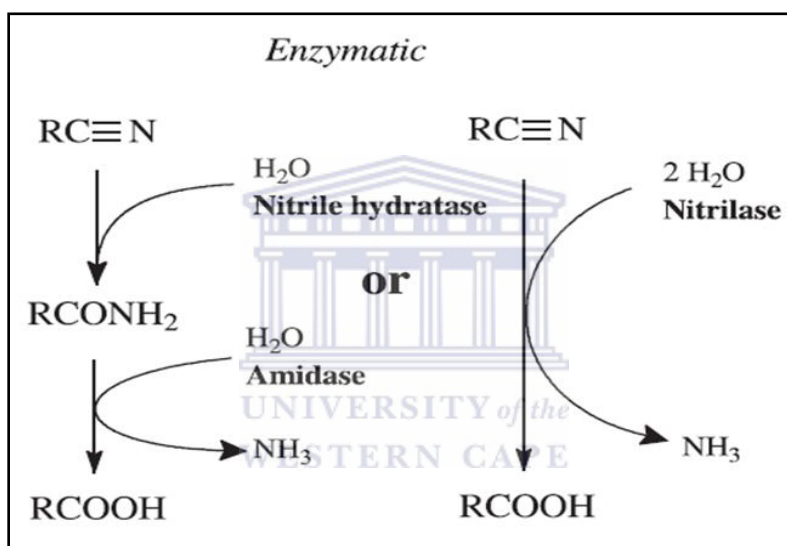


Figure 1-9 Catalytic mechanisms of nitrile hydrolysing enzymes

Figure was taken from Fournand and Arnaud (2001).

Following the discovery of an amidase from *Rhodococcus rhodochrous* J1 which hydrolyses nitriles, an hydrolytic mechanism that includes the nitrile hydrolysis was proposed by Kobayashi and colleagues (Kobayashi *et al.*, 1998b). In the proposed mechanism, the carbonyl group of the amide undergoes nucleophilic attack, resulting in

the formation of a tetrahedral intermediate. When ammonia is removed, the intermediate is converted to an acyl enzyme complex which is then hydrolyzed to an acid upon the addition of a water molecule (Kobayashi *et al.*, 1998b) (Figure 1-11).

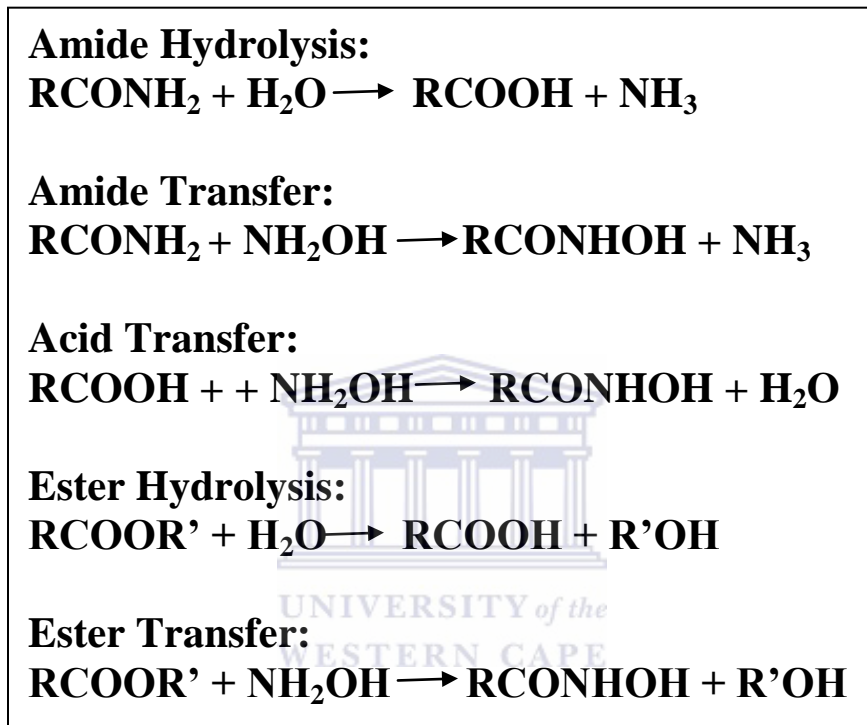


Figure 1-10 Reactions catalyzed by amidases

Figure was reproduced from Fournand *et al.*, (2001).

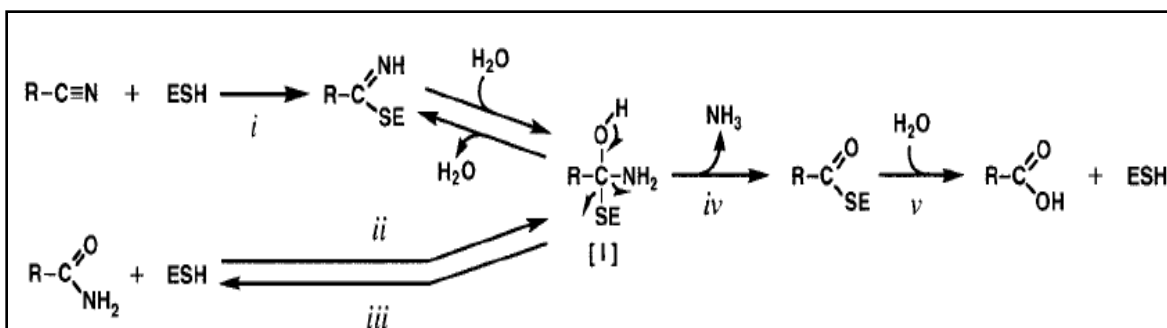


Figure 1-11 Possible catalytic mechanism for the nitrilase reaction.

The active sulfhydryl group is indicated by ESH. The tetrahedral intermediate is indicated by [I]. i: formation of a tetrahedral intermediate via an enzyme-thioimidate, ii: formation of intermediate from nitriles, iii: intermediate could be broken down to produce amide for selective substrates, iv: ammonium is removed from the tetrahedral intermediate to yield an acyl enzyme, v: hydrolysis of acyl enzyme to acid, non-reversible reaction. Figure was taken from (Kobayashi *et al.*, 1998b).

1.4 Applications of Nitrile Metabolising Enzymes

Microorganisms containing nitrile metabolising enzymes have considerable potential as synthetic biocatalysts in chemical industries as well as in environmental bioremediation (Banerjee *et al.*, 2002). Using nitrile metabolising enzymes as biocatalysts may be advantageous compared to the chemical based nitrile conversion method since the chemical hydrolysis of nitriles requires harsh conditions (Banerjee *et al.*, 2002; Fournand & Arnaud, 2001). Chemical hydrolysis of nitriles requires the heating under strong acidic or alkaline pH which may result in low yields. The formation of toxic inorganic cyanides and salts as by-products make this process even less desirable. Biocatalytic conversions are more attractive as the hydrolysis occurs at a mild pH and temperature.

Most nitrile compounds are highly toxic, carcinogenic and mutagenic and can cause harm to environments. Microorganisms with nitrile metabolising enzymes can be used to biodegrade nitriles before they are released into the environment. The most successful NHase biotransformation process is the conversion of acrylonitrile to acrylamide (Cowan *et al.*, 2003) as well as the development of transgenic cotton crops expressing a bacterial bromoxynil-specific Nase, which induces resistance of plants against the herbicides (Stalker *et al.*, 1996). Amidases are widely used with NHases in producing acrylic acid, p-aminobenzoic acid, nicotinic acid and pyrazinoic acid (Banerjee *et al.*, 2002).

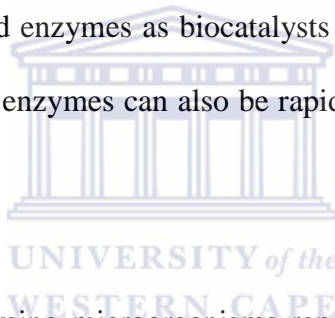
1.5 Cold Adaptation

Microorganisms are largely successful colonisers due to their ability to adapt to a wide range of environmental stress factors. With 80% of the Earth's biosphere below 5°C, a wide range of strategies to adapt to this condition have been evolved by microorganisms in order to maintain cellular functions at cold temperatures. Microbes living in cold terrestrial habitats generally encounter more than one stress factor, including desiccation, high osmotic pressure, high and/or low pH and low nutrient availability (Morgan-Kiss *et al.*, 2006).

Temperature is one of the most important factors controlling biochemical reactions. Hence, enzymes need to be suitably adapted in order to perform catalytic activities optimally. Enzymes from cold adapted organisms typically show high conformational

flexibility, which contributes to their high catalytic activity by allowing better accessibility to substrates to the enzyme's active site (D'Amico *et al.*, 2001). The flexibility of the enzymes is acquired by modifications such as decreased core hydrophobicity, decreased arginine: lysine ratios, increased glycine residues near functional domains, fewer salt bridges, additional surface loops, substitution of proline residues by glycine in surface loops and a decreased number of aromatic residue interactions (D'Amico *et al.*, 2001).

The properties of cold adapted enzymes are attractive for both applied and basic research because the use of cold adapted enzymes as biocatalysts can potentially eliminate costly processes such as heating. The enzymes can also be rapidly deactivated by heating when required (Gerday, 1997).



The majority of nitrile hydrolysing microorganisms reported were either mesophilic or thermophilic organisms. To date no NHases or Nases have been purified from cold adapted microorganisms and only one aliphatic amidase has been reported (Nel, 2009) from the same organism studied in this project.

1.6 Protein Structure Modeling

Comparative or Homology modeling uses experimentally determined protein structures (templates) to predict the structure of other proteins (targets) with similar amino acid sequences (Sánchez & Sali, 1997b). The method aims to build three dimensional (3D) models of unknown protein based on one or more atomic structures (Sánchez & Sali, 1997b). The proteins may have low sequence identities, but as long as they share similar folds and a closely related function, their models may be built due to the fact that 3D structures of related proteins are more conserved than their amino acid sequences. There is a variety of servers and programs that automate the modeling process such as WHATIF (www.cmbi.kun.nl/whatif), SWISS-MODEL (www.Expasy.ch/swissmod/SWISS-MODEL.html) and PDB-Blast (www.Bioinformatics.Burnham-inst.org/pdb_blast). However, a more accurate and useful model is obtained by a manual method where the choice of appropriate templates, construction of the best alignment, specification of constraints and the loop modeling can be controlled by the researcher.

Comparative modeling usually consists of four steps (Figure 1-12). The first step is to identify the 3D structures that have similar amino acid sequences to the target sequence. The second step involves aligning the multiple sequences and choosing the proteins that are best suited to be the templates. The third step is to build a target structure based on the alignment with the template structures. The final step involves the evaluation of the

model. If desired, the entire process can be repeated to produce a model with the satisfactory evaluation results.

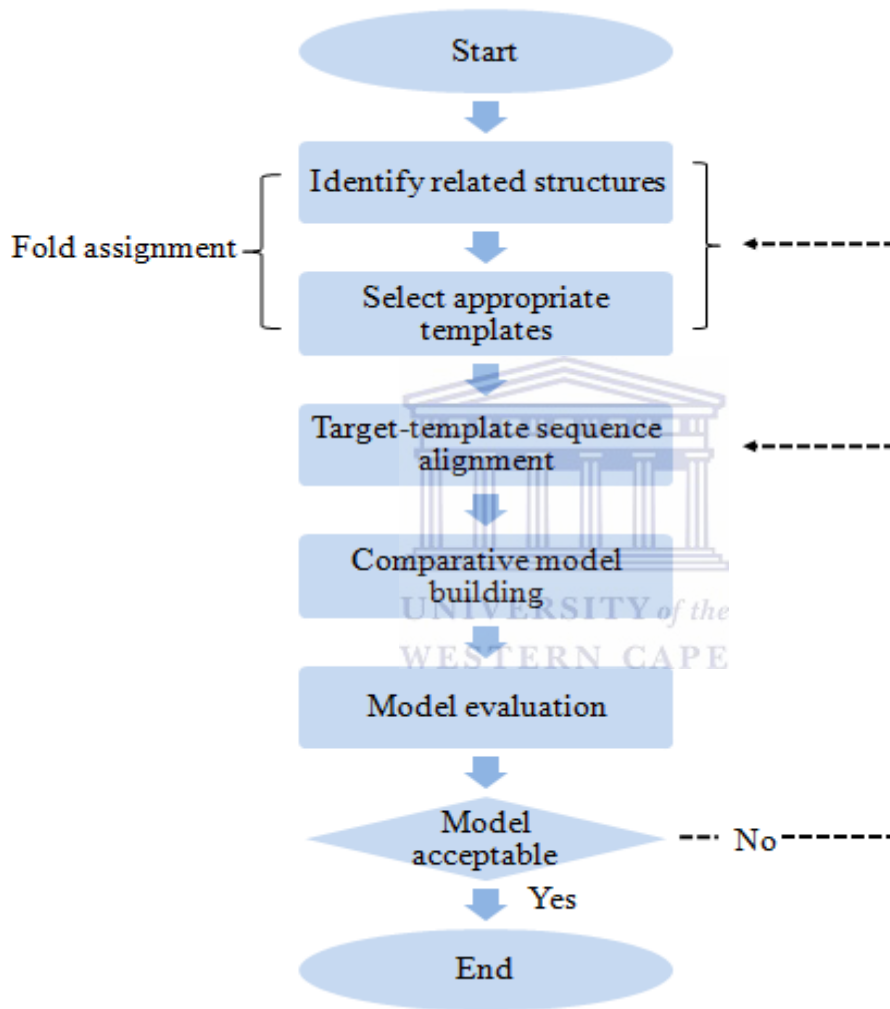


Figure 1-12 Steps in comparative protein structure modeling.

1.6.1 Fold Assignment

Fold assignment involves identifying related structures and selecting the appropriate templates for the target sequence by using Protein Data Bank (<http://www.rcsb.org/pdb>). This search can be done in several ways. The first method involves a pair-wise sequence alignment method which compares the target sequence with each of the sequences in the data base independently. The second method involves a multiple sequence alignment methods which find all the sequences in the databases that are related to the target sequence and aligns them. These methods can provide the insight into which regions are conserved, and into their evolutionary relationship. The third method involves threading the proteins and evaluates the compatibilities between the target sequence and template structures. This method is useful when very small numbers of known sequences are available as templates. Programs such as GenTHREADER (Jones, 1999) and FUGUE (Shi *et al.*, 2001) can be used for threading.

1.6.2 Target-Templates Alignment

Alignment is a fundamental but important step of the entire modeling process. The objective of this step is to establish the best possible sequence alignment between the target and the templates. This can be done in two different ways. Firstly, the alignment can be generated by using automated programs such as CLUSTALX (www.cluster.org), T-COFFEE (www.tcoffee.org) and MULTIALIGN

<http://cbrg.inf.ethz.ch/Server/MultAlign.html>) which produce local alignments according to the best score that correlates both the target and the template sequences. The alignments from the automated programs can be directly used for the modeling process. However the programs are optimised to find remote relationships between homologues but not to produce an optimal alignment. The alignment is usually accurate for the sequences which share more than 40% sequence identity (Fiser, 2004). Otherwise, gaps appear in the alignment, which requires user input to minimize misalignment. The second method generates more accurate alignments than those generated from automated programs. Initially the alignment can be generated using one of the programs specified. The alignment can then be significantly improved by the removal of the gaps in the secondary structure elements. Programs such as GenTHREADER (Jones, 1999) and FUGUE (Shi *et al.*, 2001) are designed to predict secondary structures of the target sequence. This is important because misalignment by only one residue position results in an error of approximately 4Å in the model (Sánchez & Sali, 1997a). This is because the comparative modeling programs cannot detect or recover from the errors in the alignment. Hence, the correct alignment of the target and the template sequences is by far the most important step of homology modeling.

1.6.3 Model Building and Side Chain Optimization

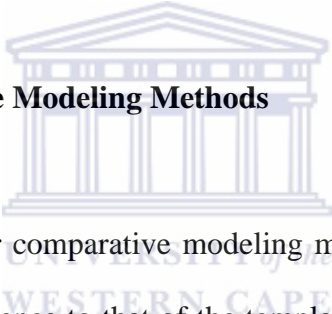
Model building can be automated or performed manually. When it is done with automated programs, only the target sequence needs to be provided as the fold assignment, target-template alignment and the model building steps are automated. Once the model is built, side chain optimisation can be done to improve the model. When the model is built manually using a program such as MODELLER (Fiser *et al.*, 2003), the alignment file obtained from the previous step needs to be converted into PIR format for subsequent use by the program. If the target and the templates sequences have more than one chain, each chain can be kept separate using chain breakers (“/”) in the sequence alignment file (Fiser *et al.*, 2003). The number of cycles of model building can be chosen by the person building the model. During the cycle, the log file generated in each cycle is continuously scanned for errors. This process can be iterated until an acceptable model is obtained. The number of models produced can also be adjusted by the user. The model that has the lowest MODELLER objective function score is an indication of how well that particular model satisfies the restraints used to calculate it (Fiser *et al.*, 2003). However, this value does not necessarily indicate the quality of the model. The models are generated as PDB files which can be viewed with the programs such as PyMOL and Chimera (Eric *et al.*, 2004). As the MODELLER focuses more on building a backbone of the model, it is advisable to optimize the side chains. This can be done using automated programs such as SCWRL (Bower *et al.*, 1997). The side chains with the most favorable

energy status would be chosen and the side chains of the models are adjusted accordingly (Bower *et al.*, 1997).

1.6.4 Model Evaluation

The accuracy of the model can be evaluated using MolProbity, PROCHECK and RAMPAGE which generates Ramachandran plot to indicate the outlier residues. The final model is structurally aligned with templates to assess the root mean square deviation of the atom coordinates between the model and the templates.

1.7 Reliability of Comparative Modeling Methods



The programs that are used for comparative modeling methods are designed to arrange the backbone of the target sequence to that of the template by using sequence alignment to determine the position of each corresponding residue. It is not a trivial process, as there may be many residue substitutions, insertions and deletions among the aligned sequences (Fiser *et al.*, 2003). This suggests that the comparative modeling method can be unreliable when the side chain conformation predictions and/or structure prediction of sections of target sequence do not align well with the template sequences.

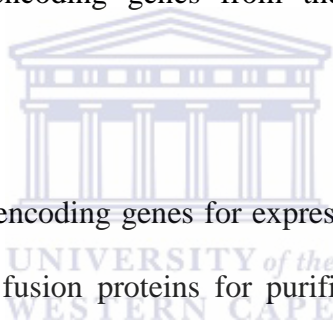
To assure the quality of the model, there are two things that need consideration. Firstly, the alignment between the target and the template sequences should be used to guide the modeling process. Errors in the sequence alignment will result in errors in spatial arrangement of the residues in the model (Fiser, 2004). Secondly, the sequence identity as well as the structural resemblance between the target and the templates sequences also contributes to the quality of the model (Marti-Renom *et al.*, 2002). A model is considered “acceptable” when the root mean square deviation (RMSD) is within the spread of deviations observed for structures used in the alignment (Marti-Renom *et al.*, 2002). The RMSD gives the overall differences in the protein backbone structure by assessing the root mean standard deviations of the positions of the main chain atoms. It was reported that with 50% sequence identity between the target and the templates, the RMS error for the main chain atoms would be about a 1Å (Baker & Sali, 2001). When the sequence identity is between 50% and 30%, the RMS error would be about 1.5Å (Baker & Sali, 2001). The sequence alignment would contain more errors as the sequence identity lowers, hence the higher the RMS value. The structural differences between the models and the solved structures may arise due to differences in molecular environment and the data collection methods used to determine crystal structures. For instance, X-ray crystal structures show differences in data collection, data refinement, experimental conditions and crystal packing.

All protein structure prediction methods are known to have some errors. Therefore the verification of the quality of the model is a pre-requisite for the good results.

1.8 Aim

Although many amidases have been characterised, it is evident that no cold-adapted amidases have been isolated. The aim of this research was to identify and isolate the cold adapted amidases. The specific objectives were:

- 1) To identify amidase encoding genes from the whole genome sequence of *Nesterenkonia* sp.
- 2) To isolate the amidase encoding genes for expression in an *E.coli* system for the production of His-Tag fusion proteins for purification and subsequent enzyme characterisation.



Chapter 2

General Materials and Methods

2.1 Chemicals and Reagents

Unless otherwise stated, Merck Chemicals and Laboratory supplies, Sigma Aldrich Chemical Company and Kimix Chemical and Laboratory supplies supplied all chemicals of analytical/reagent grade.

DNA size markers, protein size markers and all DNA modifying enzymes (polymerases and restriction endonucleases) were purchased from Fermentas Life Sciences Ltd.

Oligonucleotide primers for polymerase chain reaction (PCR) were synthesized by Inqaba Biotech or Integrated DNA Technologies (IDT) Inc.

2.2 Bacterial Strains and Plasmids

Table 2-1 Bacterial Strains Used in This Study

Bacterial Strains	Relevant Genotype	Supplier
<i>E. coli</i> Gene Hog	F- mcrA Δ (mrr-hsdRMS-mcrBC) ϕ 80lacZ Δ M15 Δ lacX74 recA1 araD139 Δ (ara-leu)7697 galU galK rpsL (StrR) endA1 nupG	Invitrogen (USA)
<i>E. coli</i> BL21(DE3)	HsdS gal ompT (λ cIts857 ind 1 Sam7 nin5 lacUV-T7 gene1)	Stratagen (USA)
<i>E. coli</i> BL21(DE3)pLysS	F-, ompT, hsdSB (rB-, mB-), dcm, gal, -(DE3), pLysS, Cmr	Invitrogen (USA)
<i>E. coli</i> Rosetta2(DE3)pLysS	F- ompT hsdSB(rB- mB-) gal dcm (DE3)pLysSRARE (CamR)	Novagen (USA)

UNIVERSITY of the
WESTERN CAPE

Table 2-2 Plasmids Used in This Study

Plasmids	Characteristics	Supplier
pET 17a+	Expression vector containing an ampicillin resistance gene	Novagen (USA)
pET 28a+	Expression vector containing a kanamycin resistance gene	Novagen (USA)
pACYC 177	Cloning vector containing both ampicillin and kanamycin resistance gene	Fermentas (USA)

2.3 Media

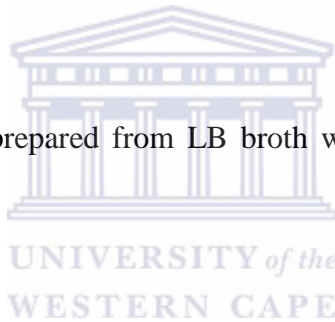
Oxoid Ltd and Biolabs supplied culture media. All media was autoclaved at 121°C for 20 minutes unless otherwise stated.

Luria-Bertani (LB) broth

These media were routinely used to grow bacterial strains.

Constituent	L ⁻¹
Tryptone	10g
Yeast Extract	5g
NaCl	10g

Luria-Bertani (LB) agar was prepared from LB broth with the addition of 1.3% [w/v] bacteriological agar.



2x TY media

2x TY media was used to prepare competent cells

Constituent	L ⁻¹
Tryptone	16g
Yeast Extract	10g
NaCl	5g

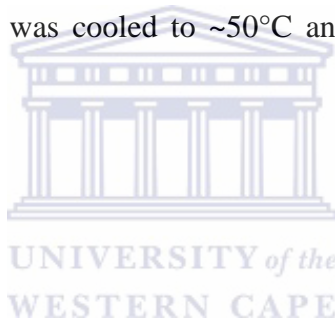
SOC media

SOC media was used for the recovery of newly transformed *E.coli* strains.

Constituent	L ⁻¹
Tryptone	20g
Yeast Extract	5g
NaCl	0.5g
KCl (250mM)	10ml

After autoclaving the medium was cooled to ~50°C and the following filter sterilized solutions added aseptically.

MgCl ₂ (2M)	5ml
Glucose (1M)	20ml



Modified Castenholtz media for the growth of *Nesterenkonia sp.*

Constituent	L ⁻¹
Castenholz salt	500ml
YE/Glycerol	500ml

Nitsch solution per L (not autoclaved)

Constituent	L ⁻¹
H ₂ SO ₄	0.5ml
MnSO ₄ H ₂ O	2.46g
ZnSO ₄ 7H ₂ O	0.89g
H ₃ BO ₄	0.5g
CuSO ₄ 5H ₂ O	0.025g
Na ₂ MoO ₄ 2H ₂ O	0.03g
CoCl ₂ 6H ₂ O	0.1g

Castenholz salts (Autoclave)

Constituent	L ⁻¹
FeCl ₃ (0.03% solution)	2ml
CaSO ₄ 2H ₂ O	0.12g
MgSO ₄ 7H ₂ O	0.2g
KCl	0.016g
Nitsch Solution	1ml
Na ₂ CO ₃	0.5g
1M KH ₂ PO ₄	pH to 9.6



Carbon and Nitrogen Source (Autoclave)

Constituent	L ⁻¹
Glycerol	2%
Yeast Extract	0.1%

After the media were autoclaved and cooled to ~50°C, the appropriate filter sterilized antibiotic was added aseptically. Final concentrations of antibiotics were: (unless otherwise stated) chloramphenicol (cam) 34 µg/ml; kanamycin (kan), 50 µg/ml; ampicillin (amp) 100 µg/ml.



2.4 DNA extractions

Genomic DNA extraction was adapted from Zhou *et al.* (1996). A 50ml cultures were centrifuged and the pellets resuspended in 10ml of extraction buffer (25mM Tris-HCl pH 8.0, 50mM Glucose, 10mM EDTA, 25mg lysozyme/ml, 50µg RNase A/ml). The samples were incubated for 30mins at 37°C. 15µl per ml of 20% SDS was added to the samples and incubated at 65°C for 30mins. 100µg per ml of proteinase K was added to the sample and incubated at 37°C for 30mins. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was mixed with the supernatants, then centrifuged at 16 000 × g and an equal volume of chloroform was mixed with the upper aqueous phase and the mixture was centrifuged. Chloroform extractions were repeated with subsequent centrifugation

until the extraction was clear. The DNA was finally precipitated with 0.6 vol of isopropanol by overnight incubation at 25°C. The pellets were collected by centrifugation at $16\ 000 \times g$ for 30 min and washed twice with 70% ethanol, dried and resuspended in 50 μL of sterile H_2O . The concentration of gDNA was estimated using the Nanodrop 1000 (Thermo Scientific) or the Qubit™ Quantitation System (Invitrogen) and the purity of DNA was evaluated by separation on a 0.7% agarose gel containing 0.5 $\mu\text{g}/\text{mL}$ of ethidium bromide and visualised using the AlphaImager HP (Alpha Innotech) gel imaging system.

2.5 DNA Purification

The Illustra™ GFX™ PCR DNA and gel band purification kit (GE Healthcare Limited, Buckinghamshire, UK) was used to purify DNA from solutions and agarose gels.



2.6 Amplification of Amidase Genes

In order to amplify target DNA, a GeneAmp PCRsystem 2700 (Applied Biosystem) or Eppendorf Mastercycler gradient thermocycler was used. Unless otherwise stated the Finnzyme's Phusion kit was used. A 20 μl PCR reaction contained the following reagents: 1x GC buffer, DNA template (20ng plasmid or 50ng chromosomal DNA), 0.5 μM of the forward and reverse primers each, 200 μM of dNTPs (dATPs, dCTPs, dGTPs and dTTPs)

and 1U of Phusion polymerase (Table 2-3). Reaction mixture was made up to 20 μ l with sterile ddH₂O.

2.7 Preparation of Electro-Competent Cells

Electrocompetent Genehog and BL21 (DE3) pLysS *E. coli* cells were prepared as outlined in Sambrook and Russell (2001) with slight modification. All glassware was thoroughly acid-washed with 70% EtOH, rinsed and autoclaved prior to use. A single colony of the *E. coli* strain was inoculated into 5ml of LB-broth and incubated at 37°C with shaking overnight. An overnight culture of 2.5ml was transferred to 250ml of LB-broth and incubated at 37°C until mid-logarithmic phase (OD₆₀₀ of 0.4). The flask was rapidly cooled on ice for 20mins and the cells were collected in polypropylene tubes by centrifugation at 3000 x g for 10mins in an Eppendorf 5810R fixed rotor centrifuge. The supernatant was decanted and the cells were resuspended in equal volume of 10% glycerol. The cells were harvested as above. The glycerol wash step was repeated 6 times. The harvested pellet was resuspended in 1ml of 15% glycerol, 2% sorbitol. The cells were aliquoted into 40 μ l volumes and stored at -80°C.

Table 2-3 Primers used and PCR condition

Genes Amplified	Primers used	PCR conditions	Reference
Ami F	Ami 2F: TTTGGGCATATGAGCCTGCCAACCGA Ami 2R: TGCGAATTCTCACAGCCCTGCGAAGTCT	98°C for 30s 25 x (98°C for 10s, 55°C for 10s, 72°C for 20s) 72°C for 10mins	This Study
Ami S	Ami F: TTTGGGCATATGGACCGGGTCACGCAGCAAT Ami R: TGCGAATTCTCTAGGTACTCTGACTGACTGAG CTC	98°C for 30s 25 x (98°C for 10s, 55°C for 10s, 72°C for 30s) 72°C for 10mins	This Study
Bacterial 16S rRNA	E9f: GAGTTTGATCCTGGCTCAG U1510R: GGTTACCTTGTTACGACTT	95°C for 3mins 25 x (95°C for 30s, 50°C for 30s, 72°C for 1min) 72°C for 10mins	(Hansen <i>et al.</i> , 1998) (Reysenbach & Pace, 1995)
RT	RTF: ACATCGCGGTGCGCAACTGT RTR: GAGAGCGCGAGCTCATCGGT	98°C for 10s 2 x (98°C for 10s, 65°C-55°C for 10s, 72°C for 10s) 72°C for 10mins	This Study

2.8 Preparation of Chemically Competent Cells

Chemically competent Genehog, BL21 (DE3) and Rosetta2 (DE3) pLysS *E.coli* cells were prepared according to the method in Sambrook and Russell (2001) with slight modification. All glassware was thoroughly washed with 70% EtOH, rinsed and autoclaved prior to use. A single colony of *E. coli* stain was inoculated into 5ml LB-broth and incubated overnight at 37°C with shaking. 2.5ml of overnight culture was aseptically transferred into 250ml of LB broth and incubated at 37°C until mid logarithmic phase (OD₆₀₀ of 0.4). The flask was rapidly cooled on ice for 20mins. The cells were harvested in polypropylene tubes by centrifugation at 3000 x g for 10mins in an Eppendorf 5810R fixed rotor centrifuge. The supernatant was decanted and the cells were resuspended in 30ml of 0.1M CaCl₂ and incubated on ice for 20mins. The cells were harvested as above and the pellet was resuspended in 2ml of 0.1M CaCl₂ and 15% glycerol. The cells were aliquotted into 100µl volumes and stored at -80°C.

2.9 Cloning

A 10µl reaction mixture was prepared with appropriate concentrations of vector and insert at a 1:3 molar ratio, reaction buffer at 1X final concentration, 10U of T4 DNA ligase and water to final volume of 10µl. The reaction mixture was incubated overnight at

4°C. Both Ami F and Ami S were cloned into pET 17 b+ and pET 28 a+ using the same multiple cloning sites, *NdeI* and *EcoRI*.

2.10 Electroporation

A 40µl volume of electrocompetent cells in a 1.5ml tube was removed from -80°C storage and allowed to thaw on ice. 1µl of ligation mix was added to the thawed cells. The mixture was transferred to pre-chilled 0.1 cm sterile electroporation cuvette (Bio-Rad Laboratories, Hercules, CA, USA). The following conditions were used to perform the electroporation: 1.8kV, 25µF, 200Ω. Immediately after the electroporation, 1ml of SOC broth was added to the cuvette. The cells were transferred to 2ml tubes and incubated at 37°C for 1hr with agitation. A 100µl volume of cells were plated out onto LB agar plates containing appropriate antibiotic(s), and incubated overnight at 37°C.

2.11 Heat Shock

A 1.5ml tube containing 100µl of chemically competent cells was removed from -80°C and allowed to thaw on ice. 5µl of the ligation mix was added to the thawed cells and incubated on ice for 20mins. The cells were heat shocked at 42°C for 90 s in water bath, then immediately returned to ice with addition of 1ml of SOC broth. The cells were then

incubated at 37°C for 1hr with agitation. A volume of 100µl of cells were plated out onto LB agar plates containing with appropriate antibiotic(s) and incubated overnight at 37°C.

2.12 Minipreps

2.12.1 Alkaline Lysis Method

Single colonies were picked from the agar plates and inoculated into 5ml LB broth supplemented with the appropriate antibiotic(s). The LB was incubated overnight at 37°C with agitation. Plasmid DNA was isolated from the overnight cultures by alkaline lysis method (Sambrook & Ressel, 2001), with slight modifications. 2ml of overnight culture was transferred into 2ml tubes and the cells were harvested by centrifugation at 10000 x g for 1min at room temperature. The supernatant was discarded and the pellet was resuspended in 200µl of solution 1 (50mM glucose, 25mM Tris-HCl pH 8.0 and 10mM EDTA pH 8.0). 200µl of solution 2 (1% [w/v] SDS and 0.2M NaOH) was added to the mixture and the tube was mixed well by inversion and incubated for 5mins on ice. Following the addition of 300µl of 7.5 M ammonium acetate (pH 7.6), the tubes were incubated on ice for 10mins then centrifuged at 13,000 x g for 15mins at room temperature. The supernatant was transferred to new tube, 500µl of chloroform/isoamylalcohol (24:1) was added and the samples centrifuged at 13,000 x g for 10 mins at 4°C. The supernatant was transferred to new tubes and the plasmid DNA

was precipitated by an adding equal volume of isopropanol. The tubes were incubated at -80°C for 15mins and centrifuged at 13,000 x g for 10mins at 4°C. The pellet was dried and redissolved in TE containing RNase A to the final concentration of 20µg/ml.

2.12.2 Miniprep Kits

Plasmid DNA extractions performed for subsequent nucleotide sequence analysis was performed using the Zymo miniprep kit (Zymo research, USA).

2.13 Screening

2.13.1 Restriction Fragment Analysis



All restriction enzyme digestions were performed in sterile 1.5ml tubes in 15µl reactions unless otherwise stated. The reaction mixture contained 1x buffer supplied by the manufacturer for the specific enzyme and 5-10U of enzyme per µg of plasmid DNA. The reactions were incubated 2-24 hrs in a water bath at 37°C unless stated otherwise. The digestion products were analyzed by gel electrophoresis in 0.8% or 1% agarose gel as described in section 2.15.1.

2.14 Sequencing

Sequencing of cloned insert DNA was performed using a MegaBACE 500 Automated Capillary DNA sequencing System (Amersham Biosciences) by Stellenbosch University.

2.15 Electrophoresis

2.15.1 Agarose Gel Electrophoresis

Analysis of DNA was performed using agarose gel electrophoresis (Sambrook & Ressel, 2001). Horizontal 0.7% - 1% (w/v) TAE agarose gels were casted and electrophoresed at 100V in 0.5 TAE buffer (40mM Tris-HCl, 1mM EDTA, 10mM glacial acetic acid, pH 8.5). To allow visualization of the DNA on a UV transilluminator, the gels were supplemented with $0.5\mu\text{g}\cdot\text{ml}^{-1}$ ethidium bromide. The DNA fragments were sized according to their migration in gels in comparison to standard DNA molecular markers (100bp ladder and 1kb ladder, Fermentas, USA).

2.15.2 Denaturing SDS-polyacrylamide Gel Electrophoresis

Denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to the method of Laemmli (1970) using a Mighty small TM

SE 280 vertical slab unit (Hoefer Inc, USA) with 1mm gels containing 10% acrylamide. Protein samples were prepared by suspending in equal volumes of SDS-PAGE loading buffer (0.5M Tris-HCl pH 6.8, 10% SDS, 50% Glycerol, 0.05% Bromophenol blue and 0.05% β -mercaptoethanol) and were denatured by boiling for 10mins. The samples were loaded onto the gel and electrophoresed with running buffer (0.02M Tris-HCl pH 6.8, 0.192M glycine and 0.1% SDS) at room temperature under constant voltage of 80V until the dye migrated into the separating gel, followed electrophoresis at 120V. Pre-stained SDS-PAGE page marker or unstained page molecular marker from Fermentas were used.

Following electrophoresis, SDS-PAGE gels were stained with Coomassie Brilliant Blue staining solution. The gels were heated for 30 sec in a microwave then incubated at room temperature with agitation for 2hrs. The Coomassie solution was then discarded, the gel was rinsed with sterile water followed by soaking in destaining solution (50% methanol (v/v), 40% dH₂O (v/v), 10% acetic acid (v/v)) for overnight with agitation.

2.15.3 Native Polyacrylamide Gel Electrophoresis

Native polyacrylamide gel electrophoresis was carried out using a Mighty small TM SE 280 vertical slab unit (Hoefer Inc, USA). The native PAGE gel was made using the same method in section 2.14.2 but the SDS was left out. Protein samples were prepared by suspending in equal volumes of loading buffer (50% glycerol, 25% 0.5M Tris-HCl pH 6.8 and trace of bromo phenol blue) and kept cold on ice. The samples were loaded onto

the gel and electrophoresed under constant voltage of 80V until the dye migrated into the separating gel, followed electrophoresis at 120V. The running buffer used contained 0.05M Trizma (w/v) and 0.38M Glycine. The pH of the running buffer was adjusted to 8.9.

Following electrophoresis, Native gels were stained as it was described in section 2.15.2.



Chapter 3

Sequence Analysis and Homology Modeling

3.1 Introduction

Protein sequence analysis can be used to identify distantly related proteins and their possible functions. The structure of a protein also provides insights into the location of possible active sites and regions of the protein involved in maintaining the structure. Solving the structures of the proteins experimentally is a complex and time-consuming process. Hence, the importance of the homology modeling is that it is a less time consuming and more convenient way to predict the protein structure. In addition, with the increasing use of whole genome sequencing, solving the structures of all the proteins identified experimentally are not feasible. However, it is widely accepted that homology models are not as reliable as those determined experimentally.

This chapter describes the modeling processes and the model structures of AmiF and AmiS.

3.2 Sequence Analysis

The genome of *Nesterenkonia* sp. was sequenced using the Illumina DNA sequencing technology at the University of the Western Cape for the previous studies done on the same organism (Nel, 2009). This study also revealed that the *Nesterenkonia* sp. genome sequence showed high percentage sequence identity with *Arthrobacter* species genomes. Hence, *Arthrobacter* amidase sequences were used to search, by BLAST algorithm, semi-aligned contigs generated from the sequenced *Nesterenkonia* genome. When the AmiS ORF was searched against NCBI database, the AmiS sequence showed the highest similarity towards an amidase from *Halothiobacillus* and an amidotransferase subunit A (*gatA*) from *Pseudoalteromonas* (73% and 72% respectively). Amidotransferases comprise of three subunits (*gatA*, *gatB* and *gatC*). In an operon, these genes are positioned in a sequence of *gatC*, *gatA* and *gatB* (Curnow *et al.*, 1997). Although GatA has the highest sequence identity towards the amidase family, GatA does not express, unless it is co-expressed with GatC (Curnow *et al.*, 1997). The upstream sequence of the AmiS was investigated to confirm the identity of AmiS as a *gatA* as well as to identify the *gatC* sequence. However, the upstream sequence showed 50% identity to an aceta/formamidase from *Mycobacterium* sp. as well as 49% towards an aceta/formamidase from *Geobacillus* sp. This aceta/formamidase was named AmiF.

The full length 1140bp of *amiF* gene encoded 380 amino acids (Figure 3-2) and the full length 1365 bp of *amiS* gene encoded 455 amino acids (Figure 3-4). The results of

reverse transcription PCR (section 4.3.4) showed that *amiF* and *amiS* are co-transcribed. Further sequence analysis resulted in identification of another protein about 500 bp downstream from *AmiS*. The BLAST result indicated this gene was a probable amidotransferase class V, belonging to the amidohydrolase family. This cluster of amidases could possibly indicate the presence of an amidase operon (Figure 3-1).

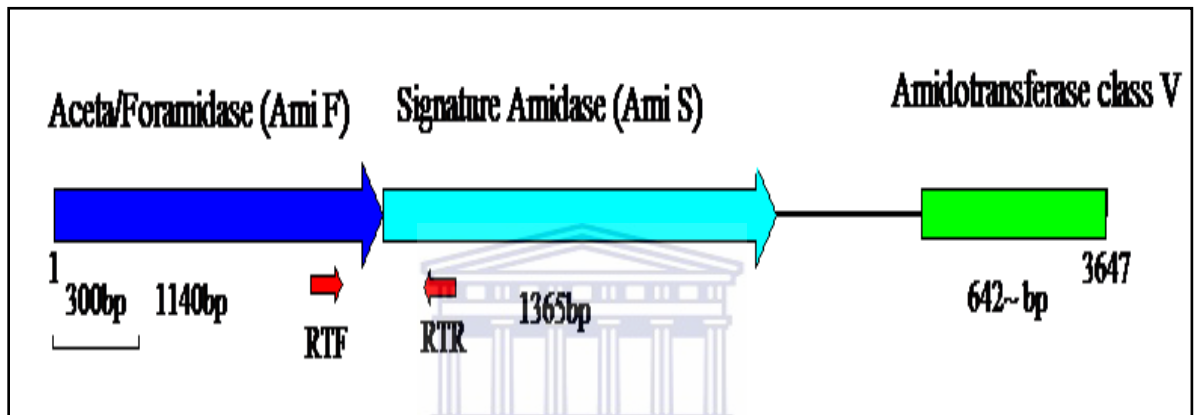


Figure 3-1 Genetic map of two amidases and neighboring ORFs identified on *Nesterenkonia* sp.

RTF: Forward primer for the reverse transcription PCR, RTR: Reverse primer for the reverse transcription PCR, Red arrows indicate the approximate binding sites for the primers. Reverse transcription PCR is described in Chapter 4. 300bp with underline indicates the scale of the map.

Initially, *AmiF* was assigned to nitrilase superfamily due to the fact that it lacked the GGSS signature sequence, and that BLAST results indicated that the protein was aceta/formamidase belonging to branch 2 (aliphatic amidases) of the nitrilase superfamily. However, further sequence analysis showed the absence of the novel catalytic triad EKC which all the members of the nitrilase superfamily share (Brenner, 2002). Although many studies categorize the amidases into two groups (signature

amidases and nitrilase superfamily), AmiF clearly did not belong to either of these groups. Pfam server (pfam.sanger.ac.uk) identified AmiF as an aceta/foramidase family member which is a member of amidohyrolase family. The Pfam server listed two proteins in this family, an acetamidase from *Mycobacterium smegmatis* and a formamidase from *Methylophilus methylotrophus*. Although their substrates are known, these amidases are not well characterised and their structures nor their catalytic residues are known. When AmiF sequences were aligned with the acetamidase and formamidase sequences, they showed few conserved residues (Figure 3-3), further studies such as site directed mutagenesis would be required for the identification of the catalytic residues.

For AmiS, both the BLAST results and the alignments with proteins of the signature amidase family clearly showed that it belongs to this group. AmiS also showed a conserved catalytic triad Ser131-Ser155-Lys54 (the numbering is according to AmiS sequence) (Figure 3-4).

Open Reading Frame of AmiF

```

1      ATGAGCCTGCCAACCGACCAGCAGTCCAGCACCCGCATCAGCGTGGAGTCCAGCCCCGAG
1      M S L P T D Q Q S S T R I S V E S S P E

61     CCCGTCCTGCAGCCCGGCGAGCACCTGCCCGAGGGGGCCCGTTACCTCCCGGCCACCGCG
21     P V L Q P G E H L P E G A R Y L P A T A

121    GAGAACATCTTCTGGGGGCGCCTGCCGTGCGCCAGTGACGTTCCGGTGCTGCACATCCAG
41     E N I F W G R L P C A S D V P V L H I Q

181    CCCGGGGAGCAGGTCGTTCATCGACACCCTCTCGCATGAAGGCATCCTGGAGGATCAGGGC
61     P G E Q V V I D T L S H E G I L E D Q G

```

241 CGGGACCCCGCCGGATACTTCGCCGGCCACGGCGTCCCGGCCGAGCAGGTGCTCAACGAC
 81 R D P A G Y F A G H G V P A E Q V L N D

301 GCGATCGAGCTGGCCGCCAGCGATCACGCCCGGGACCCGAAGCTCGATGGACCGCATGTG
 101 A I E L A A S D H A R D P K L D G P H V

361 AACACCGGCCCCGATCCATGTGGCCGGCGCGCAGCCCGGAGACCTGCTGAAGGTCACCATC
 121 N T G P I H V A G A Q P G D L L K V T I

421 GACAAGCTTGAACGCCGGGTGCCCTACGGAATCATCTCCAACCGGCACGGCAAGGGCGCG
 141 D K L E R R V P Y G I I S N R H G K G A

481 CTGAACGGGGAGTACCCGCGCGGGGAGGCCAACGTGTCGATCTTCGCCGCCGTGGAGCCC
 161 L N G E Y P R G E A N V S I F A A V E P

541 GACGCCGAGGGCGTCCCGCGCGGCTGGATGGCCCCGAAGGCAGGGCGCCCCGCGACGGCTG
 181 D A E G V P R G W M A P K A G A P R R L

601 AGCTTCGACCTCGCCCCGTTCTTCGGGACCATCGGGGTGGCCGTGGCCGGCGAGGTCCGG
 201 S F D L A P F F G T I G V A V A G E V R

661 CCGCACTCGGTGCCCCCGGGCGCCACGGCGGGAACATCGACATCAAGGTGCTCACCGAA
 221 P H S V P P G A H G G N I D I K V L T E

721 GGAACCTCGCTCTACCTCCCGGTCCAGGTGCCCGGGCGCGCTGGCCTTCGTCGGGGACCCA
 241 G T S L Y L P V Q V P G A L A F V G D P

781 CACTTCGCCCAGGGCGACGGCGAGGTTGCGCTCACCGCCATGGAAGCCTCGCTGCGCGGA
 261 H F A Q G D G E V A L T A M E A S L R G

841 CACCTGAGCTTCGAGGTGGTTCCCGTGGAGGCGCTGCGAGCCTTCGGCGAACAGATC
 281 H L S F E V V P R E E A L R A F G E Q I

901 GGCCCCGCTCGCCGAGACCACCGATCACCTGATCCCCACCGGGATGGATGAGGACCTCGAC
 301 G P L A E T T D H L I P T G M D E D L D

961 ATCGCGGTGCGCAACTGTGTGCACAACGCGATCGTGCTGCTTCAGGCCCGCTGGGGGCTC
 321 I A V R N C V H N A I V L L Q A R W G L

1021 GACGCCGAACACGCCTACGCGTATCTCTCCGCGGCCACCGACTTCAACATCTCCCAGGTG
 341 D A E H A Y A Y L S A A T D F N I S Q V

1081 GTCGACATCGTCAAGGGCGTCCATGCGAAGGTCTCCGTCAAAGACTTCGCAGGGCTGTGA
 361 V D I V K G V H A K V S V K D F A G L *

Figure 3-2 Nucleotide and deduced amino acid sequences of AmiF

Total amino acid number: 379, MW=40138, Max ORF: 1-1137, 379 AA, MW=40138

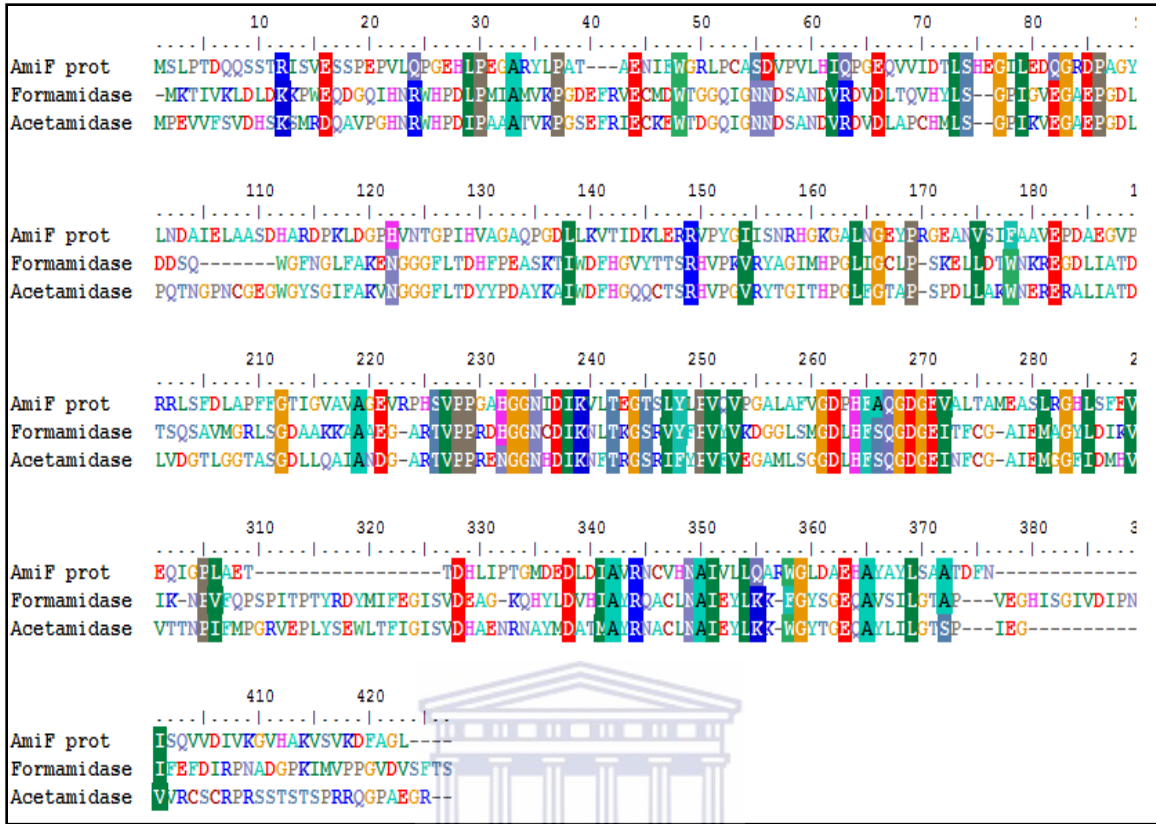


Figure 3-3 Alignment of AmiF, Formamidase from *Methylophilus methylotrophus* and Acetamidase from *Mycobacterium smegmatis* showing conserved sequences.

Open Reading Frame of AmiS

```

1           ATGGACCGGGTCACGCAGCAATCACCCGCGATGCCCCCGTCCCAGGCCGCGGCAGAGCAC
1           M D R V T Q Q S P A M P P S Q A A A E H

61          CCGATCGACGCGACGGTCTGGCGGGTCGGCGGGGACGCCGGCGGCCACTGGTGCCCCGCC
21          P I D A T V W R V G G D A G A P L V P A

121         ACCGGAACGGGGAGTCTGGACGGGCACACCCCTCGCGGTGAAGGACCTCTACGCCGTGGCA
41          T G T G S L D G H T L A V K D L Y A V A

181         GGGCATCGGATCGGCGCCGGGAACCCGAGCTGGCTCGCCGACGCGCCGGTTC AAGACACC
61          G H R I G A G N P S W L A D A P V Q D T

241         ACCGCCACCGGATCACCCAGCTGCTCGAGGCCGGGGCAGCCATCGCCGGCATCGCGCAG
81          T A T A I T Q L L E A G A A I A G I A Q

```


301 ACCGATGAGCTCGCGCTCTCCCTGGCCGGAACCAACCAGCACTACGGCACCCCGCCCAAT
101 T D E L A L S L A G T N Q H Y G T P P N
361 CCAGCCGCACCGCAGCGGATCAGCGGGGGTCCAGCTCCGGCTCCGCATCCGCCGCGGCG
121 P A A P Q R I S G G S S S G S A S A A A
421 CTGGGCCAGGCGAGCATCGGACTGGGCAGCGACACCGGGCTCCATCCGGGTGCCCGCC
141 L G Q A S I G L G S D T G G S I R V P A
481 TCCTACCAGGGGCTCTGGGGCATCCGCAGCACCCACGATGCCATCGCCCTGGACGGAGTG
161 S Y Q G L W G I R S T H D A I A L D G V
541 CAGCCGCTGGCGCAGAGCTTGCACACCATCGGGTGGATGACCCGCGACGCGGACACCCTC
181 Q P L A Q S F D T I G W M T R D A D T L
601 GCCCGCTCGCACAGGTGCTGCTGCGCCACCCCGCCACTTCTCCAGGTGATCCGGTCCTC
201 A A V A Q V L L R H P A T S P G D P V L
661 AACTCCGAGCCGGCGCTGAAGATCATCCCACCCTTTTCGACCAGGTGCAGCCGGAGGTC
221 N S E P A L K I I P T L F D Q V Q P E V
721 ACCGAGGCCGTGCGCGCGGCACTGACCGCCAGTCCGGCTCCGGAGCTGACGGCAGTGCC
241 T E A V R A A L T A Q S G S G A D G S A
781 GAACGCTCGCCGCAGAGCCCCGGGAGTCCGATCGAAGAGATCGAAGGCGTCACCCCGGAG
261 E R S P Q S P G S P I E E I E G V T P E
841 ATGCACGCGGCCTGGTTCGAGGCGTTCGGGATCATTTCAGTTCGCGGAGGCCTGGGCCAAC
281 M H A A W F E A F R I I Q F R E A W A N
901 CACGGAGAATGGATCGCCACCCTGGGAGACGATGGCCTCCGACGTGCGCGCCAGGTTCC
301 H G E W I A H H W E T M A S D V G A R F
961 CGGATGGCCAGCGAGCTCACGGCAGACCAGGAGATGCAGGCCCGGGCGCTGGCCGCCGGA
321 R M A S E L T A D Q E M Q A R A L A A G
1021 GCCCGCCGGATCATCCGCCGCTGGGTGCGGACGAGATCCTCGCCGTGCCCTCCGCCGCC
341 A R R I I R R W V G D E I L A V P S A A
1081 GGGCCGGCACCGCTTCGTGCGGATGCCGCCTTGGGCGGCACCGTGATCGAAGAGCATCGC
361 G P A P L R A D A A L G G T V I E E H R
1141 CGCAGCACCATGCTGCTGACCTGCCTTGCAGGCCTTGCCGGTCTTCCCGCGGTCAACGTC
381 R S T M L L T C L A G L A G L P A V N V
1201 CCGCTGCGCACCCGCGACCACCTGCCACCAGGCGTCTGCCTGATCGGCCCGCTGGCAGC
401 P L R T R D H L P T G V C L I G P A G S
1261 GATATCGCGCTGATCCATCGGGCCGCCCTTCCATGCCAGAATCCCGGCCACCAGAGC
421 D I A L I H R A A A F H A R I P A H Q S
1321 TATGGGTCCGGAGTCACGGAAACGGAGATCCATGAGACTGACTGA

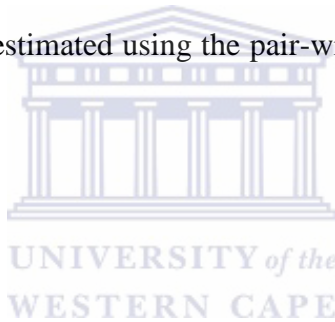
Figure 3-4 Nucleotide and deduced amino acid sequences of AmiS

Total amino acid number: 454, MW=46930, Max ORF: 1-1362, 454 AA, MW=46930

Invariant GGSS motif sequence and catalytic triad sequences are underlined and in bold.

3.3 Phylogenetic Analysis

The phylogenetic trees for AmiF and AmiS were each constructed with 10 sequences, which have the highest sequence identity and belong to the same group of amidases. AmiF and AmiS were each used as the outlier for the alternative tree. The evolutionary distances of the proteins were estimated using the pair-wise sequence comparison tool in MEGA4 (Tamura *et al.*, 2007).



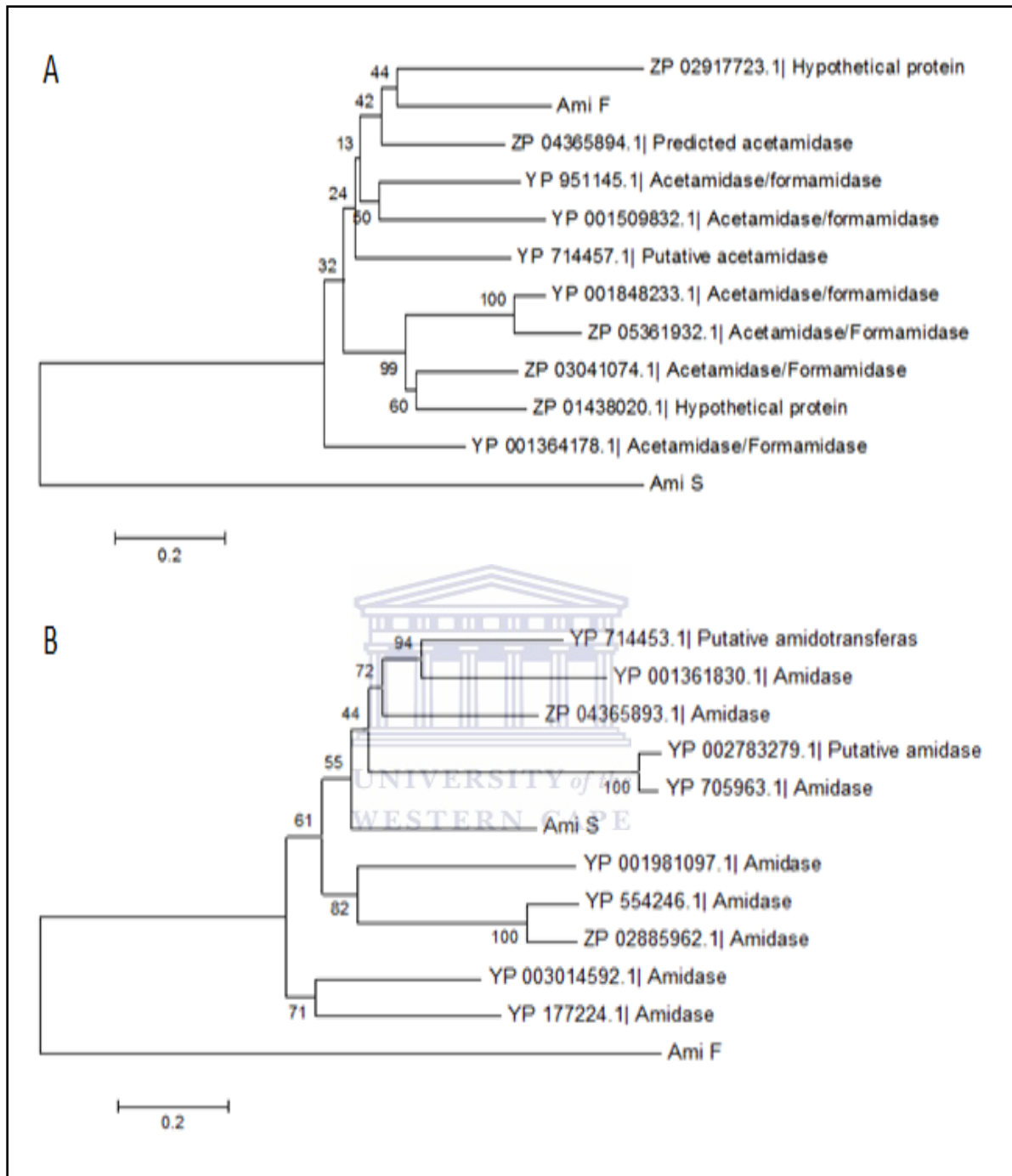


Figure 3-5 Neighbour-joining phylogenetic dendrograms of AmiF (A) and AmiS (B).

AmiF and AmiS were used for each other's outlier.

Figure 3-5 was generated using the Neighbour-joining method (Saitou & Nei, 1987). The bootstrap consensus tree inferred from 500 replicates (Felsenstein, 1985), is taken to represent the evolutionary history of the taxa analysed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

The evolutionary distances were computed using the Poisson correction method (Zuckerkindl & Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset. There were a total of 337 positions for AmiF and 323 positions for AmiS in the final dataset. Phylogenetic analysis was conducted in MEGA4 (Tamura *et al.*, 2007). The dendrograms showed that AmiF and AmiS belong to the expected families, and also clearly demonstrated that these two proteins do not belong to the same family.

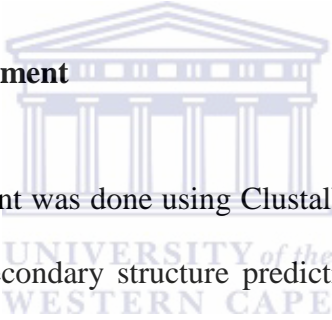
3.4 Homology Modeling

As the quaternary structures of neither protein are known, the models were constructed as monomers.

3.4.1 Search for Templates

The protein sequences of AmiF and AmiS were submitted to GenTHREADER (Jones, 1999) and FUGUE (Shi *et al.*, 2001) to identify structural homologues. The templates suggested by both servers as significant matches were used for the modeling process. The template chosen for AmiF were crystal structure of acetamidase from *Bacillus Halodurans* at 1.95 Å resolution (PDB ID: 2III, chain D) and the structure of tRNA dependent amidotransferase GAT CAB (PDB ID: 2f2a, chain A) for AmiS (Nakamura *et al.*, 2006).

3.4.2 Pairwise Sequence Alignment



The pairwise sequence alignment was done using ClustalW in BioEdit. Alignments were adjusted manually based on secondary structure prediction of the target proteins from PSIPRED (McGuffin *et al.*, 2000) to avoid the presence of any gaps in their secondary structures. Figure 3-6 and Figure 3-7 show the results of the alignment between AmiF and 2III and AmiS and 2F2A respectively. Both alignments resulted in about 25% sequence identity which was regarded as low score.

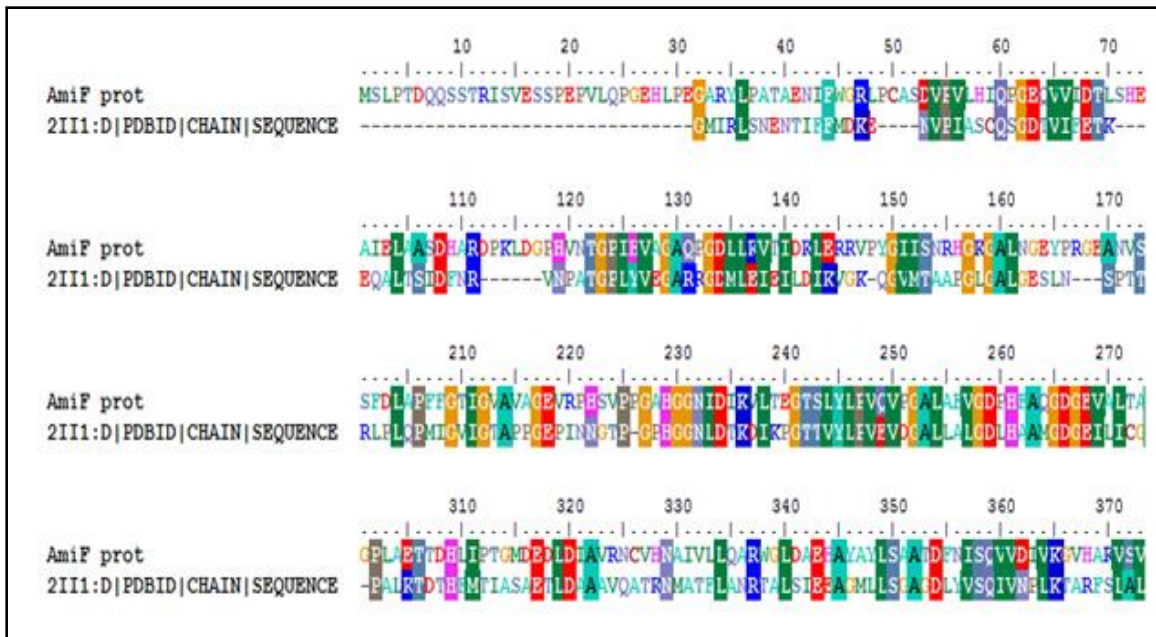


Figure 3-6 Alignment between AmiF and 2iI1

The sequence identity between AmiF and the template 2iI1 is 25%

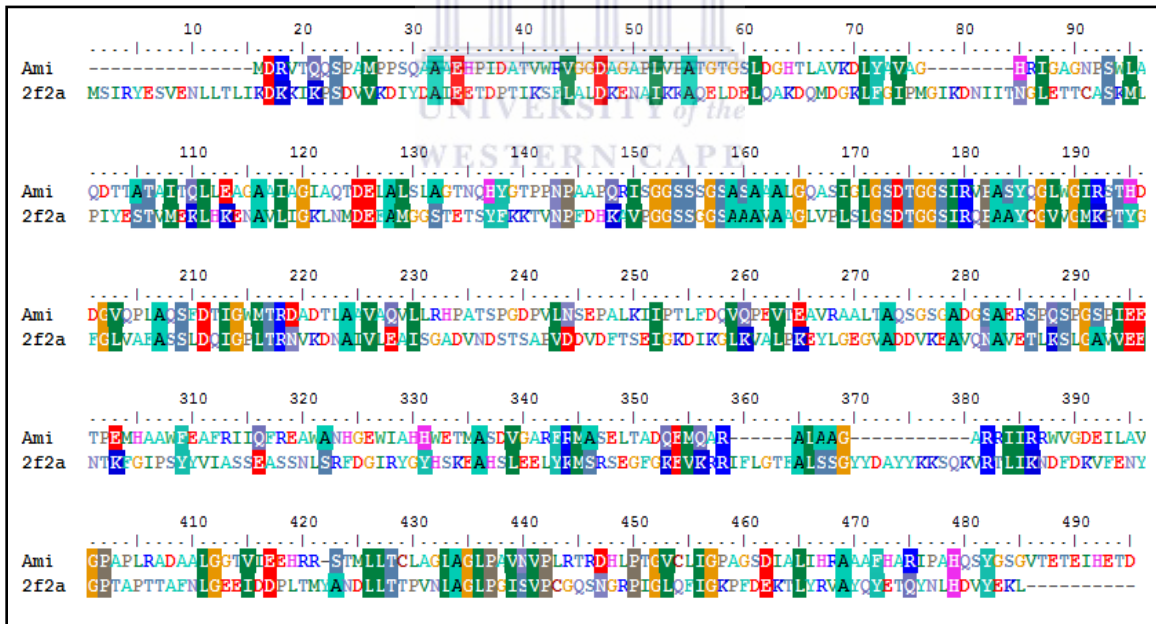


Figure 3-7 Alignment between AmiS and 2f2a

The sequence identity was 24%.

3.4.3 Model Building and Side-Chain Optimisation

The alignment files were converted into PIR format for subsequent use by MODELLER 9v7 (Sali & Blundell, 1993). Twenty cycles of modeling were executed with continuous scanning of the log files for errors. This process was iterated until acceptable models were generated. The model was visualized with Chimera (Eric *et al.*, 2004). The orientation of the side-chains was optimised using SCWRL 4.0 (Bower *et al.*, 1997). To use the SCWRL server, the user needs to provide a template structure in the PDB format and an alignment of the target sequence with the template sequence. The server outputs the alignment between the full template sequence and the template sequence corresponding to atom entries in the template structure. This alignment shows which regions of the template structure that are not known. Before model building the original target-template alignment is automatically adjusted to take their discrepancies into account. Then the model is built by replacing all template residues with target residues according to the target-template alignment.

For both proteins, there were visible changes in the orientation of the side chains after optimisation using SCWRL 4.0 (Bower *et al.*, 1997) (Figure 3-8 & Figure 3-9). The models before and after the side-chain optimisation were evaluated.

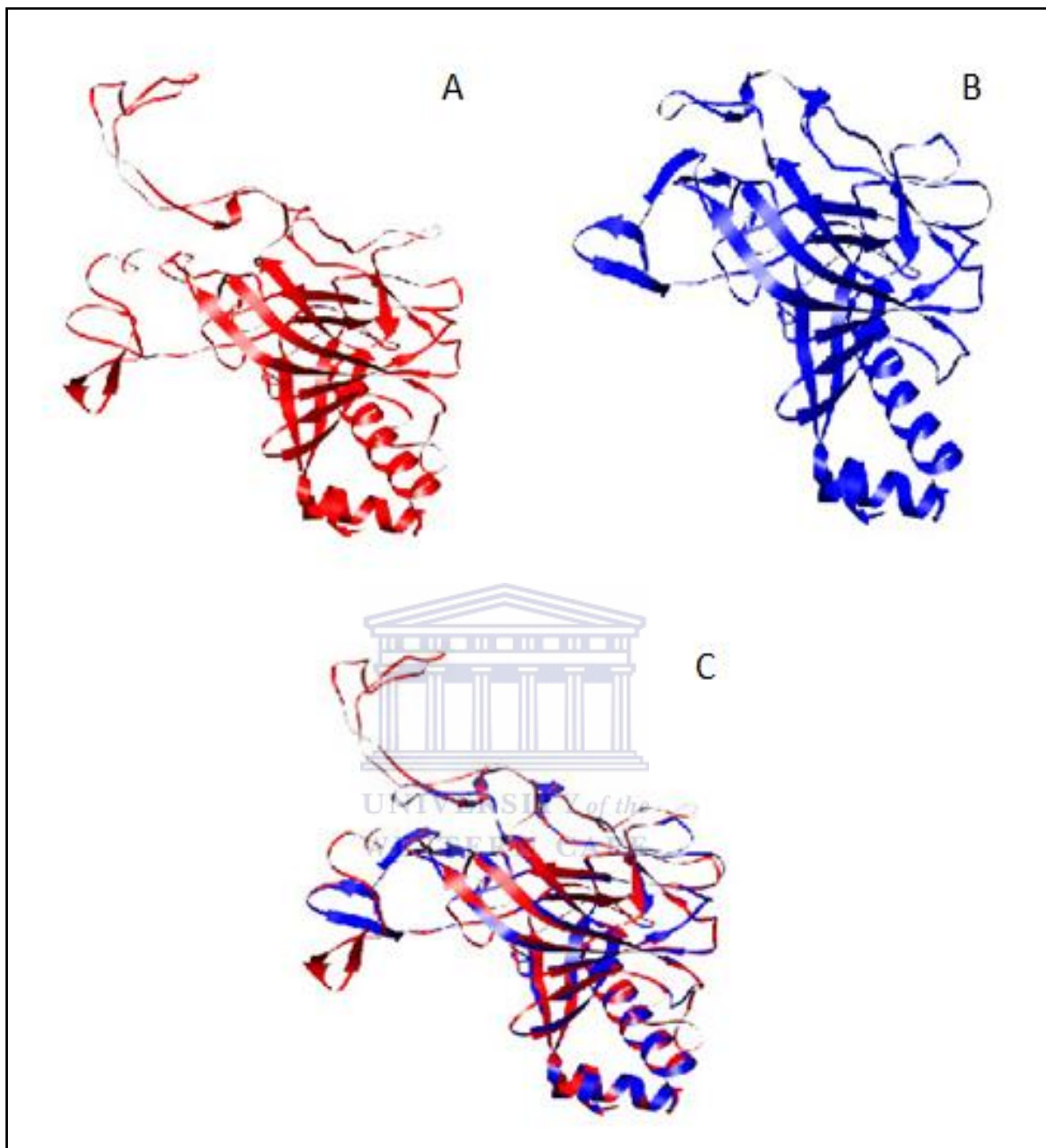


Figure 3-8 Cartoon representation of AmiF before (A) and after (B) side-chain optimisation (c) Superimposed (A) and (B)

Figures generated by Chimera (Eric *et al.*, 2004)

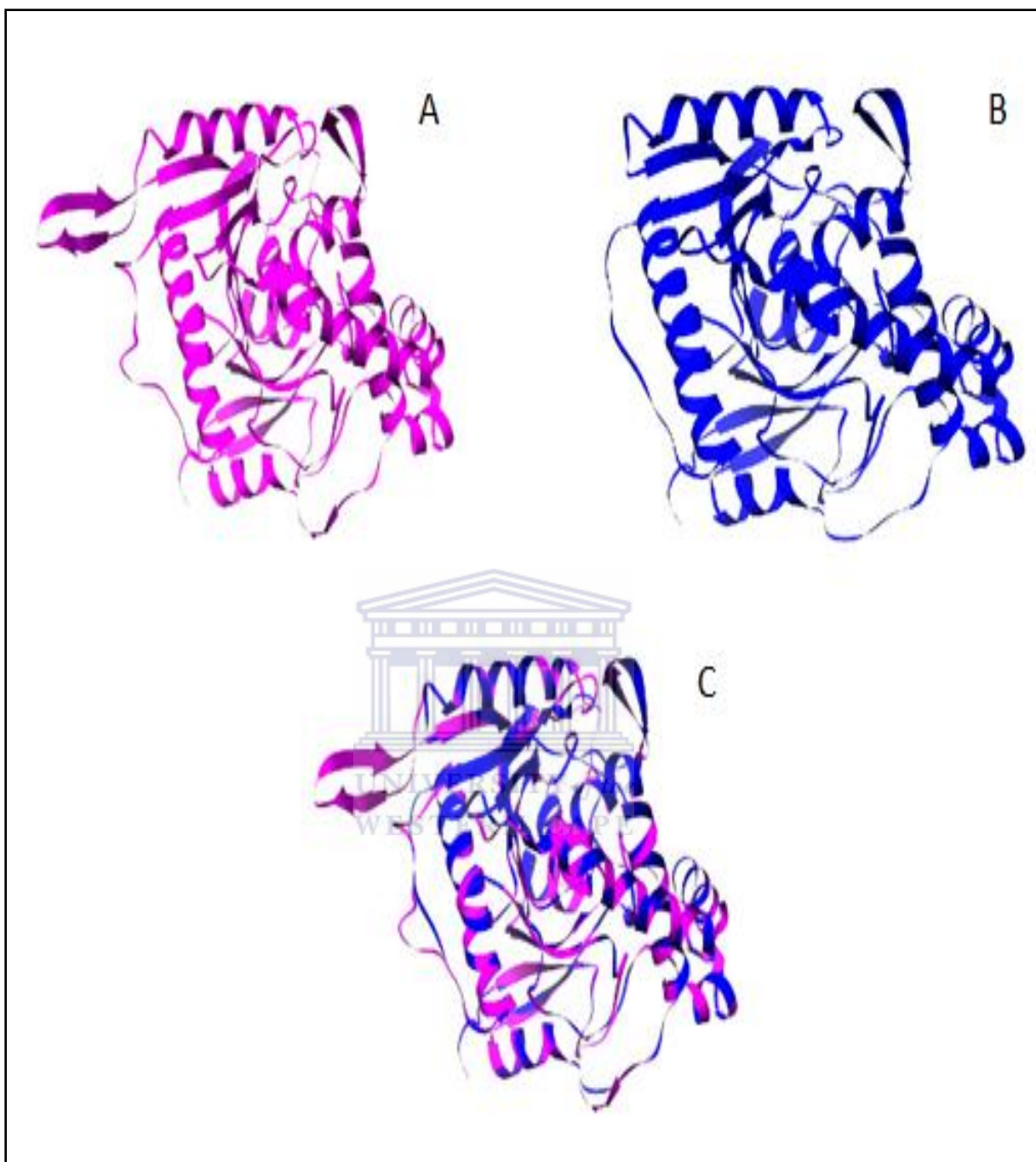


Figure 3-9 Cartoon representation of AmiS before (A) and after (B) side-chain optimisation (c) Superimposed (A) and (B)

Figures generated by Chimera (Eric *et al.*, 2004)

3.4.4 Model Evaluation

With low sequence identities between target and template sequences, it is expected that the accuracy and reliability of the models would be low. In addition the structural deviations between family members would reduce the reliability of the model. Hence, the accuracy of the models were assessed using MOLPROBITY (Davis *et al.*, 2004), which plots the stereochemistry of residues on a Ramachandran plot. The percentage of the residues of AmiF model in the favored, allowed and outlier region were 97.45, 2.13 and 0.42 %, respectively (Figure 3-10). AmiS had values of 93.91, 4.15 and 1.94 % respectively (Figure 3-11). Values for both proteins were slightly improved compared to the models generated before the side-chains were optimised (data not shown). Residues that lied in the outlier regions were found to be located in the loop regions of the protein and not in the core structures. None of the ‘outlier’ residues were located in the catalytic triad region. Together, these data indicate that the final models for the proteins are acceptable. The presence of outlier residues could be due to errors in the alignment, wrong orientation of side chains and variability of the loop regions.

The validation of the AmiF and AmiS homology models indicated that they were suitable for 3D comparative analysis. The analysis was done using the SSM server (<http://www.ebi.ac.uk/msd-srv/ssm/>). The SSM server compares 3D structures of AmiF and AmiS to PDB database and calculates Root Mean Square Deviation (RMSD) values which represent the distances between C α -atoms of matched residues at best 3D

superposition of the query and target structures; i.e, RMSD simply gives an indication of the positional accuracy of a pair of matched C α -atoms. The larger RMSD values indicate the lower accuracy of the models.

For the AmiF model, the best match suggested from the PDB database was the 1.95Å crystal structure of acetamidase from *Bacillus halodurans* (PDB ID; 2ii1, chain B). The RMSD values between the model and the crystal structure was only 0.23Å suggesting that despite the low sequence identity (25%), the structure homology was very high. This might suggest that the structures of these family members have been conserved. For the AmiS model, the best match suggested by the server was a crystal structure of signature amidase (PDB ID; 2dc0, chain B). The RMSD value was 0.00Å which suggests that the two structures are identical, despite the low sequence identity of 20%. The SSM server did not indicate the template (PDB ID: 2f2a) as one of close match to the AmiS model although the 2f2a was suggested by two different servers. The RMSD between AmiS and 2f2a was 0.95Å. Considering RMSD for proteins share 50% residue identity are expected to be 1Å, 0.95Å between 2f2a and the model of AmiS is a good value as they only share 24% sequence identity.

The 3D structure analysis results also indicated that the both models were reasonably well-built and acceptable.

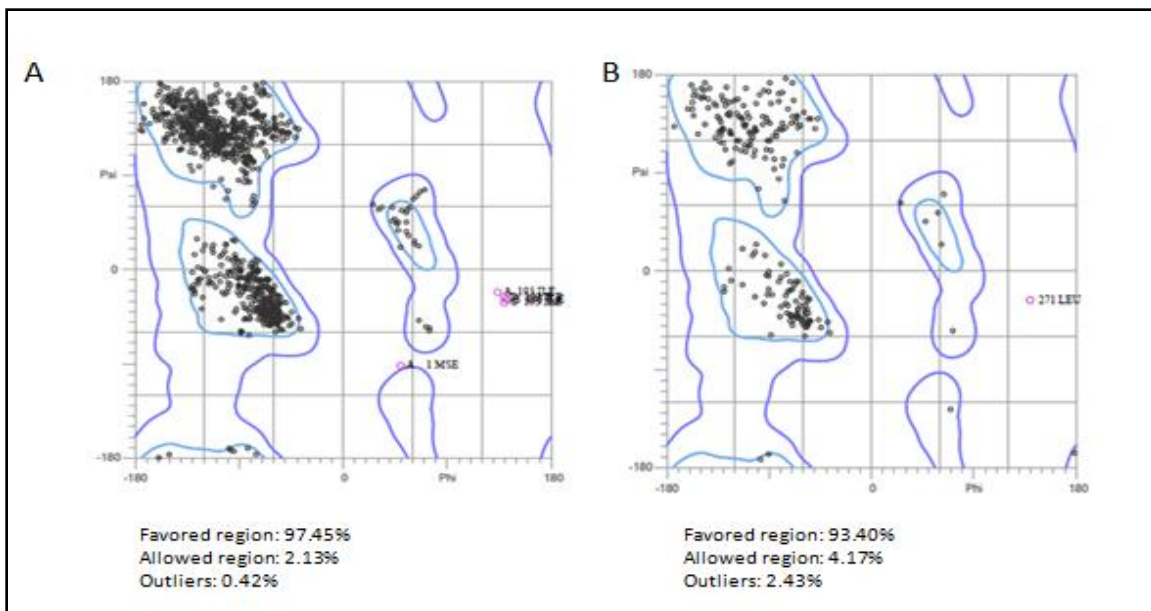


Figure 3-10 Ramachandran plots for the template (A, PDB ID: 2ii1) and the AmiF model (B).

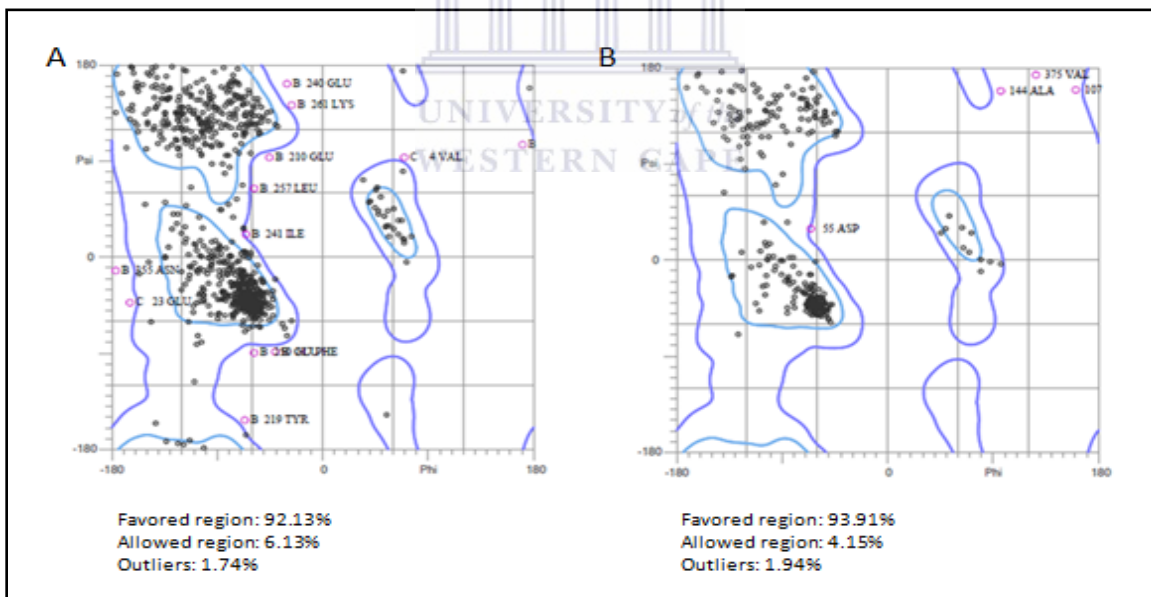


Figure 3-11 Ramachandran plots for the template (A, PDB ID: 2f2a) and the AmiS model (B).

3.4.5 Description of AmiF and AmiS homology models

Few crystal structures of aceta/formamidase family members are available and none of these structures are supported by any publications after the submission of the structures to the PDB database. Hence, the comparison of the AmiF structures to a wider range of known structures was not possible at this point. However, when the model of AmiF and its template was superimposed, they showed good structure conservation of core β -sheet structure as well as α -helices C-terminal structure (Figure 3-12, panel A) despite the low sequence identity. Catalytic residues for this family are yet to be found.

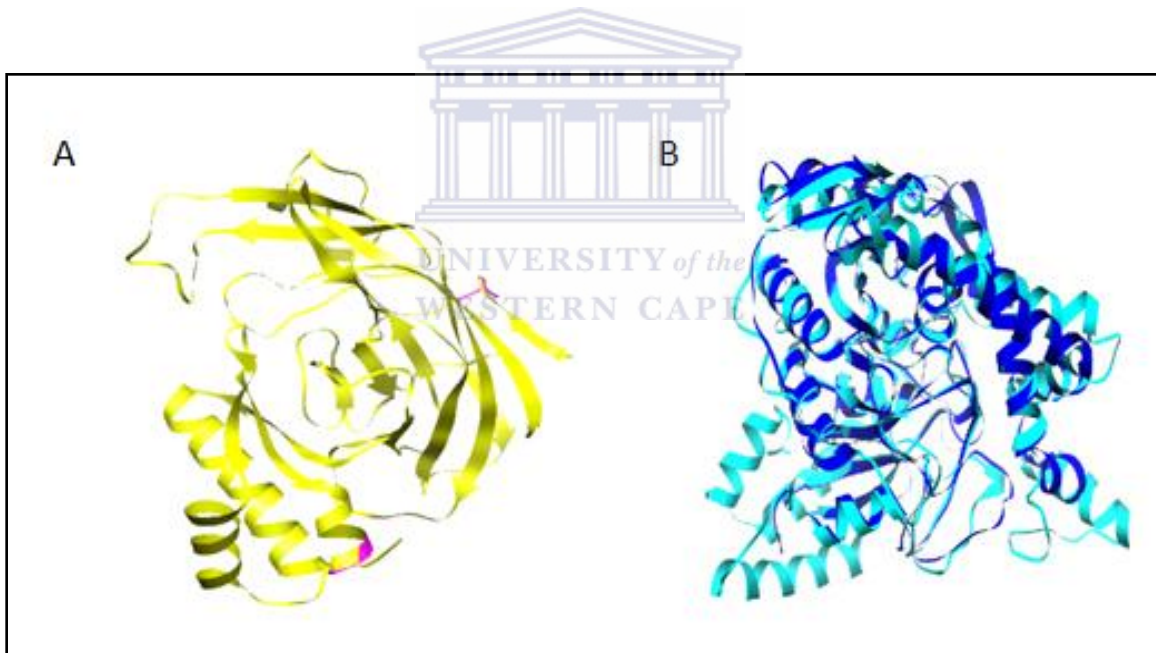


Figure 3-12 Superimposed model and template structures.

A: AmiF (pink) and 2ii1 (yellow), B: AmiS (blue) and 2f2a (cyan)

For AmiS, the superimposition of two structures showed good structural conservation of the structure although the template structure has longer N-terminal helices that did not match the AmiS model (Figure 3-12, panel B).

AmiS consists of a central 10 β sheet core covered by double layers of α helices on the top and bottom (Figure 3-13). The conserved signature amidase sequence forms the core of the structure.

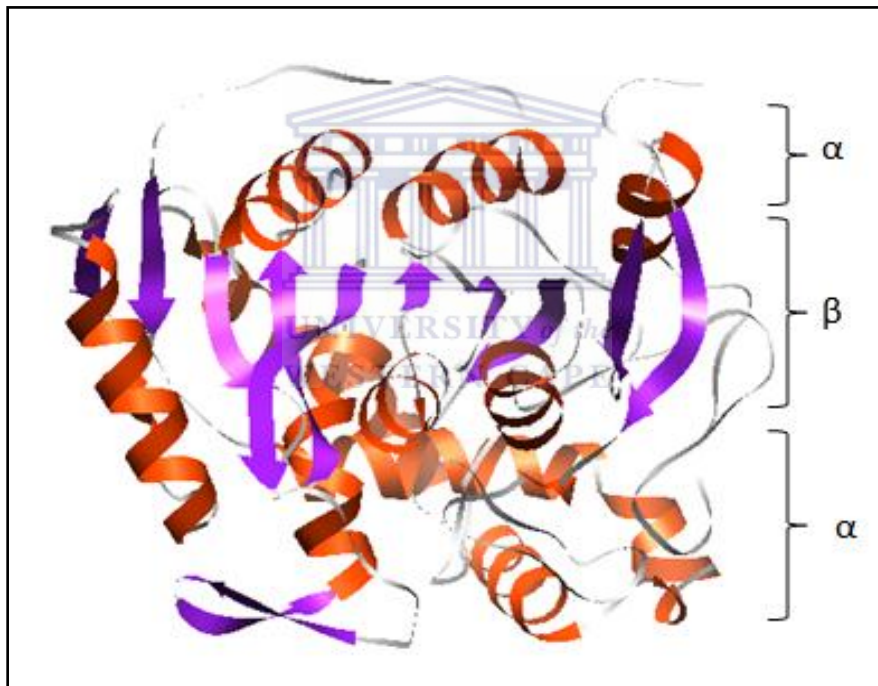


Figure 3-13 AmiS monomer fold.

The α/β sandwich fold with a central β -sheet core surrounded by α -helices is conserved across the members of the amidase signature family.

Sequence analysis using the Pfam server (pfam.sanger.ac.uk) indicated a possible catalytic triad of residues Ser131-Ser155-Lys54. When the model and the template were superimposed, the catalytic triad showed exact conservation of residue positions between the two structures (Figure 3-14). Furthermore, Ser 155 residue showed a novel *cis*-Ser conformation which is found in some signature amidase family members (Nakamura *et al.*, 2006; Shin *et al.*, 2003). This residue is located in the loop enriched in small residues (glycine, serine and alanine).

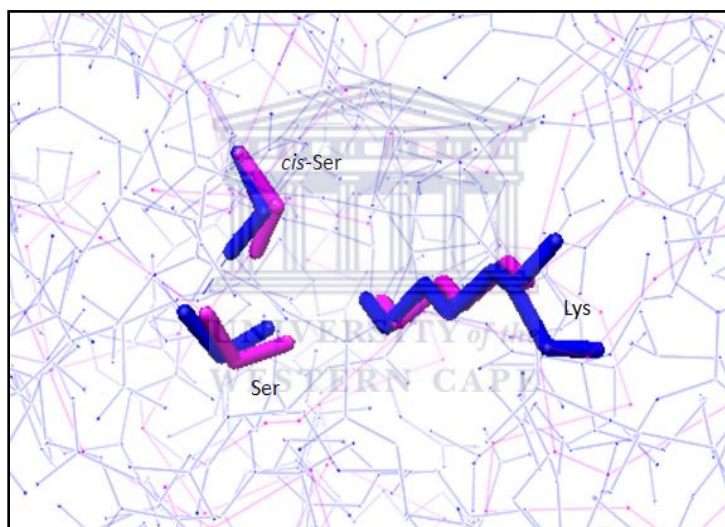


Figure 3-14 Superimposed catalytic triad of the AmiS model (pink) and the template (blue, PDB ID: 2f2a).

The catalytic residues namely Ser131, *cis*-Ser 155 and Lys54 are shown as sticks (the numbering is according to the target sequence). Figures generated by Chimera (Eric *et al.*, 2004).

The signature amidase catalytic mechanism suggests that there are 2 more residues that are involved in the amide catalysis process (Ala111 and Arg158, numbering according to

the structure of Malonamidase. PDB ID: 1OCK) (Shin *et al.*, 2002). The superimposition of 1OCK and the AmiS model shows good positional agreement in the core structure, however the structural alignment of the additional residues involved in the catalysis showed good agreement for the alanine residue but not for the arginine residue (Figure 3-15). In AmiS, a valine residue was found instead of the arginine residue.

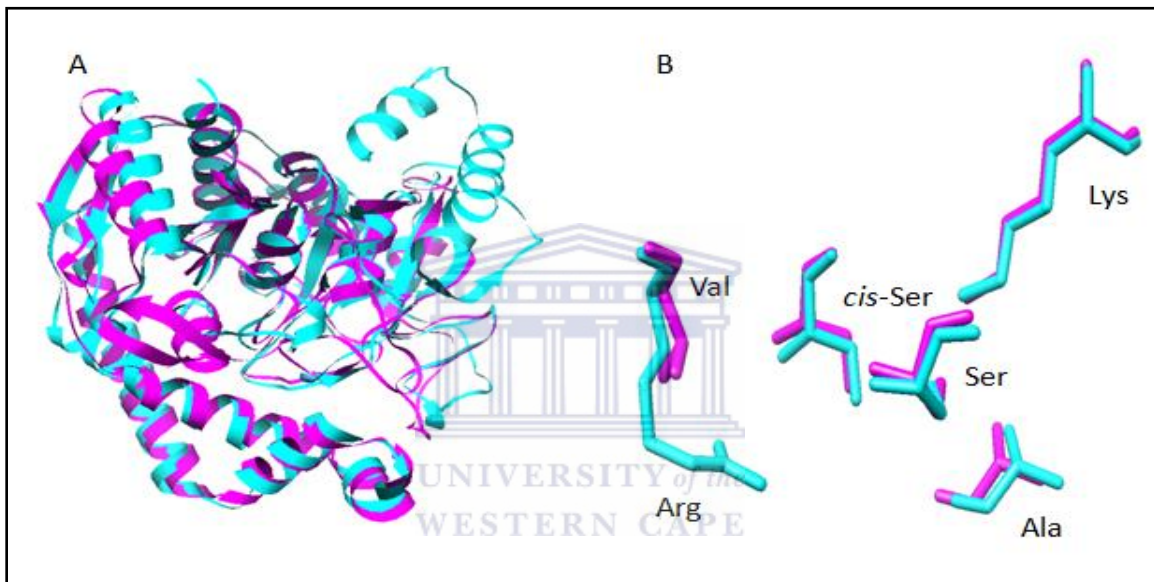


Figure 3-15 Superimposition of 1OCK (cyan) and AmiS model (pink) (A) and the residues involved in catalysis (B)

3.5 Discussion

The genome of *Nesterenkonia* sp. was sequenced and assembled using Illumina DNA sequencing technology at the University of the Western Cape. Two ORFs of 1140 bp and

1365bp, encoding for the aceta/formamidase (AmiF) and signature amidase (AmiS) respectively were identified. The alignments of the translated sequences of two proteins with database sequences with the highest identities indicated close phylogenetic relationships of the two proteins to their family members.

The 3D structure model of AmiF and AmiS were constructed based on the closest similarity to the experimentally determined structures of the amidase for AmiF and the subunit A of amidotransferase for AmiS. The structural model generated was found to be in a good agreement with the template structure, suggesting a high level of accuracy for the models.

Although the solved structures of the same family members of AmiF were available, the proteins were not further investigated. Hence it was not possible to make the comparison of structures in detail or predict the possible catalytic residues. However, the superimposed structures of the AmiF and the template structure suggested that there was good conservation in their structures.

The catalytic triad residues of AmiS were shown to be conserved between the model and the template structures. AmiS shared the novel catalytic triad Ser-*cis*Ser-Lys that is absolutely conserved throughout the family. Ser 131 is the nucleophile initiating the catalysis. The unusual *cis* configuration also allows two direct hydrogen bonding

interaction between the two serine residues in the catalytic triad. Upon the nucleophile attack of Ser 131 on the carbonyl carbon of the substrate, *cis*Ser 155 is proposed to function as catalytic acid that provides a proton to the leaving amino group (Shin *et al.*, 2002). The proton transfer should be facilitated by the amino group of Lys 54 that can stabilize the deprotonated form of *cis*Ser 155 and this residue is also involved in maintaining the structure as the substitution of the residue disrupts the active site's structure completely (Shin *et al.*, 2003). The ammonia produced from the catalysis leaves the active site and a water molecule occupies the vacated site. Activated by *cis*Ser 155, the water molecule attacks the carbonyl carbon atom of the acyl enzyme intermediate to finally regenerate Ser 131 (Shin *et al.*, 2003). Since the pKa of serine is around 13, *cis*Ser 155 is most likely in the protonated form at physiological pH. The chemical environment around Ser 155 is not unusually polar to promote deprotonation of the residue. Rather Ala 111 renders the environment partly hydrophobic (Shin *et al.*, 2003). Ser 155 is in the deprotonated state mainly by virtue of two direct hydrogen bonding interactions provided by the backbone –NH and the side chain –OH groups of *cis*Ser 155 and by the additional dipolar interaction provided by the guanidine group of Arg 158 (Shin *et al.*, 2003). Although Ala 111 showed a good match with Ala 105 from AmiS (position deduced from the alignment between 1OCK and AmiS), Arg158 was aligned with Val158 from AmiS. The lack of a guanidine group in the the valine residue may affect the deprotonated state of catalytic residue Ser 155 (Figure 3-15, panel B). In addition, the substitution of the hydrophobic valine residue to the hydrophilic arginine residue might affect the core

structure environment. The possible disruption of the core environment may prevent the formation of bonds between the catalytic residues which would in turn affect the activity of AmiS.



Chapter 4

Expression and Characterisation of Amidases

4.1 Introduction

Amidases belong to the hydrolase family, a subclass of acrylamide amidohydrolases (EC 3.5.1, 3.5.2). Amidases are still not sufficiently investigated and their classification is not definitively formulated (Pertsovich *et al.*, 2005). However, the amidases can be roughly categorised into two families (Fournand *et al.*, 1998b; Kimani *et al.*, 2007; Pertsovich *et al.*, 2005). First, enzymes containing Gly-Gly-Ser-Ser (GGSS) motif in the primary structure which are called signature amidases. The second group of enzymes lack the GGSS motif and they generally belong to the nitrilase superfamily whose members all share a novel catalytic triad Glu-Lys-Cys (EKC). Some aliphatic amidases which lack both characteristic sequences roughly belong to aceta/formamidase family (EC 3.5.1.-)

Two ORFs with conserved amidase sequences were identified from the complete genome sequence of the nitrile hydrolysing *Nesterenkonia* sp. Two genes, named *amiF* and *amiS*, were assigned to two different families: the hydrolase family and amidase signature family, respectively. *AmiS* belongs to the amidase signature family and it shared the characteristic signature motif with the amino acid sequence of GGSS. *AmiF* roughly

belongs to aceta/formamidase family since it lacks the both GGSS and EKC sequences. Few amidases from each family have been isolated and heterologously expressed in *E.coli*.

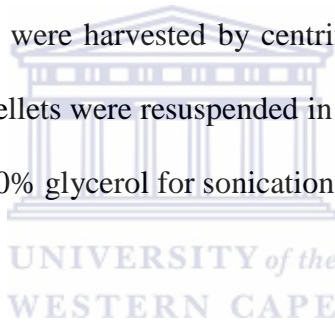
This chapter describes the cloning, expression and purification of the two amidase encoding genes and their subsequent characterisation.



4.2 Materials and Methods

4.2.1 Expression of AmiF

Protein expression was carried out in Rosetta2 (DE3) pLysS. For AmiF, a 5ml culture was inoculated overnight and used to inoculate 500ml LB broth containing 50µg/ml kanamycin and 34µg/ml chloramphenicol. The cells were incubated at 37°C until the optical density (600nm) reached 0.4. The cells were induced by addition of 0.4mM isopropyl thio-β-D-galactoside (IPTG), and incubation at room temperature (RT) for 4hrs. After induction, the cells were harvested by centrifugation and washed with Tris-HCl pH 7.9. The washed cell pellets were resuspended in 50ml of Tris-HCl buffer pH 7.9 containing 200mM NaCl and 10% glycerol for sonication.



4.2.2 Expression of AmiS

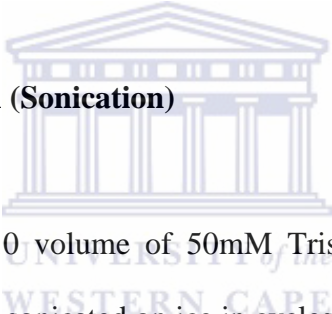
AmiS was expressed under various conditions. Cells were grown at RT until OD₆₀₀ was 0.4 then IPTG was added at concentrations ranging from 0.1mM to 1mM. Each culture was induced for 4hrs at room temperature. Aliquots were taken at 1, 2, 3 and 4hr intervals to identify optimum expression conditions for Ami S. Ami S was also expressed without adding IPTG by incubating at room temperature for 24hrs. After expression, Ami S was treated the same as Ami F, as described in section 4.2.1.

4.2.3 Cell Lysis Methods

4.2.3.1 Enzymatic Methods

Cultures were centrifuged at 8000 x g for 10mins and the pellets were resuspended in Bugbuster (Novagen, USA) (5ml per 1g of cell pellet) and Benzonase nuclease (Novagen, USA) [1U/ml Bugbuster]. The mixture was incubated at room temperature for 30mins with gentle agitation. The lysed cells were centrifuged at 13000 x g for 10mins and the supernatant was transferred to a sterile tube.

4.2.3.2 Mechanical Disruption (Sonication)



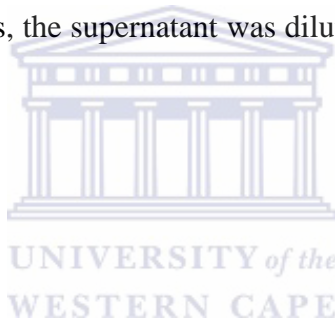
Cells were resuspended in 1/10 volume of 50mM Tris-HCl (pH8.0) containing 10% glycerol and 200mM NaCl, and sonicated on ice in cycles of 30sec pulse and 30sec pause for 8mins per 50ml of culture volume.

4.2.4 Solubilisation and Refolding of AmiS

AmiS was expressed as indicated unless stated otherwise. Culture was grown at 37°C until OD₆₀₀ reached 0.4 then IPTG was added to a final concentration of 0.4mM. Cells were incubated for a further 4hrs at RT.

4.2.4.1 Urea Solubilisation

Pellets were washed with 50mM Tris-HCl (pH 8) containing 1mM EDTA and 100mM NaCl and resuspended in urea concentrations ranging from 3M to 6M containing 50mM Tris-HCl and 10mM DTT. The mixtures were incubated at RT for 30mins with agitation. After centrifugation at 4°C for 10mins, supernatants were collected. The supernatants were dialysed against the dialysis buffer (50mM Tris-HCl, pH 8, 10% glycerol and 200mM NaCl) at 4°C for 2 days changing the buffer every 12 to 18 hrs. The dialysed protein was purified using His-Tag purification column according to manufacturer's recommendation. In some cases, the supernatant was diluted 33-folds with dialysis buffer instead of dialysis.

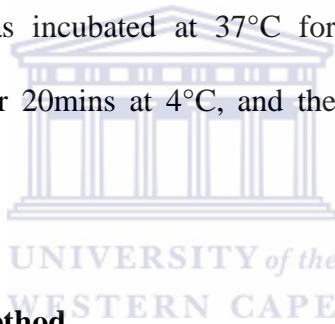


4.2.4.2 Two Step Refolding

The pellets were washed with the same buffer then resuspended in solubilisation buffer (6M Urea, 30mM Tris-HCl, pH 8, 100mM NaCl, 1mM DTT and 20% glycerol) and incubated at RT for 30mins with agitation. After incubation the sample was filtered twice through a 0.45µm filter then dialysed against dialysis buffer with 3M Urea and 1mM DTT for 12hrs. The buffer was changed and the dialysis was continued for further 24hrs. The dialysed sample was subjected to His-Tag purification.

4.2.4.3 Chloramphenicol Method

This solubilisation method was adapted from Villaverde *et al.*, (2002). A 5ml overnight culture was inoculated in LB containing 50µg/ml kanamycin and 34µg/ml chloramphenicol. 1ml of the overnight culture was used to inoculate 1L LB containing kanamycin but not chloramphenicol. The culture was incubated at 37°C for 6hrs. IPTG was added to the culture to a final concentration of 1mM and incubated at 30°C overnight. Cells were pelleted by centrifugation at 4,000 x g for 20mins. The pellet was resuspended in 500ml LB. Chloramphenicol was added to a final concentration of 200µg/ml and the culture was incubated at 37°C for 2hrs. Cells were pelleted by centrifugation at 4,000 x g for 20mins at 4°C, and the pellet was resuspended in the dialysis buffer for sonication.



4.2.4.4 β-mercaptoethanol Method

This solubilisation method was adapted from Panda *et al.*, (1997). The pellet was washed twice with 50mM Tris-HCl (pH 8) then resuspended in solubilisation buffer (30mM Tris-HCl, pH 8, 100mM NaCl, 20% glycerol, 8M β-mercaptoethanol and 2M Urea). The sample was incubated for 1hr at RT, then filtered twice through a 0.45µM filter before dialysed against 30mM Tris-HCl (pH 8), 30mM NaCl, 10% glycerol and 2M Urea) at 4°C for 24hrs. The buffer was changed and dialysis was continued for a further 24hrs.

The sample was subjected to His-Tag purification under denaturing condition using 2M urea and the eluted fraction was dialysed against dialysis buffer without any urea for 24hr at 4°C.

4.2.4.5 Sodium Deoxycholate Method

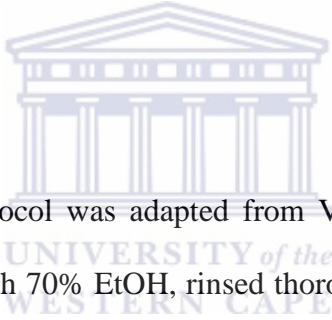
The cell pellets were resuspended in 50mM Tris-HCl (pH 8.0) containing 5mM EDTA then sonicated for 8 cycles. The sample was centrifuged at 10,000 x g at 4°C for 20mins, the pellets were resuspended in 50mM Tris-HCl (pH 8.0) containing 1% deoxycholate and sonication was repeated for 8 cycles. The pellet was collected by centrifugation, resuspended in the same buffer and incubated overnight at 37°C. After incubation the sample was sonicated for another 8 cycles then centrifuged at 10,000 x g for 20mins at 4°C. The pellet was resuspended in 50mM Tris-HCl (pH 8.0) then incubated at room temperature for 30mins. The purified inclusion bodies were collected by centrifugation, solubilised using 50mM Tris buffer (pH 12.5) containing 2M urea, 10% glycerol, 100mM NaCl and 0.5mM EDTA and dialysed against standard dialysis buffer.

4.2.5 His-Tag Affinity Chromatography

The His-Bind resin and Buffer kit (Novagen, USA) was used for the affinity chromatography purification of His-tag fusion proteins. Cell lysates were prepared using

either Bugbuster protein extraction reagent (Novagen) or by sonication. After sample preparation, the purification was followed according to the manufacturer's instructions with the exception of the elution step. Elution buffer volume was decreased to 2ml instead of recommended 6ml volume to concentrate the elution fraction. The eluate was dialysed overnight in a 5ml Slide-A-Lyzer Dialysis cassette (Thermo Fisher Scientific) against 50mM Tris-HCl (pH 8.0), 10% glycerol and 200mM NaCl. The purified protein was stored at 4°C. The fractions of each step of His-Tag affinity chromatography were analyzed by SDS-PAGE.

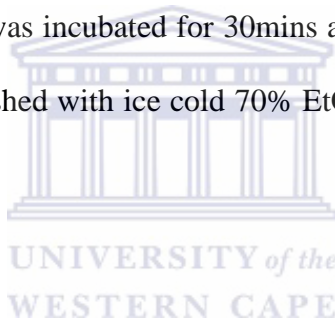
4.2.6 RNA Extraction



The total RNA extraction protocol was adapted from Verwoerd *et al.* (1989). All the glassware used was washed with 70% EtOH, rinsed thoroughly with DEPC treated Milli Q (MQ) water. All solutions and tips were autoclaved twice before use.

A 50ml culture was harvested by centrifugation at 9500x g for 5mins. The pellet was resuspended in 500µl of ice cold 0.01M NaAc/0.3M sucrose (pH 4.5). The solution was split into two 2ml tubes then 500µl of 0.01M NaAc/2% SDS (pH 4.5) was added to each tube. The samples were heated at 65°C for 90sec then the equal volume of hot phenol (65°C, pH 4) was added. Tubes were vortexed briefly and incubated for 3mins at 65°C. The samples were rapidly cooled in ice then centrifuged at RT for 10mins. The aqueous

phase was recovered and the hot phenol extraction step was repeated three times. 1/10 volume of 3M NaAc (pH 5.2) and 3 volumes of 100% EtOH were added to the samples then incubated for 30mins at -80°C. The samples were centrifuged for 15mins at 4°C and the pellet was washed with 70% ice cold EtOH. The pellet was dried and resuspended in 200µl of DEPC treated water. An equal volume of phenol:chloroform (1:1) was added then centrifuged at for 2mins. The aqueous phase was transferred into a new 1.5ml tube then the equal volume of chloroform:isoamylalcohol (24:1) was added to the tube. The pellet was collected by centrifugation at RT for 2mins. The top layer was transferred to a new tube and 1/10 3M NaAc (pH 5.2) and three volumes of 100% EtOH were added to precipitate RNA. The sample was incubated for 30mins at -80°C then centrifuged at 4°C for 15mins. The pellet was washed with ice cold 70% EtOH then resuspended in 40µl of DEPC treated MQ water.



4.2.7 Reverse Transcription PCR (RT-PCR)

For RT PCR, four reactions were set up as follows;

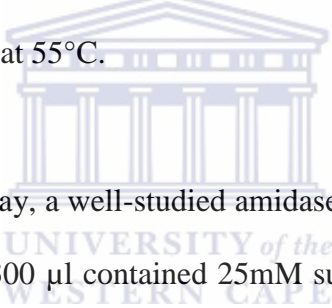
Table 4-1 Reverse Transcription PCR Reaction Set Up

	Positive Control	Reverse Transcriptase Negative Control	No Template	Experimental
RNA	GAPDH RNA 2µl (50ng/µl)	GAPDH RNA 2µl (50ng/µl)	-	2µl (50ng/µl)
Primers	GAPDH Reverse Primer 1 µl	GAPDH Reverse Primer 1 µl	GAPDH Reverse Primer 1 µl	RTR 1 µl
5X Reaction Buffer	4 µl	4 µl	4 µl	4 µl
RiboLock RNase Inhibitor (20u/µl)	1 µl	1 µl	1 µl	1 µl
10mM dNTPs	2 µl	2 µl	2 µl	2 µl
Reverse Transcriptase (20u/µl)	2 µl	-	2 µl	2 µl
Water	8 µl	10 µl	8 µl	8 µl
Total Volume	20 µl	20 µl	20 µl	20 µl

All the reactions were incubated at 37°C for 60mins then at 70°C for 5mins to stop the reaction.

4.2.8 Activity Assay

Enzyme activity was determined by the release of ammonia using the phenol-hypochlorite ammonia detection method (Weatherburn, 1967). A standard reaction mixture contained 50mM Tris-HCl (pH 7.9), 200mM NaCl, 10% glycerol, 10-100mM substrate and 10-50µl of cell free extract. The reaction was terminated by the addition of 3.5volume of reagent A (0.59M phenol and 1mM sodium nitroprusside) followed by addition of an equal volume of reagent B (2M sodium hydroxide and 0.11M sodium hypochlorite) to develop color. Ammonia release was measured spectrophotometrically at OD_{600nm} after 5mins incubation at 55°C.



As a positive control for the assay, a well-studied amidase from *Geobacillus pallidus* was used. The reaction mixture of 300 µl contained 25mM substrate, 10 µl partially purified amidase and 50mM potassium buffer (pH 7.2). The mixture was incubated for 1 min at 50°C. This assay was carried out in triplicates.

Negative control reactions were carried out under the same conditions with the enzyme replaced with the same volume of buffer. Standards were prepared using ammonium chloride.

4.3 Results and Discussion

4.3.1 Cloning of AmiF and AmiS

Primers were designed to amplify the two amidase coding genes *amiF* and *amiS*. Both amplified genes were cloned into pET 28 using *NdeI* and *EcoRI* sites on the vector. The transformants containing the vector with inserts were digested with *EcoRV* and the restriction fragment patterns were examined (data not shown). The plasmids producing right size bands were selected. The selected clones were sequenced and the clones with the correct sequences were expressed.

4.3.2 Expression and Purification of Ami F

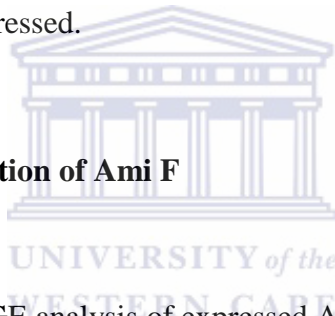


Figure 4-1 shows the SDS-PAGE analysis of expressed Ami F fusion protein in Rosetta 2 (DE3) pLysS cells. A band migrating at ~50kDa was present in induced fractions and absent in the uninduced fraction. The molecular mass of the Ami F fusion protein indicated by SDS-PAGE were slightly different from the calculated value 40kDa without His-Tag and 41kDa with His-Tag using DNAMAN.

To determine whether the protein was expressed in the soluble fraction, SDS-PAGE analysis was used to compare the presence of the ~50kDa protein band in the soluble and

the insoluble fractions. Only a small proportion of the Ami F was present in the soluble fraction.

Cell lysates that were prepared by sonication was subjected to His-Tag affinity chromatography. A protein of expected size was eluted with wash buffer with the majority of the protein eluted with elution buffer. By decreasing the volume of the elution buffer to 2ml, a more concentrated eluate was acquired. Figure 4-2 shows the steps involved in His-Tag purification.

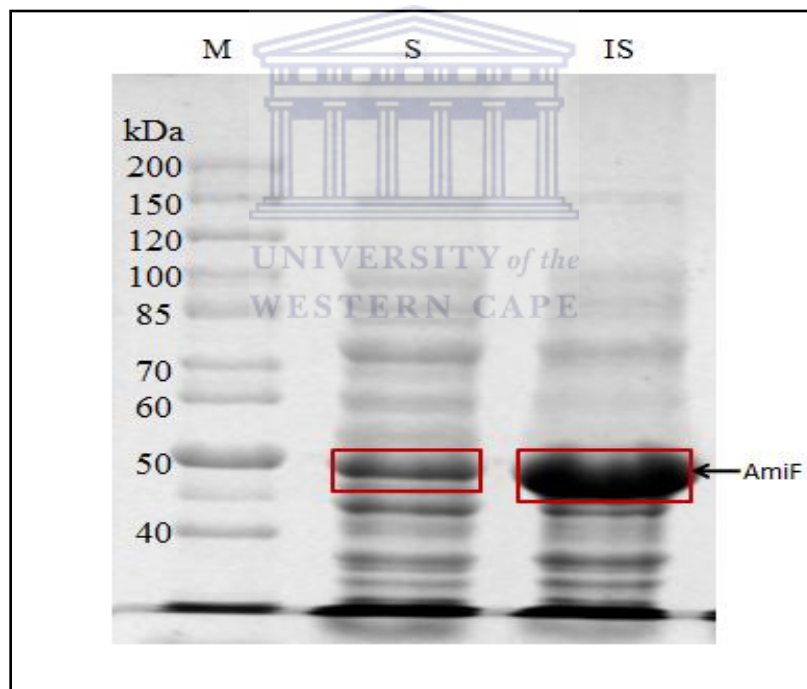


Figure 4-1 Expression of Ami F in *E.coli* Rosetta2 (DE3) pLysS with 0.4mM IPTG

M: Protein Ladder #661, Fermentas, S: Soluble Fraction, IS: Insoluble Fraction. Majority of AmiF expresses in insoluble fraction however there is small amount of AmiF in soluble fraction which was subjected to His-Tag purification.

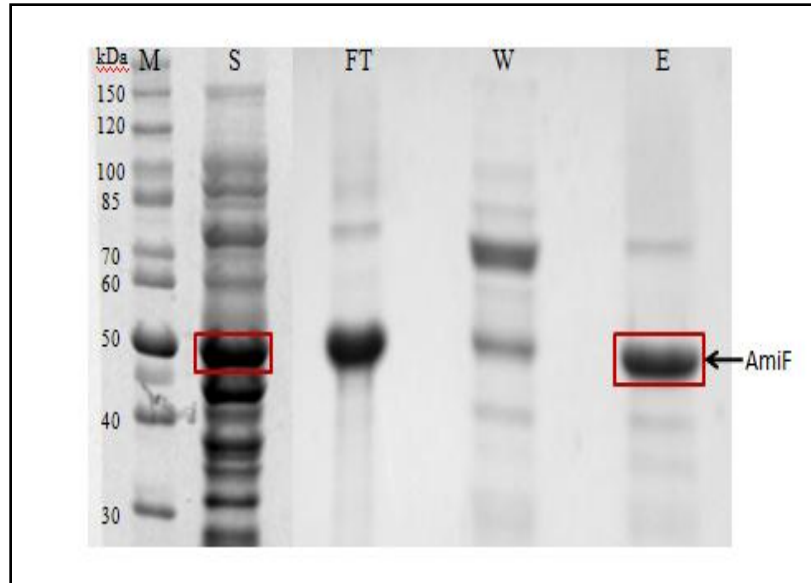
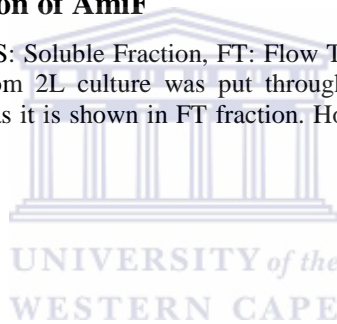


Figure 4-2 His-Tag Purification of AmiF

M: Protein Ladder #661 Fermentas, S: Soluble Fraction, FT: Flow Through, W: Washing Buffer Fraction, E: Elution. The soluble fraction from 2L culture was put through His-Tag column. Large amount of proteins did not bind to the column as it is shown in FT fraction. However, a reasonable amount of AmiF was present in elute fraction.

4.3.3 Expression of Ami S



When *amiS* was expressed with 0.4mM IPTG, SDS-PAGE analysis suggested that all the protein was present in the insoluble fraction, probably in the form of inclusion bodies (Figure 4-3). Expression was assessed at different temperature to determine an optimum expression condition. At 20°C, the protein did not express while at 25 and 30°C, the protein was expressed but was only present in the insoluble fraction (Figure 4-3). The induced fused protein band migrated at ~50kDa, which was very close to the predicted molecular mass of ~47kDa.

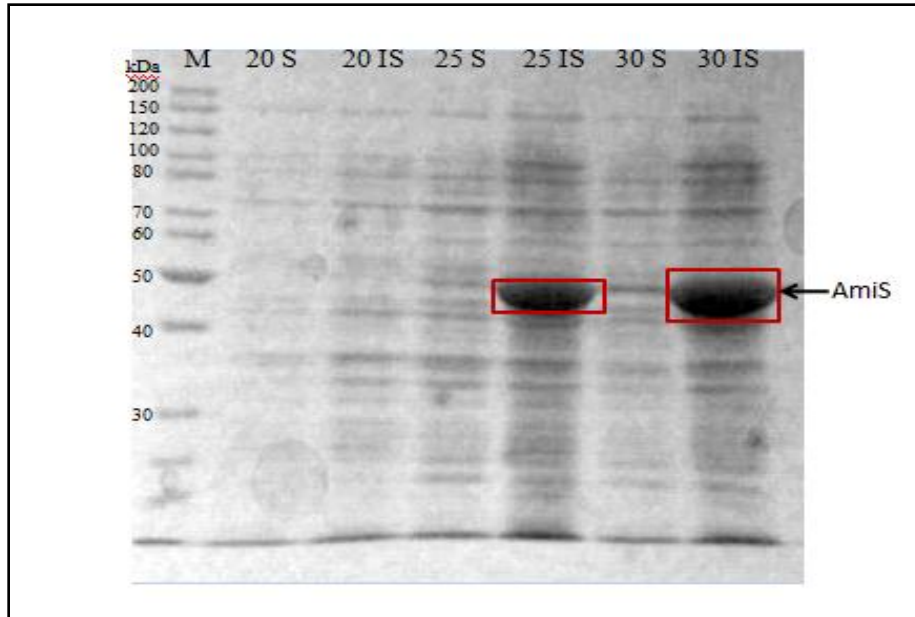
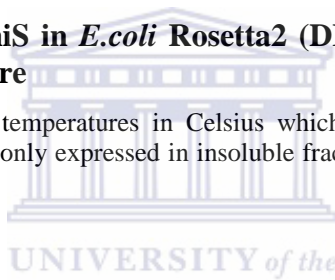


Figure 4-3 Expression of AmiS in *E.coli* Rosetta2 (DE3) pLysS with 0.4mM IPTG induced at various temperature

M: Marker, Number represents the temperatures in Celsius which protein was induced at, S: Soluble fraction IS: Insoluble fraction. AmiS only expressed in insoluble fractions when the proteins were induced at 25 and 30°C.



To express the protein in the soluble fraction, an optimisation of the IPTG concentration was attempted. Figure 4-4 shows that the protein did not express in the soluble fraction at any concentration of IPTG at any time interval. However, AmiS was present in the insoluble fraction at all concentrations of IPTG. The amount of the protein expressed increased as the expression time was extended. In addition, more protein was expressed at the lower the IPTG concentration. This suggests that for AmiS, a high concentration of IPTG inhibits the expression of the protein.

Despite numerous attempts made to find the optimal expression condition, AmiS failed to express in the soluble fraction. Hence, attempts were made to solubilise the AmiS protein.

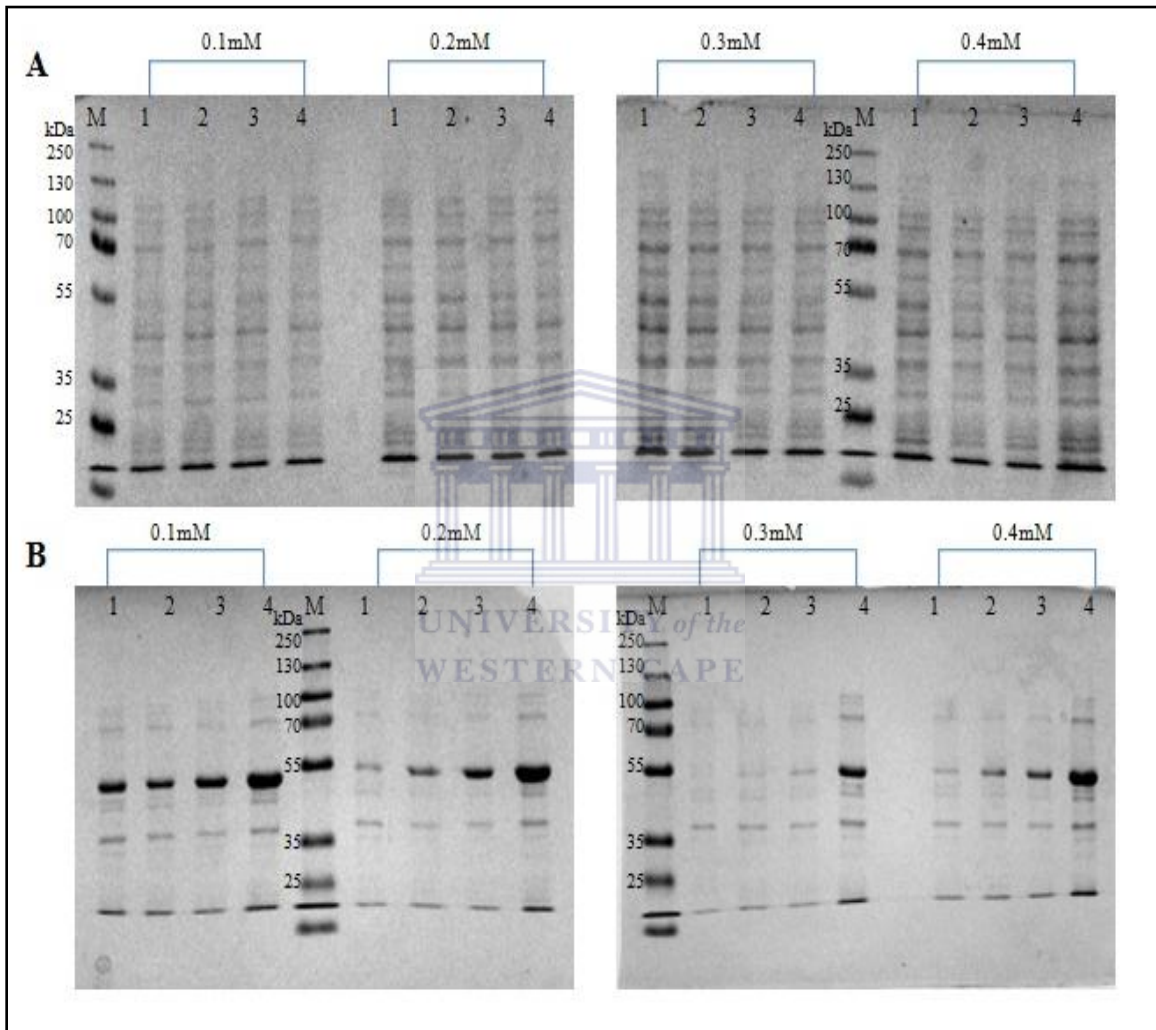
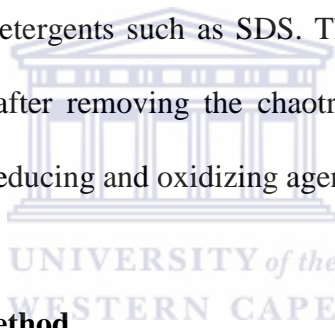


Figure 4-4 Optimisation of IPTG concentration for AmiS

A: Soluble Fractions, B: Insoluble Fractions, 1, 2, 3 and 4 represents period of time in hours the protein was expressed. AmiS only expressed in insoluble fractions only.

4.3.4 Solubilisation of Ami S

Inclusion bodies are protein aggregates commonly occurring when a recombinant targeted gene is over-expressed. It is generally assumed that high level expression of non-native, highly hydrophobic proteins are more susceptible to formation of the inclusion bodies (Bennion & Daggett, 2003). Proteins with disulphide bonds may also form inclusion bodies since the formation of the disulphide bonds are inhibited by the reducing environment of the bacterial cytosol (Singh & Panda, 2005). Inclusion bodies are typically solubilised by the use of high concentration of chaotropic agents such as urea, guanidine hydrochloride and detergents such as SDS. The solubilised proteins are then refolded to their native state after removing the chaotropic agents other by dialyzing against the buffers containing reducing and oxidizing agents (Singh & Panda, 2005).



4.3.4.1 Urea Solubilisation Method

Urea is a chemical denaturant which is used primarily to assess protein stability, the effects of mutations on stability and protein unfolding (Bennion & Daggett, 2003). Urea is also often used to solubilise proteins when they are present as inclusion bodies and were used to the solubilisation of AmiS.

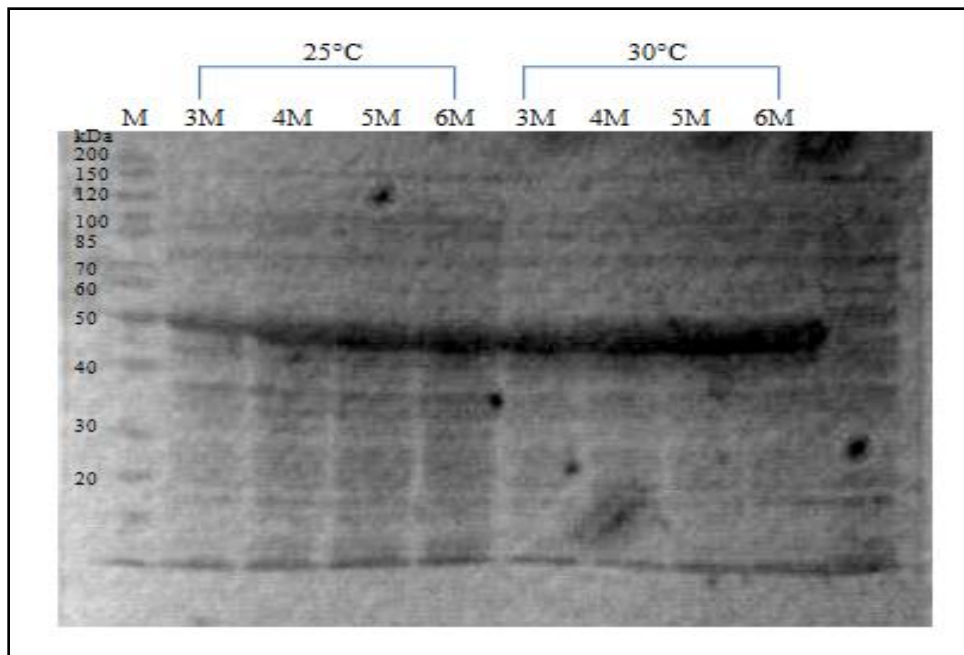


Figure 4-5 SDS-PAGE of Solubilised of AmiS by Urea

M: Marker, 3M, 4M, 5M and 6M: Concentrations of Urea used, 25°C and 30°C: Temperatures used for induction. The cultures were incubated at 37°C until O.D was 0.4. IPTG was added then the cultures were incubated at above mentioned temperature for 4hrs. The solubilisation using the urea was successful at all the concentrations used. The higher concentration of urea released larger amount of protein from the inclusion bodies.

UNIVERSITY of the
WESTERN CAPE

Figure 4-5 shows the results of solubilisation by urea. As the concentration of the urea increases, the amount of solubilised protein was increased. The AmiS was induced at 25 and 30°C. More protein was present when the protein was expressed at 30°C, hence resulted in solubilisation of larger amount of protein. In addition, the higher concentration of urea was used, more protein was solubilised (Figure 4-5).

Solubilisation of proteins using high concentration of chaotropic reagents results in the loss of secondary structure, causing proteins to form random coil structures (Singh & Panda, 2005). Hence, at lower concentrations of the chaotropic reagent, the more of

native protein structure is retained and the refolding the protein may be simpler. To solubilise the protein using the minimum amount of chaotropic agent, the lowest concentration of urea (3M) was used first. Although the protein was solubilised using 3M urea, the solubilised protein blocked the His-Tag purification column on loading (data not shown). Even after the protein sample was filtered through the 0.45µm filter, the blocking of the column persisted.

The protein was then solubilised with a high concentration of urea (6M) with dithiothreitol (DTT), a reducing agent, added to the denaturing solution to maintain cysteine residues in a reduced state so as to prevent non-native intra- or inter- disulphide bond formation in highly concentrated protein solutions at alkaline pH (Singh & Panda, 2005). However, DTT interferes with His-Tag purification so the protein samples were dialysed against the dialysis buffer to remove DTT before the sample was loaded to the column. The solubilisation and the purification was found to be successful (Figure 4-6).

The proteins with the His-Tag sequence bind to Ni²⁺ cations that are immobilised on the His-Bind resin. After unbound proteins are washed away, the target protein is recovered by elution with imidazole. Figure 4-6 shows the His-Tag purification steps of AmiS solubilised by 6M urea. The flow through fraction shows the unbound protein including a small amount of AmiS. The target protein was not only eluted in the elution fraction but also found in washing buffer fraction. Binding buffer, washing buffer and elute buffer contain 5mM, 60mM and 1M imidazole respectively.

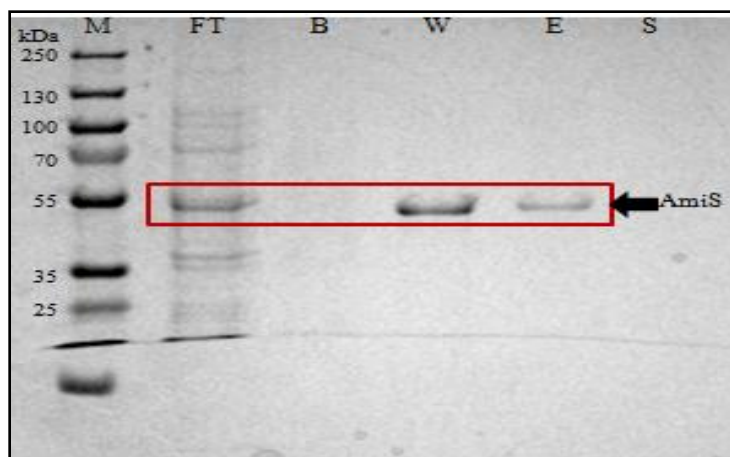


Figure 4-6 His-Tag purification of AmiS solubilised by 6M urea

M: Marker, FT: Flow Through, B: Binding Buffer Fraction, W: Washing Buffer Fraction, E: Elution, S: Strip Buffer Fraction. Some of the target protein did not bind to the column and some of the bound proteins were released by the washing buffer. Small amount of the target protein was collected in a elute fraction.

Although this result indicates that 60mM imidazole was enough to release the target protein, other purification trials showed that with washing buffer, large amounts of other unbound proteins were released with the target protein, whereas in the eluted fraction less undesirable proteins were released with the target protein (data not shown). Therefore further purification steps were carried out with 1M imidazole. Both the washing buffer fraction and the elute buffer fraction were dialysed against the dialysis buffer for 2 days to refold the protein into its native form. Since both proteins pI was lower than 7.0 (pI of Ami F: 4.8 and AmiS: 5.04, calculated by DNAMAN), the basic native gel protocol was used. However, when the refolded protein was eletrophoresed on a native-PAGE gel, all the proteins appeared to be remained in the well (Figure 4-7). The separation of proteins on the native gels depends on hydrodynamic size, shape and native charges. Since the

charges were taken into the considerations already, it was deduced that the reason the samples remained in the well was due to the aggregation of the proteins.

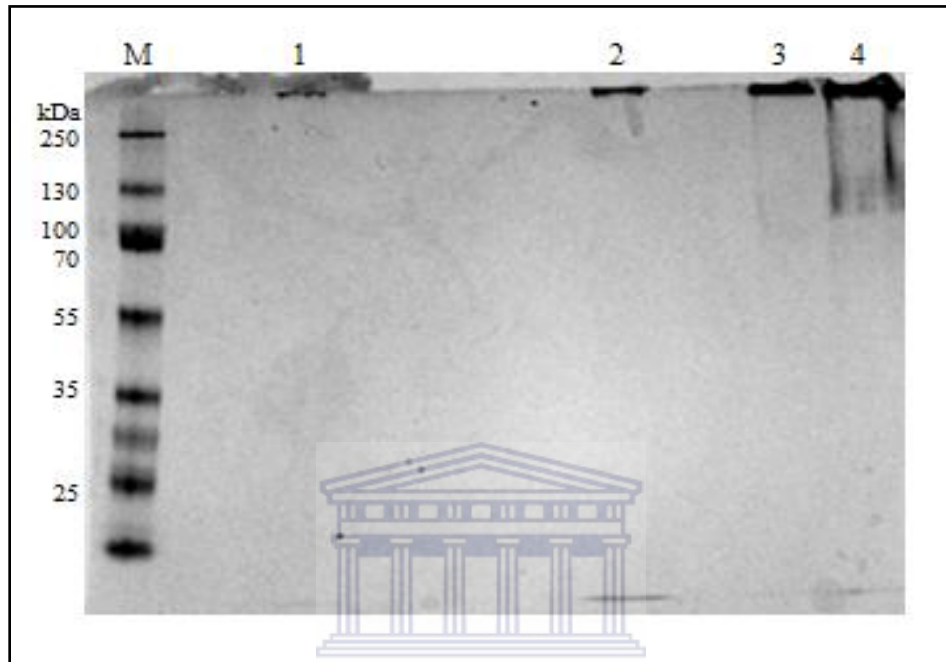


Figure 4-7 Native-PAGE of refolded AmiS

M: Marker, 1: 10ul, 2: 20ul and 3: 30ul of refolded AmiS, 4: cell free extract. Protein samples remained at the well

This approach was repeated using 8M and 10M urea. The protein was successfully solubilised, but each time failed to refold. The observed aggregation could be the results of the rapid refolding method. Protein aggregation is a higher order reaction while refolding is a first order reaction, favouring the aggregation of solubilised protein rather than refold into the native form (Singh & Panda, 2005). When the high concentration of

chaotropic agents is removed rapidly, the proteins have insufficient time to refold properly. Therefore the slower refolding method was also attempted.

4.3.4.2 Two Step Refolding Method (Urea based)

The AmiS was expressed and solubilised as outlined in section 4.2.2. The solubilised protein was dialysed against the dialysis buffer containing 3M urea for 12hrs. The buffer was changed to fresh buffer without any urea and the protein was dialysed for further 24hrs. The refolded protein was subjected to the His-Tag purification (Figure 4-8).



Figure 4-8 His-Tag steps of two step refolding of AmiS solubilised by 6M urea

M: Marker, 1: Soluble fraction, 2: Pellet washing buffer, 3: Insoluble fraction, 4: Flow through fraction, 5: Binding buffer fraction, 6: Washing buffer fraction, 7: Elution, 8: Strip buffer fraction. Protein was expressed in the insoluble fraction (lane 3) only. Most of the protein did not bind to the column (lane 4) and small amount of the protein was released in the elute fraction (lane 7).

The purification was successful, although a large amount of AmiS was lost. The eluted fraction (lane 7 from Figure 4-8) was run on the native-PAGE to confirm the successful refolding. Figure 4-9 shows that even when AmiS was refolded in a slower manner, the aggregation problem was not solved.

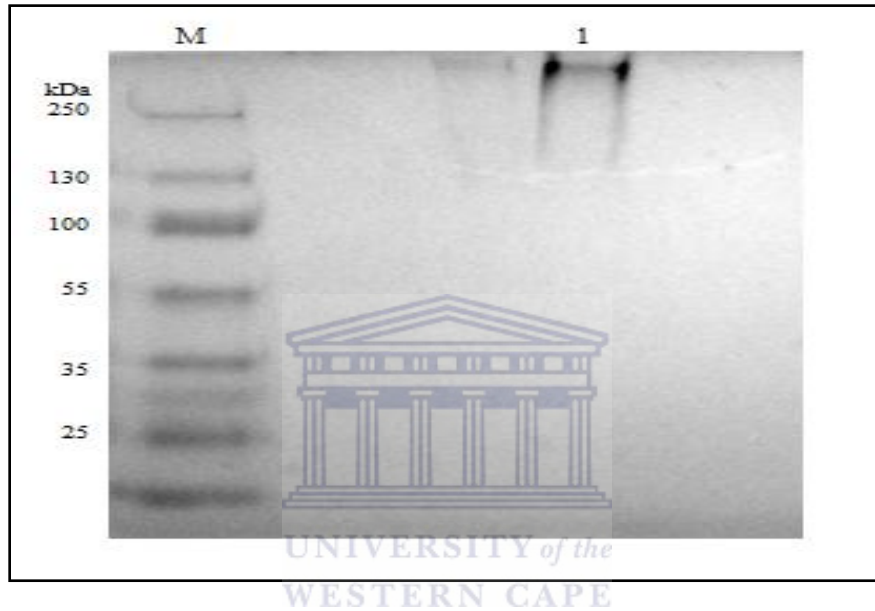


Figure 4-9 Native-PAGE gel of AmiS refolded by two step refolding method

M: Marker, 1: Refolded AmiS. The protein was aggregated after the slow refolding.

4.3.4.3 Chloramphenicol Method

It has been shown that the formation of the inclusion bodies is the results of unbalanced equilibrium between protein aggregation and solubilisation (Carrió & Villaverde, 2001).

It has also been found that the equilibrium can be spontaneously displaced toward protein refolding when protein synthesis is arrested. A high concentration of chloramphenicol (up

to 200 μ g/ml) was added to arrest the protein synthesis after overnight expression at 30°C. The culture was then incubated for 2 hrs to allow the proteins to refold properly.

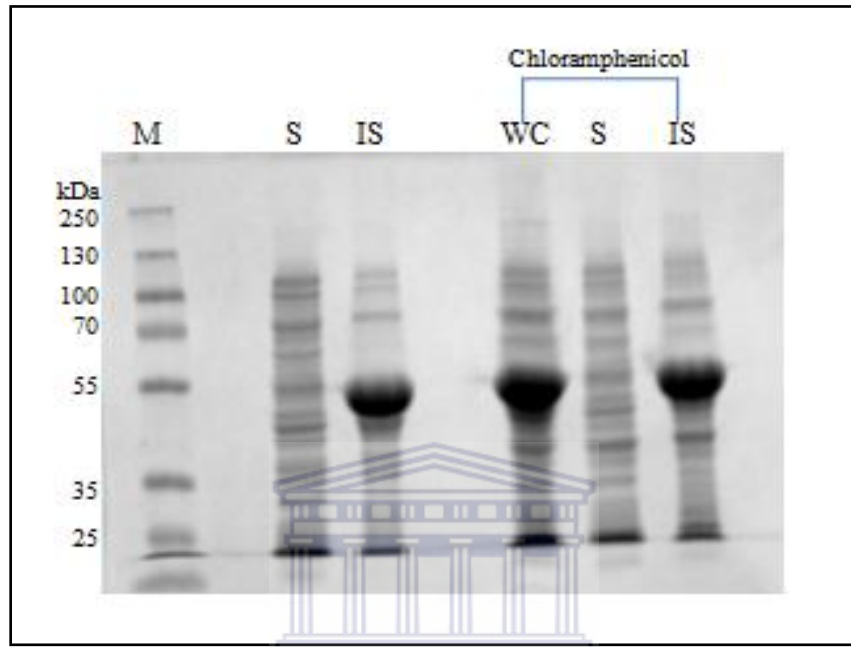


Figure 4-10 Solubilization of AmiS by Chloramphenicol Method

M: Marker, S: Soluble fraction, IS: Insoluble fraction, WC: whole cell. The soluble and the insoluble fraction were expressed using the standard method for the comparison. Using the chloramphenicol method did not improve the expression of the target protein.

Figure 4-10 shows the comparison of two expression methods. The soluble and insoluble fractions were obtained after the expression following the method outlined in section 4.2.4 and the whole cell extract, soluble and insoluble fraction were collected following the method describe in Villaverde *et al.*,(2001). Although arrest of protein synthesis has been shown to result in almost complete inclusion body disintegration (Carrió &

Villaverde, 2001), this method did not assist in solubilising AmiS. The Figure 4-10 shows that there was no conversion of the inclusion bodies to the soluble form of AmiS.

4.3.4.4 β -mercaptoethanol Method

A new method has been developed to solubilise proteins from inclusion bodies using β -mercaptoethanol. Inclusion bodies were treated with a denaturing solution consisting of mercaptoethanol solvent at a concentration of 6-8M and urea at a concentration of 1-2M (Panda *et al.*, 1997). The solubilised protein was dialysed against a refolding buffer to obtain a protein in its bioactive form.

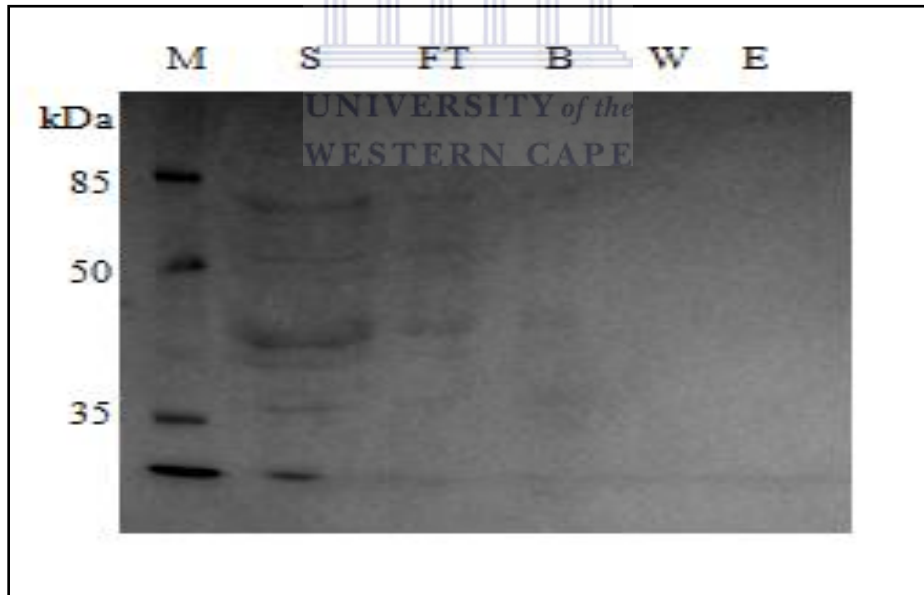


Figure 4-11 His-Tag Purification of the solubilised protein by beta-mercaptoethanol

M: Marker (Fermentas #SM441), S: Solubilised Protein, FT: Flow Through, B: Binding Buffer Fraction, W: Washing Buffer Fraction, E: Elution. The lane S shows that the solubilisation of the target protein using the mercaptoethanol was not successful.

Figure 4-11 shows the solubilised protein fraction and His-Tag purification steps. The protein band in lane S demonstrates that the combination of beta-mercaptoethanol and urea denaturing solution did not solubilise AmiS from the inclusion bodies. When the solubilised fraction that was applied to the His-Tag column, no protein in the eluted fraction was observed, suggesting that AmiS was not present in the solubilised fraction.

The theory of this method is that the high concentration of mercaptoethanol in combination with a lower concentration of urea would assist in disrupting disulphide bonds and hydrophobic interaction of the proteins. In the past, high concentration of Urea and very low concentration of mercaptoethanol (10-20mM) were used for the solubilisation. This method has some potential advantages. Firstly, the combination assists solubilisation without unfolding the protein completely, hence resulting in higher recovery of the bioactive form of the protein. Secondly, the low aggregation during refolding of the proteins also helps with a higher recovery rate. Thirdly, the solubilisation of the inclusion body is done without using high concentration of chaotropic solvent which completely unfolds the protein molecules into random structures. Although this method did not work for this instance, it has been proven to be effective for other proteins such as human growth hormone, polyketide synthase and enolase (Panda *et al.*, 1997).

4.3.4.5 Purification of Inclusion Bodies using Na. Deoxycholate

Inclusion bodies of AmiS were successfully purified using sodium deoxycholate (DOS). The purified inclusion bodies were solubilised using a low concentration of chaotropic agent (2M urea) at alkaline pH (pH 12.5).

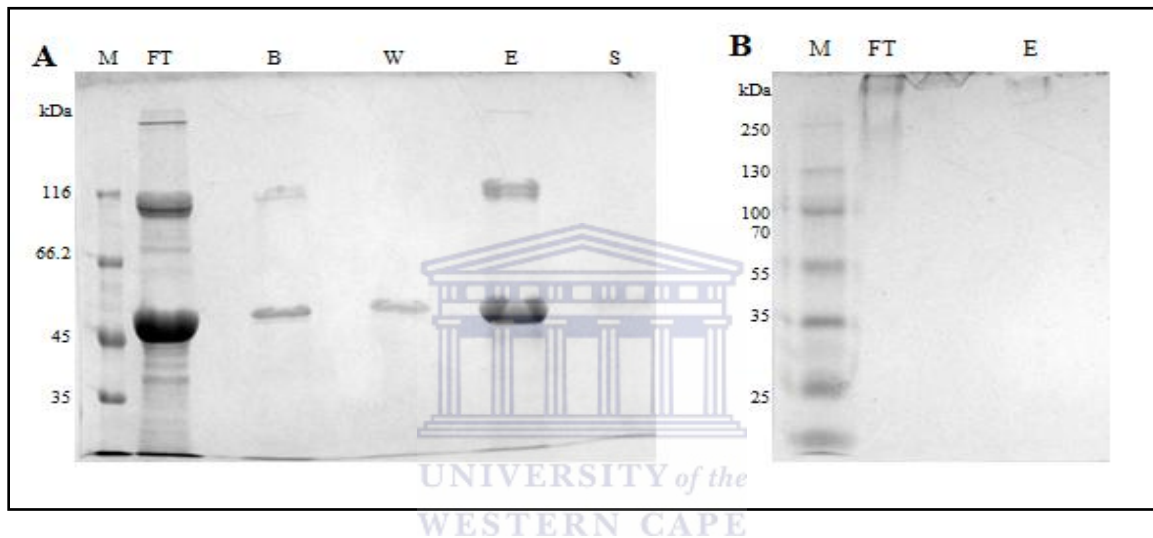


Figure 4-12 SDS-PAGE analysis of inclusion body purification and solubilisation of AmiS using DOS

A: His-Tag purification steps, B: Native-PAGE of solubilised AmiS

FT: Flow through fraction, B: Binding buffer fraction, W: Washing buffer fraction, E: Elution, S: Strip buffer fraction

Although large amount of protein did not bound to the column possibly due to overloading the column with proteins (FT lane of panel A), sufficient amount of protein was purified as it is shown in lane E, panel A. The proteins from elution fraction was refolded and analyzed on Native-PAGE (panel B) which shows that the majority of the protein did not refold correctly.

The flow through fraction (lane FT on panel A, Figure 4-12) shows the results of the purification of the inclusion bodies. When this lane is compared to the flow through

fraction (lane FT) from Figure 4-6, it can be noted that there was less background protein after purification. The His-Tag purification of solubilised protein under the denaturing condition (2M urea) was successful. The elution fraction and the flow through fraction were dialysed against the standard dialysis buffer. When the dialysed proteins were electrophoresed on Native-PAGE, the refolding seemed to be largely unsuccessful. However, a faint band was seen which could suggest that very little AmiS had refolded correctly. The purified AmiS fractions were pooled and subjected to the activity assays.

When the cell pellets are collected after the induction, a high level of background proteins is evident along with the over-expressed target protein. These impurities have been shown to interfere with subsequent steps such as solubilisation and refolding, i.e. human growth hormone (Singh & Panda, 2005). The purification of the inclusion bodies could also help with the removal of high molecular weight aggregates (Singh & Panda, 2005), hence increasing the recovery of bioactive proteins. Use of 2M urea generally does not unfold the protein completely, and preserve the native-like secondary structure (Singh & Panda, 2005) while the use of a high pH buffer also assists solubilisation as long as the isoelectric point of the target protein is further away from the pH of the buffer being used creating pH shock effect. A combination of 2M urea and pH shock destabilises both ionic and hydrophobic interactions which are the major cause of protein aggregation (Singh & Panda, 2005).

The purification of the inclusion bodies did improve the refolding of AmiS as a small amount of the protein appeared refolded.

4.3.4.6 Effect of His-Tags on Solubility of AmiS

In studies aimed at establishing the effects of His-Tag on protein solubility, 24 proteins were expressed *in vitro* with both N- and C-terminal His-Tags (Busso *et al.*, 2003). The results indicated that His-Tag affects solubility of some target proteins but not all (Busso *et al.*, 2003). In some proteins, solubility was enhanced by the presence of N-terminal His Tag while other proteins were more soluble when the N- or C- His-Tags were removed (Woestenenk *et al.*, 2004). Since AmiS was provided with a N-terminal His-Tag, it was sub-cloned into pET 17 b to remove His-Tag from the protein. However, the removal of the His-Tag did not increase the solubility of the AmiS which still expressed in the insoluble fraction (data not shown).

4.3.4.7 Co-factors

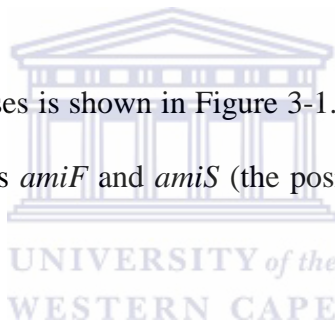
Previous studies have shown, the many proteins do not refold properly without the presence of their co-factors (d'Abusco *et al.*, 2001; Pabarcus & Casida, 2002). In this study, only Mg²⁺ was added to the dialysis buffer to determine the effect on the refolding of the AmiS protein. Mg²⁺ did not improve the refolding process or solve the aggregation problem (data not shown). Although there are many other co-factors available to be

tested, it was not pursued due to the fact that few studies done on amidases have shown some co-factors increase the activity but they are not necessary for their activity or correct refolding and also some co-factors inhibited the enzyme activity (d'Abusco *et al.*, 2001; Pabarcus & Casida, 2002).

4.3.5 Reverse Transcription PCR

The spatial orientation and the intergenic distance (1bp overlap) of *amiF* and *amiS* suggested that these two ORFs could possibly be co-transcribed.

The genetic map of two amidases is shown in Figure 3-1. Two primers were designed to amplify a 500bp product across *amiF* and *amiS* (the position of the primers is indicated by red arrows on the figure).



RNA was successfully extracted from *Nesterenkonia* sp. using the hot phenol method described in section 4.2.6.

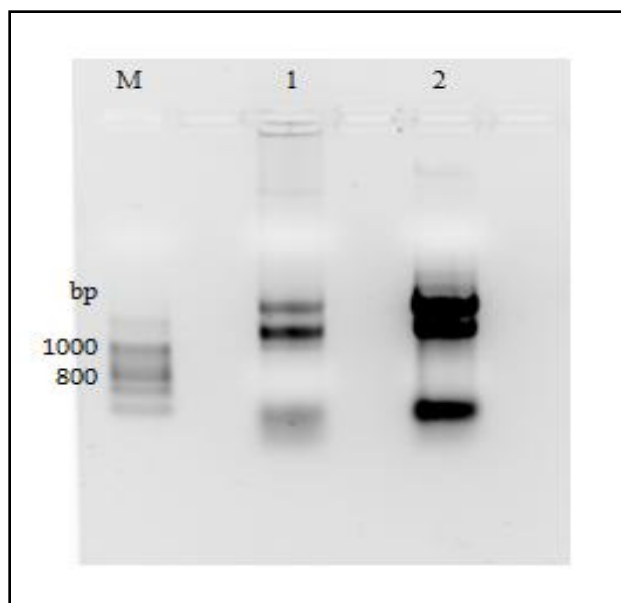


Figure 4-13 Analysis of RNA extraction on an agarose gel

M: RNA marker, 1: RNA from *Nesterenkonia* sp. 2: RNA from *E. coli*

Although the secondary structure of RNAs prevent it from migrating strictly according to its size on the agarose gel, it is sufficient to make rough assessment of the integrity of RNA samples. This gel shows the intact RNAs from both *Nesterenkonia* sp. and *E. coli*. The 18S and 28S rRNA is clearly visible and distinguishable. Although comparably less amount of RNA was extracted from *Nesterenkonia* sp. (lane 1) it was enough to carry out the subsequent experiments.

UNIVERSITY of the
WESTERN CAPE

Figure 4-13 shows evidence of intact RNA from both *Nesterenkonia* sp. and *E. coli*. The RNA marker was too degraded to distinguish the size of the bands. PCR reactions were carried out on RNA with 16S rRNA gene primers and primers designed for reverse transcription PCR (namely, RTF and RTR). The 16S PCR results showed that there was DNA present to be amplified (Figure 4-14, lane 3). RNA was therefore treated with DNase. The 16S rRNA PCR was carried out once again to confirm the removal of DNA (Figure 4-15, lane 4).

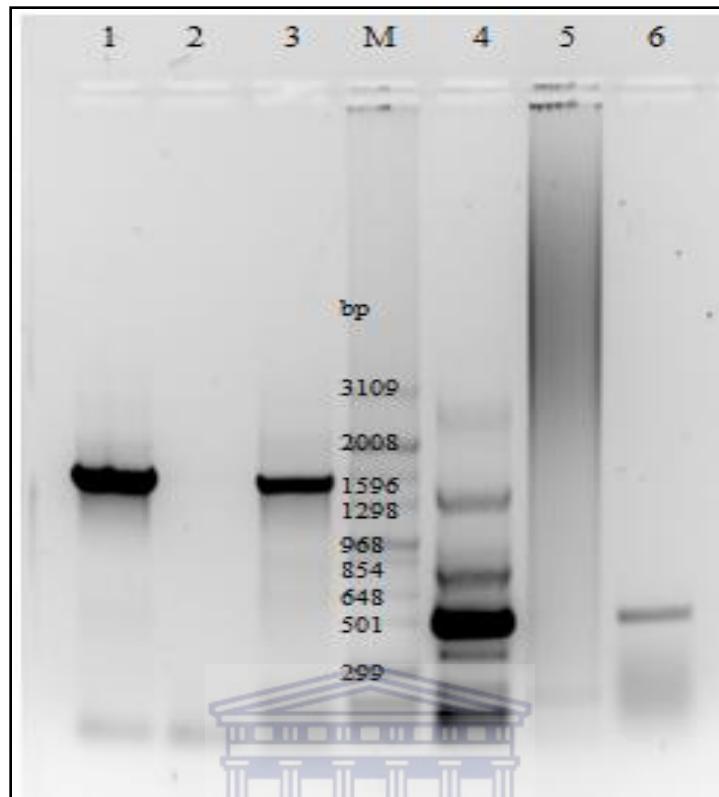


Figure 4-14 PCR on RNA before DNase treatment

M: Marker, 1: 16S positive control, 2: 16S negative control, 3: 16S RNA, 4: RT primers positive control, 5: RT primers negative control, 6: RT primers on RNA. PCR reactions were set up using 16S and RT primers before RNA was treated with DNase to confirm the presence of the DNA in the sample. Both reactions gave the positive results indicating that there was DNA contamination in RNA sample.

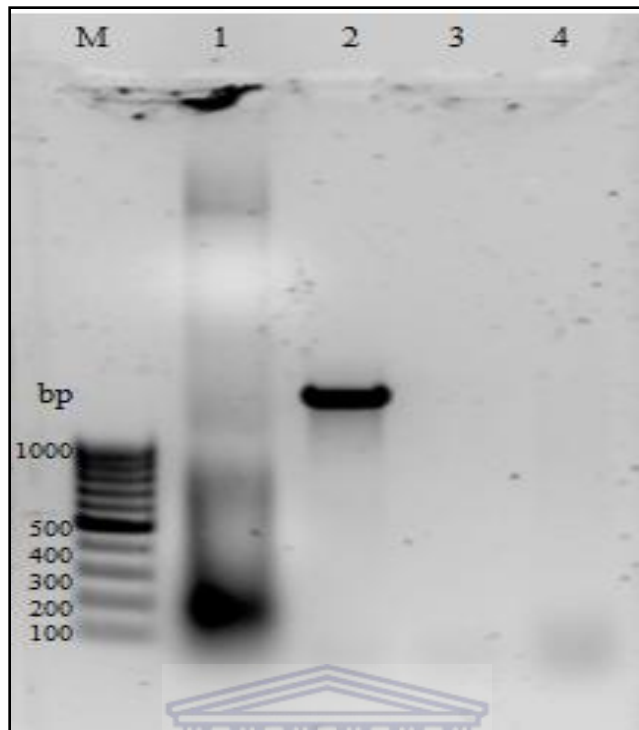


Figure 4-15 PCR reactions on RNA after DNase treatment

M: DNA Marker, 1: DNase treated RNA, 2: 16S rRNA positive control, 3: 16S rRNA negative control, 4: 16S rRNA on DNase treated RNA. After the DNase treatment on RNA sample, the 16S PCR was carried out to confirm the removal of the DNA from the sample. The results of the PCR indicated the complete removal of DNA from the sample.

WESTERN CAPE

Figure 4-15 shows the RNA after DNase treatment (lane 1) and the results of the 16S rRNA PCR (lane 4). Considering that the same concentration of RNA was used for PCR before and after DNase treatment, these results suggest that DNA was effectively degraded by DNase treatment. This RNA preparation was subjected to the reverse transcription PCR.

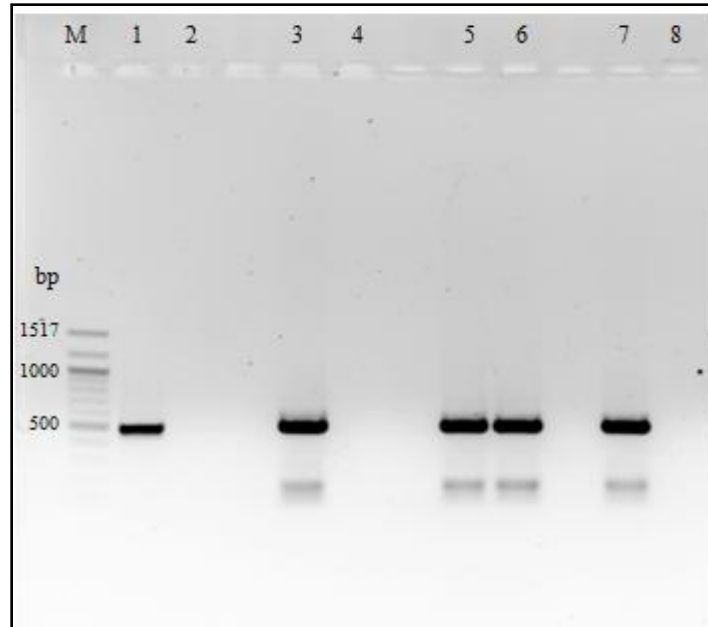


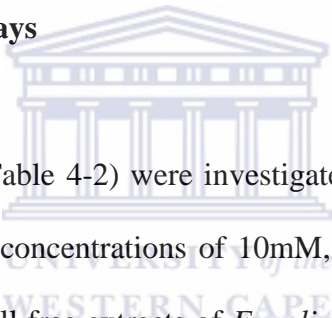
Figure 4-16 Reverse transcription PCR

M: Marker, 1: + control from cDNA kit (496bp), 2: + Avian Myeloblastosis Virus (AMV) reverse transcriptase, RT PCR with no template, 3: + control for PCR (500bp), 4: - control for PCR, 5: +AMV, PCR with RNA no DNase, 6: -AMV, PCR with RNA no DNase, 7: +AMV, PCR with DNase treated RNA, 8: -AMV, PCR with DNase treated RNA

The positive control provided in the cDNA synthesis kit was included in the reaction. This reaction amplified a 496bp product (lane 1), confirming the RT reaction was successful. A negative control was set up to ensure that all the reagents provided in the kit were not contaminated; this reaction did not give any amplicon bands (lane 2). The amplification with the RTF and RTR primers was successful (lane 3) and showed no sign of contamination in the PCR reagents used (lane 4). The RNA sample that was not treated with DNase amplified 500bp product for the positive and the negative reverse transcriptase reaction (lanes 5 & 6). This result clearly indicates that there was genomic DNA present in the RNA sample before the DNase treatment. To confirm the absence of

genomic DNA contamination in the RNA sample after the DNase treatment, a reverse transcriptase minus control reaction was set up. The absence of amplified bands (lane 8) indicated that the sample was not contaminated. When the DNase treated RNA was used for the RT reaction, only the reverse transcriptase positive sample amplified a 500bp product (lane 7). This result confirmed that the two genes are co-transcribing in their natural host and therefore suggests that the enzymes may be functionally linked. A northern blot can be done on the RNA sample to observe the expression patterns of two amidase genes.

4.3.6 Ammonia Detection Assays



A range of amide substrates (Table 4-2) were investigated in order to establish enzyme substrate specificity. Substrate concentrations of 10mM, 25mM, 50mM and 100mM were used for the assay. The cell free extracts of *E. coli* cultures expressing the proteins, partially purified AmiF and refolded AmiS were used. None of the samples showed any detectable activity. Given the fact that two ORFs might be functionally linked, a mixed sample of AmiF and AmiS was also tested for activity. None of the assays showed any enzyme activity.

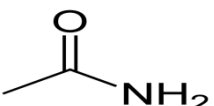
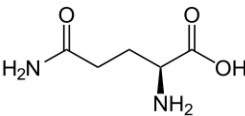
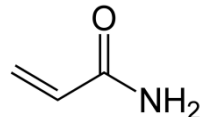
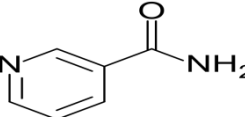
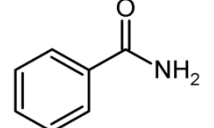
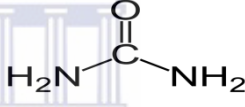
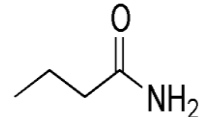
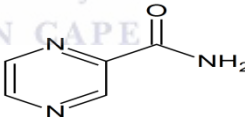
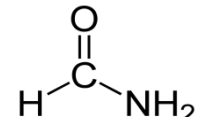
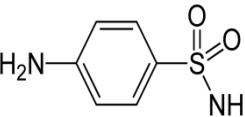
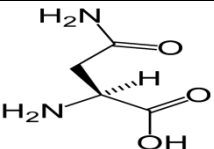
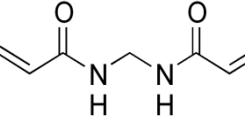
As a positive control for the assay, a well-studied amidase from *Geobacillus pallidus* was used. Although this amidase is known to be active only on short chain amides substrates, hence acetamide was used for the assay as it was one of the preferred substrate of the

enzyme (Makhongela *et al.*, 2007). The amidase had K_M of 20mM which indicated that the assay itself was working properly (data not shown).

The control assay confirmed that the assay system was functional. The absence of apparent activity in AmiF and AmiS could be due to amino acid substitution, incorrect folding of the proteins, inhibition of the enzyme from the conditions used (pH range, concentration of salts, temperatures) and the absence of co-factors as some co-factors are known to improve the activity of amidases (d'Abusco *et al.*, 2001). Further studies to optimise the solubilisation process and the assay conditions would be required to identify the possible reasons for the absence of activity for both proteins.



Table 4-2 Amidase substrates used for ammonia detection assay

Substrate Structure	Name	Substrate Structure	Name
	Acetamide		L-Glutamine
	Acrylamide		Nicotinamide
	Benzamide		Urea
	Butyramide		Pyrazinamide
	Formamide		Sulfanilamide
	L-asparagine		N,N'- Methylene bisacrylamide

Chapter 5

General Discussion

The *in silico* gene mining of the partially sequenced genome of *Nesterenkonia* sp. led to the identification of two amidase genes, *amiF* and *amiS*. Both proteins belonged to amidohydrolases family. AmiF was assigned to the aceta/formamidase subfamily while AmiS belonged to the signature amidase family, and shared the invariant GGSS sequence as well as the novel catalytic residues, Ser-*cis*Ser-Lys.

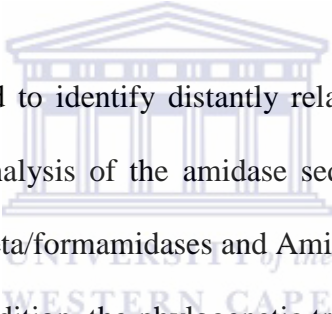
Two genes (*amiF* and *amiS*) were cloned into pET 28 a (+) and expressed in Rosetta 2 (DE3) pLysS. Majority of AmiF was expressed in the soluble fraction and successfully purified. AmiS was expressed in an insoluble form. Solubilisation of the protein was successful using various chaotropic reagents but refolding of the denatured AmiS to its native form was largely unsuccessful.

In sequence analysis of the two genes, both the spatial orientation and the intergenic distance (1bp overlap) suggested that these two ORF's could possibly be co-transcribed. Primers were designed to amplify a 500bp product across the *amiF* and *amiS* for reverse transcription PCR. RNA was successfully extracted from *Nesterenkonia* sp. using the hot

phenol method. The reverse transcription PCR successfully amplified a 500bp product indicating that these two genes are co-transcribed in their natural host and two genes may be functionally linked. Further sequence analysis revealed the partial sequence of another member of amidohydrolases family ~500bp downstream from *amiS*, indicating the possible presence of an amidase ‘operon’ (Figure 3-1). Further reverse transcription PCR could be performed to identify the extent of co-transcription of these amidases. In addition, completion of the whole genome sequence would allow a more accurate identification of the third gene as well as possible identification of other amidases from upstream and downstream of *amiF* and *amiS*.

The partially purified AmiF and a small amount of refolded AmiS were subjected to the amidase activity assays. Twelve different amide substrates at four different concentrations were used to test the activity of the enzymes. The pH range and the temperatures used were based on previous studies done on different but related amidases from the same organism (Nel, 2009). No positive assay results were obtained. Since it has been established that the two genes co-transcribe, it is possible that two gene products may have to present together to have activity. In this study, although the mixing of two amidases for the activity assay was attempted, the proper refolding of AmiS was not confirmed. In further studies, the confirmation of the expression of AmiS as well as AmiF should be done to make sure the over-expressed proteins are AmiF and AmiS. This

could be done using either western blot or 6x His protein tag staining kit. In this study, 6x His protein tag staining kit from Pierce was used to confirm the target proteins. However, the kit stained neither the control protein provided by the kit nor AmiF and AmiS. This could be due to many reasons including the age of the reagents, possible degradation of the control protein. Hence, it would be advisable to repeat this experiment with new reagents and fresh control protein. Western blotting with His-tag antibody would be equally effective. In addition, further studies of optimisation of expression and/or refolding of AmiS would be required for the better assessment of the enzyme activity.



Protein sequences can be used to identify distantly related proteins and their possible functions. The phylogenetic analysis of the amidase sequences confirmed the BLAST results. AmiF was related to aceta/formamidases and AmiS was related to the members of signature amidase family. In addition, the phylogenetic tree also indicated that two genes are not related to each other. The structure of a protein also provides insights into the proteins' catalytic residues and reaction mechanisms. Homology models of AmiF and AmiS were built using MODELLER 9v7 (Sali & Blundell, 1993) and side-chains were optimised by SCWRL 4.0 (Bower *et al.*, 1997). Although the templates chosen for both proteins showed low sequence identities (24% for AmiF and 25% for AmiS), which lowers the accuracy and the reliability of the models, the evaluation of the models indicated that both models were acceptable. Ramachandran plots showed that residues

lying outside of the allowed region were all in loop regions which do not affect the catalytic residues or mechanisms. No comprehensive studies have been reported on the members of the aceta/formamidase family, and the residues involved in the catalytic mechanism are not known. However, the alignment between the template and the AmiF showed the conserved regions which could be involved in the enzyme catalysis.

The sequence of AmiS showed all the conserved sequences of the signature amidase family share. The model showed conservation of the catalytic triad residues Ser-*cis*Ser-Lys that are absolutely conserved throughout the family. The conserved sequences of AmiS mostly form the core structure and to compensate the unusual *cis* conformation, *cis*Ser is located in the region enriched with small residues such as glycine, serine and alanine. A new catalytic mechanism has been proposed to accommodate the *cis*Ser conformation (Shin *et al.*, 2003). Although all three residues directly involved in the mechanism showed both sequence and position conservation, one of the residue (arginine) that is thought to contribute to catalysis was substituted by valine. This substitution was not a mutation caused by subsequent PCR of the gene as the whole genome sequence result showed the valine residue in the same position. The substitution could be the reason for inactivity of the enzyme as the arginine is involved in a dipolar interaction which maintains the serine residue in a deprotonated state. In a further study, a Val to Arg site specific mutagenesis would establish the importance of this residue. This

family of enzymes appears to be evolutionarily distinct but has diverged to acquire a wide spectrum of individual substrate specificities, while maintaining a core structure that supports the catalytic function of the unique triad (Shin *et al.*, 2003).



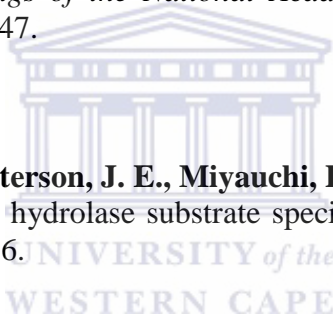
References

Baker, D. & Sali, A. (2001). Protein structure prediction and structural genomics. *Science* **294**, 93-96.

Banerjee, A., Sharma, R. & Banerjee, U. C. (2002). The nitrile-degrading enzymes: current status and future prospects. *Applied Microbiology and Biotechnology* **60**, 33-44.

Bennion, B. J. & Daggett, V. (2003). The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5142-5147.

Boger, D. L., Fecik, R. A., Patterson, J. E., Miyauchi, H., Patricelli, M. P. & Cravatt, B. F. (2000). Fatty acid amide hydrolase substrate specificity. *Bioorganic & Medicinal Chemistry Letters* **10**, 2613-2616.



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

Bower, M. J., Cohen, F. E. & Dunbrack, R. L. (1997). Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *Journal of Molecular Biology* **267**, 1268-1282.

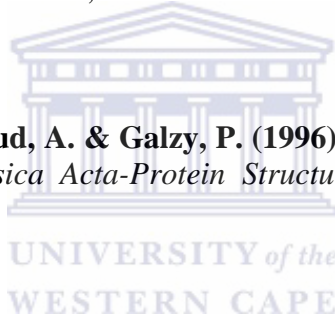
Brenner, C. (2002). Catalysis in the nitrilase superfamily. *Current Opinion in Structural Biology* **12**, 775-782.

Busso, D., Kim, R. & Kim, S. H. (2003). Expression of soluble recombinant proteins in a cell-free system using a 96-well format. *Journal of Biochemical and Biophysical Methods* **55**, 233-240.

Cameron, R. A., Sayed, M. & Cowan, D. A. (2005). Molecular analysis of the nitrile catabolism operon of the thermophile *Bacillus pallidus* RApC8. *Biochimica Et Biophysica Acta-General Subjects* **1725**, 35-46.

Carrió, M. M. & Villaverde, A. (2001). Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Letters* **489**, 29-33.

Chebrou, H., Bigey, F., Arnaud, A. & Galzy, P. (1996). Study of the amidase signature group. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1298**, 285-293.



Cowan, D. A., Cameron, R. A. & Tsekoo, T. L. (2003). Comparative biology of mesophilic and thermophilic nitrile hydratases. *Advances in Applied Microbiology, Vol 52* **52**, 123-158.

Curnow, A. W., Hong, K. W., Yuan, R., Kim, S. I., Martins, O., Winkler, W., Henkin, T. M. & Soll, D. (1997). Glu-tRNA(Gln) amidotransferase: A novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 11819-11826.

d'Abusco, A. S., Ammendola, S., Scandurra, R. & Politi, L. (2001). Molecular and biochemical characterization of the recombinant amidase from hyperthermophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* **5**, 183-192.

D'Amico, S., Gerday, C. & Feller, G. (2001). Structural determinants of cold adaptation and stability in a large protein. *Journal of Biological Chemistry* **276**, 25791-25796.

Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Research* **32**, W615-619.

Eric, F. P., Thomas, D. G., Conrad, C. H., Gregory, S. C., Daniel, M. G., Elaine, C. M. & Thomas, E. F. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612.

Farnaud, S., Tata, R., Sohi, M. K., Wan, T., Brown, P. R. & Sutton, B. J. (1999). Evidence that cysteine-166 is the active-site nucleophile of *Pseudomonas aeruginosa* amidase: crystallization and preliminary X-ray diffraction analysis of the enzyme. *Biochemical Journal* **340**, 711-714.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783-791.

Fiser, A., Sali, A. & Charles W. Carter, J. a. R. M. S. (2003). Modeller: Generation and refinement of homology-based protein structure models. In *Methods in Enzymology*, pp. 461-491: Academic Press.

Fiser, A. (2004). Protein structure modeling in the proteomics era. *Expert Review of Proteomics* **1**, 97-110.

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

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

Fournand, D., Bigey, F. & Arnaud, A. (1998b). Acyl transfer activity of an amidase from Rhodococcus sp. strain R312: Formation of a wide range of hydroxamic acids. *Applied and Environmental Microbiology* **64**, 2844-2852.

Fournand, D. & Arnaud, A. (2001). Aliphatic and enantioselective amidases: from hydrolysis to acyl transfer activity. *Journal of Applied Microbiology* **91**, 381-393.

Gerday, C., Aittaleb, M., Arpigny, J.L., Baise, E., Chessa, JP., Garsoux, G., Petrescu, I and Feller, G et al (1997). Psychrophilic enzymes: a thermodynamic challenge. *Biochimica et Biophysica Acta* **1342**, 119-131.

Hansen, H. S., Lauritzen, L., Moesgaard, B., Strand, A. M. & Hansen, H. H. (1998). Formation of N-acyl-phosphatidylethanolamines and N-acylethanolamines - Proposed role in neurotoxicity. *Biochemical Pharmacology* **55**, 719-725.

Jones, D. T. (1999). GenThREADER: an efficient and reliable protein fold recognition method for genome sequences. *Journal of Molecular Biology* **287**, 797-815.

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

Kato, Y., Ooi, R. & Asano, Y. (2000). Distribution of aldoxime dehydratase in microorganisms. *Applied and Environmental Microbiology* **66**, 2290-2296.

Kimani, S. W., Agarkar, V. B., Cowan, D. A., Sayed, M. F. R. & Sewell, B. T. (2007). Structure of an aliphatic amidase from *Geobacillus pallidus* RAPc8. *Acta Crystallographica Section D-Biological Crystallography* **63**, 1048-1058.

Kobayashi, M., Fujiwara, Y., Goda, M., Komeda, H. & Shimizu, S. (1997). Identification of active sites in amidase: Evolutionary relationship between amide bond- and peptide bond-cleaving enzymes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 11986-11991.

Kobayashi, M., Goda, M. & Shimizu, S. (1998a). Nitrilase catalyzes amide hydrolysis as well as nitrile hydrolysis. *Biochemical and Biophysical Research Communications* **253**, 662-666.

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

Kobayashi, M., Goda, M. & Shimizu, S. (1999). The catalytic mechanism of amidase also involves nitrile hydrolysis (vol 439, pg 325, 1998). *FEBS Letters* **444**, 296-296.

Labahn, J., Neumann, S., Buldt, G., Kula, M. R. & Granzin, J. (2002). An alternative mechanism for amidase signature enzymes. *Journal of Molecular Biology* **322**, 1053-1064.

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

Marti-Renom, M. A., Madhusudhan, M. S., Fiser, A., Rost, B. & Sali, A. (2002). Reliability of Assessment of Protein Structure Prediction Methods. *Structure* **10**, 435-440.

McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **19**, 874-881.

Morgan-Kiss, R. M., Priscu, J. C., Pocock, T., Gudynaite-Savitch, L. & Huner, N. P. A. (2006). Adaptation and acclimation of photosynthetic microorganisms to permanently cold environments. *Microbiololgy Molecular Biology Review* **70**, 222-252.

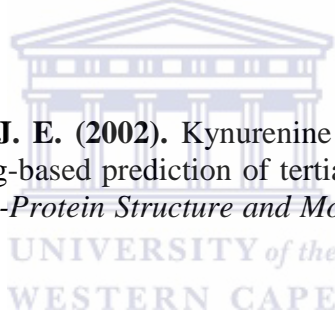
Nakamura, A., Yao, M., Chinnaronk, S., Sakai, N. & Tanaka, I. (2006). Ammonia Channel Couples Glutaminase with Transamidase Reactions in GatCAB. *Science* **312**, 1954-1958.

Nel, A. (2009). Identification of novel cold-adapted nitrile hydrolysing enzymes. In *Biotechnology*, pp. 195. Cape Town: University of the Western Cape.

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

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

Pabarcus, M. K. & Casida, J. E. (2002). Kynurenine formamidase: determination of primary structure and modeling-based prediction of tertiary structure and catalytic triad. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1596**, 201-211.



Pace, H. C. & Brenner, C. (2001). The nitrilase superfamily: classification, structure and function. *Genome Biology* **2**, 0001.0001-0001.0009.

Panda, T., Babu, P. S. R., Kumari, J. A. & other authors (1997). Bioprocess optimization - a challenge. *Journal of Microbiology and Biotechnology* **7**, 367-372.

Pertsovich, S. I. (2005). Computer modeling of the aliphatic amidase structure. *FEBS Journal* **272**, 112-113.

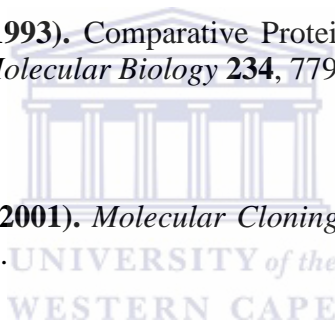
Pertsovich, S. I., Guranda, D. T., Podchernyaev, D. A., Yanenko, A. S. & Svedas, V. K. (2005). Aliphatic amidase from *Rhodococcus rhodochrous* M8 is related to the nitrilase/cyanide hydratase family. *Biochemistry-Moscow* **70**, 1280-1287.

Reysenbach, A. & Pace, N. (1995). *Archaea: a Laboratory Manual - Thermophiles*. New York: Cold Spring Harbor Laboratory.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.

Sali, A. & Blundell, T. L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **234**, 779-815.

Sambrook, J. & Ressel, D. (2001). *Molecular Cloning: A Laboratory Manual*, 3 edn: Cold Spring Harbor Laboratory.



Sánchez, R. & Sali, A. (1997a). Comparative protein structure modeling as an optimization problem. *Journal of Molecular Structure: THEOCHEM* **398-399**, 489-496.

Sánchez, R. & Sali, A. (1997b). Advances in comparative protein-structure modelling. *Current Opinion in Structural Biology* **7**, 206-214.

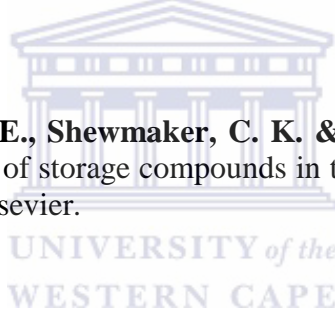
Shi, J. Y., Blundell, T. L. & Mizuguchi, K. (2001). FUGUE: Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of Molecular Biology* **310**, 243-257.

Shin, S., Lee, T. H., Ha, N. C., Koo, H. M., Kim, S. Y., Lee, H. S., Kim, Y. S. & Oh, B. H. (2002). Structure of malonamidase E2 reveals a novel Ser-cisSer-Lys catalytic triad in a new serine hydrolase fold that is prevalent in nature. *Embo Journal* **21**, 2509-2516.

Shin, S., Yun, Y. S., Koo, H. M., Kim, Y. S., Choi, K. Y. & Oh, B. H. (2003). Characterization of a novel Ser-cisSer-Lys catalytic triad in comparison with the classical Ser-His-Asp triad. *Journal of Biological Chemistry* **278**, 24937-24943.

Singh, S. M. & Panda, A. K. (2005). Solubilization and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering* **99**, 303-310.

Stalker, D. M., McBride, K. E., Shewmaker, C. K. & Kwan-Hwa Park, J. F. R. a. Y.-D. C. (1996). Manipulation of storage compounds in transgenic plants. In *Progress in Biotechnology*, pp. 189-199: Elsevier.



Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology Evolution* **24**, 1596-1599.

Thuku, R. N., Brady, D., Benedik, M. J. & Sewell, B. T. (2009). Microbial nitrilases: versatile, spiral forming, industrial enzymes. *Journal of Applied Microbiology* **106**, 703-727.

Verwoerd, T. C., Dekker, B. M. M. & Hoekema, A. (1989). A small scale procedure for the rapid isolation of plant RNAs. *Nucleic Acids Research*, 2362.

Weatherburn, M. W. (1967). Phenol-hypochlorite reaction for determination of ammonia. *Analytical Chemistry* **39**, 971-974.

Wilson, S. A. & Drew, R. E. (1995). Transcriptional analysis of the amidase operon from *Pseudomonas aeruginosa*. *Journal of Bacteriology* **177**, 3052-3057.

Woestenenk, E., Hammarstrom, M., van der Berg, S., Hard, T. & Berglund, H. (2004). His Tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors *Journal of Structural and Functional Genomics* **5**, 217-229.

Zhou, J., Bruns, M. A. & Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Applied Environmental Microbiology* **62**, 316-322.

Zuckerkindl, E. & Pauling, L. (1965). *Evolutionary divergence and convergence in proteins*. New York: Academic Press.

