

https://dspace.org/

Nitrile hydratase enantioselectivity as a function of genetic diversity and directed evolution

Schmid, Kate

2020
https://hdl.handle.net/10539/31017

Downloaded from a DSpace repository. https://dspace.org/

Nitrile hydratase enantioselectivity as a function of genetic diversity and directed evolution

Kate Schmid

A thesis submitted to the Faculty of Science

University of the Witwatersrand

Johannesburg

In fulfilment for the requirements of the degree of Master of Science

18 August 2020

Declaration

I declare that this dissertation is my own, unaided work under the supervision of Prof Dean Brady, Dr Varsha Chhiba-Govindjee and Prof Karl Rumbold and any assistance received has been explicitly acknowledged. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.



_____

On this 18th day of August 2020

Abstract

The South African soil isolate Rhodococcus *rhodochrous* ATCC BAA-870 encodes a low-molecular mass nitrile hydratase (L-NHase), similar to that of from *R. rhodochrous* J1, which may be used to produce enantiopure β-substituted amides for use in peptidomimetics and other pharmaceuticals. This study aimed to compare the protein sequence of this enzyme with other known and novel NHase, heterologously express enzyme, identify residues involved in enantioselectivity of the L-NHase for β-substituted nitrile substrates and finally to carry out directed evolution on the recombinantly expressed L-NHase to improve its enantioselectivity. Multiple South African soil isolates were genetically screened to identify any NHase genes. The NHase genes from *R. rhodochrous* ATCC BAA-870, *R. rhodochrous* J1, five South African Rhodococcal soil isolates and other reported NHase genes were compared to determine if patterns of gene sequence and substrate usage occurred. Homology modelling and previous studies were used to visualize residues within the substrate channel that may interact with substrates. Directed evolution using iterative saturation mutagenesis (ISM) and the combinatorial active site saturation test (CASTing) was done to introduce mutations to the selected residues. Chiral high performance liquid chromatography was attempted to test for enantioselectivity of the recombinant L-NHase. All of the *Rhodococcus* soil isolates were found to possess a gene for L-NHase but not H-NHase. With the L-NHases from *R. rhodochrous* J1 and BAA-870, they formed three separate homologous genetic groups, however these not appear to have associated substrate usage patterns. Database mining revealed a possible origin for the *R. rhodochrous* J1 H-NHase. ISM-PCR was successful in introducing mutations within the beta subunit of the recombinant L-NHase but unfortunately not in the target region of the gene. The L-NHase from *R. rhodochrous* ATCC BAA-870 was successfully recombinantly expressed and was able to fully hydrolyse the substrates 4-methoxybenzonitrile, 3-amino-3-*p*-tolyl propanenitrile and 3-amino-3-(4-methoxyphenyl) propanenitrile, although enantioselectivity could not be quantified due to HPLC column degradation. Although some methods used in this study require optimisation, this is a promising approach to modify and possibly improve the enantioselectivity of the L-NHase from *R. rhodochrous* ATCC BAA-870 for β-substituted amides.

Acknowledgments

Thank you to my supervisor, Prof Dean Brady of the Molecular Science Institute, for his continuous patience and support and to my co-supervisors, Dr Varsha Chhiba-Govindjee and Prof Karl Rumbold, for all their assistance. Thank you to Dr Varsha Chhiba-Govindjee of the CSIR for providing initial samples of the enzyme substrates and the lyophilised nitrile hydratase and for explaining things I didn't understand; and to Professor Karl Rumbold for the use of his lab and equipment. Thank you to Joni Frederick for her advice and her previous research which formed the foundation for my own work.

Thank you to the members of the Organic research group for all the assistance and advice, in particular Dr Moira Bode and Dr Charles Changunda for their help with the organic synthesis, Dr Daniel Pienaar for assisting with HPLC and organic reactions, Dr Peter Wanyama Juma for the use of his equipment and fume hood, Pious Shuro for the use of his fume hood, Adelaide Mashweu for the useful discussions about molecular modelling, and to everyone else for their moral support and for lending me equipment when necessary. Thank you as well to the members of the IMBL lab for all their assistance and advice.

Thank you to Dr Craig Coel of Schrödinger Inc. and the CSIR/CHPC for providing access to the suite of Schrödinger Materials Molecular Modelling software. Particular thank you to Dr Katalin Phimister of Schrödinger Inc. for conducting tutorials using the software suite and for personally assisting me to troubleshoot the NHase homology modelling and to produce a final model.

Funding from the DST Biocatalysis Initiative and the National Research Foundation (NRF) as well as the Wits Postgraduate Award was greatly appreciated.

List of Abbreviations and Formulae

Å	Angstrom
ATCC	American Type Culture Collection
bp	base pairs
CAST	combinatorial active site saturation test
DMSO	dimethyl sulfoxide
DSMZ	German Collection of Microorganisms and Cell Cultures GmbH
E	enantiomeric ratio
ee_p	enantiomeric excess (product)
ee_s	enantiomeric excess (substrate)
EtBr	ethidium bromide
EtOAc	ethyl acetate
H-NHase	high molecular mass nitrile hydratase
HPLC	high-performance liquid chromatography
hr(s)	hour(s)
ISM	iterative saturation mutagenesis
JTT	Jones-Taylor-Thornton
kDa	kilodalton
K_m	Michaelis constant
K_{cat}	turnover number
L-NHase	low molecular mass nitrile hydratase
MCS	multiple cloning site
min(s)	minute(s)
MP	melting point
MWM	molecular weight marker
NCBI	National Centre for Biotechnology Information
NHase	nitrile hydratase
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600nm
PCR	polymerase chain reaction

PDB	Protein Data Bank
RBS	ribosome binding site
R*f*	retardation factor
rpm	revolutions per minute
rRNA	ribosomal RNA
SDS-PAGE	sodium dodecyl sulphate – polyacrylamide gel electrophoresis
sec	seconds
TAE	Tris-acetate-EDTA
TLC	thin layer chromatography
T_m	melting temperature
v.	version
V	volts
V_{max}	maximum velocity (reaction rate)

IUPAC codes for nucleotides and amino acids were used

Table of Contents

Declaration i
Abstract ii
Acknowledgments iii
List of Abbreviations and Formulae iv
Table of Contents vi
List of Figures ix
List of Tables xii
CHAPTER 1: GENERAL INTRODUCTION AND LITERATURE REVIEW 1
1.1 INTRODUCTION 1
1.1.1 The genus Rhodococcus and its biocatalytic potential 1
1.1.2 Nitrile hydratase enzymology and structure 3
1.1.3 Central metal ion 3
1.1.4 Gene structure 4
1.1.5 Co-type: high and low molecular mass 5
1.1.6 Enzymology 6
1.1.7 Gene Structure 8
1.1.8 Nitrile hydratase: function and importance to industry 8
1.2 AIMS AND OBJECTIVES 11
1.2.1 Aim 11
1.2.2 Specific objectives 11
1.3 CHAPTER SUMMARY 12
1.4 REFERENCES 14
Chapter 2: COMPARING THE SEQUENCE VARIATION AND SUBSTRATE SPECIFICITY OF NITRILE HYDRATASES WITHIN AND BEYOND THE *RHODOCOCCUS* GENUS 18
2.1 INTRODUCTION 18
2.1.1 Directed evolution as a means to improve industrial enzymes 18
2.1.2 Substrate specificity and geographic origin 19
2.1.3 Substrate channel amino acids may be a target for directed evolution 20

2.2 MATERIALS AND METHODS 21
2.2.1 Materials 21
2.2.2 Methods 22
2.3 RESULTS 25
2.3.1 Identification of nitrile hydratase genes in South African agricultural soil isolates by polymerase chain reaction 25
2.3.2 Database mining overview 27
2.3.3 The low molecular mass nitrile hydratases from various South African *Rhodococcus rhodochrous* isolates form a cluster with that of *Rhodococcus rhodochrous* J1 27
2.3.4 Comparison of nitrile hydratase sequences within the *Rhodococcus* genus 34
2.3.5 Commentary on the origins of the high molecular mass cobalt type nitrile hydratase 43
2.3.6 Phylogenetic analysis of nitrile hydratases based on substrate usage 44
2.3.7 Selection of amino acid residues for mutation 52
2.4 CONCLUSIONS 53
2.5 REFERENCES 55
2.6 APPENDIX 58
CHAPTER 3: GENETIC ENGINEERING AND DIRECTED EVOLUTION VIA ITERATIVE SATURATION MUTAGENESIS 63
3.1 INTRODUCTION 63
3.1.1 Recombinant protein expression and its contribution to modern biocatalysis 63
3.1.2 Industrially useful *Rhodococci* 63
3.1.3 Rhodococcal hosts and expression vectors 64
3.1.4 Recombinant expression of nitrile hydratase from *Rhodococcus rhodochrous* 64
3.1.5 Directed evolution for the improvement of industrially useful enzymes 66
3.1.6 The megaprimer method of PCR for whole-plasmid replication 67
3.1.7 Iterative saturation mutagenesis and the Combinatorial Active-site Saturation Test (CAST) 70
3.2 MATERIALS AND METHODS 72
3.2.1 Chemicals 72
3.2.2 Expression systems 72
3.2.3 Saturation mutagenesis by megaprimer PCR 81
3.3 RESULTS AND DISCUSSION 86
3.3.1 Recombinant expression of NHase using pRSFDuet-1 and pET21a(+) 86

3.3.2 Saturation mutagenesis by megaprimer PCR 91
3.4 CONCLUSIONS 95
3.5 REFERENCES 97
CHAPTER 4: ORGANIC SYNTHESIS AND BIOCONVERSION OF BETA-AMINO NITRILES 102
4.1 INTRODUCTION 102
4.1.1 Amides and their derivatives in medicine and industry 102
4.1.2 Current synthesis techniques for production of amides 103
4.1.3 Greener amide synthesis techniques 104
4.1.4 The potential of nitrile hydratase as a biocatalyst to produce enantiopure beta substituted nitriles 105
4.2 MATERIALS AND METHODS 107
4.2.1 General 107
4.2.2 Organic synthesis of beta-substituted nitriles 107
4.2.3 Column chromatography 109
4.2.4 Biocatalytic conversion of beta-substituted nitriles using crude protein extract 109
4.3 RESULTS AND DISCUSSION 113
4.3.1 Organic synthesis of beta-substituted nitriles 113
4.3.2 Biocatalytic conversion of beta-substituted nitriles using crude protein extract 117
4.3.3 Chiral high performance liquid chromatographic analysis of enzyme reaction products 122
4.3.4 Development of a high-throughput assay to monitor enantioselectivity 125
4.4 CONCLUSIONS 127
4.5 REFERENCES 129
4.6 APPENDIX 133
CONCLUSION AND FUTURE WORK 137

List of Figures

Figure 1.1: The nitrile hydratase active site coordination complex……4

Figure 1.2: Simplified schematic of the Fe-type and Co-type NHase genes……5

Figure 1.3: Alignment of the alpha and beta subunits from *R. rhodochrous* L-NHase and H-NHase……7

Figure 1.4: Simplified schematic of the Co-type low and high molecular mass NHase genes……8

Figure 1.5: Nitrile Hydratase hydrolyses nitriles to amides……9

Figure 2.1: Confirmation of gDNA quality from *R. rhodochrous* soil isolates……26

Figure 2.2: Example of PCR results showing the presence of L-NHase genes in some of the agricultural isolates……26

Figure 2.3: Alignment of nitrile hydratase alpha subunits……29

Figure 2.4: Alpha subunit residues which show variation between the three identified *R. rhodochrous* clusters……30

Figure 2.5: Alignment of nitrile hydratase beta subunits……31

Figure 2.6: Beta subunit residues which show variation between the three identified *R. rhodochrous* clusters……32

Figure 2.7: Visualisation of substrate channel amino acids in the Co-type low molecular mass NHase from *R. rhodochrous* ATCC-BAA 870……34

Figure 2.8: Pairwise identity matrix constructed from a MAFFT sequence alignment of the alpha subunits from various *Rhodococci*……36

Figure 2.9: Alignment of *Rhodococcus* alpha subunits showing conservation of residues……37

Figure 2.10: Schematic diagram showing the Co ion in the active site of *R. rhodochrous* ATCC BAA-870 and the residues within 10 Å of the ion……38

Figure 2.11: Alignment of *Rhodococcus* alpha subunits showing hydrophobicity of residues for an excerpt of the alignment, centred on the active site……39

Figure 2.12: Pairwise identity matrix constructed from a MAFFT sequence alignment of the beta subunits from various *Rhodococci*……41

Figure 2.13: Alignment of *Rhodococcus* beta subunits showing conservation of residues for an excerpt of the alignment, excluding the N- and C-termini……42

Figure 2.14: Alignment of *Rhodococcus* beta subunits showing hydrophobicity of residues for an excerpt of the alignment, excluding the N- and C-termini……43
Figure 2.15: Alignment of alpha subunits of all NHase sequences for which activity data was available……49
Figure 2.16: Alignment of beta subunits of all NHase sequences for which activity data was available……..50
Figure 3.1: Schematic of the two plasmid systems designed to express the Rhodococcal nitrile hydratase and amidase genes for herbicide degradation……65
Figure 3.2: The MEGAWHOP (megaprimer PCR of whole plasmid) method……68
Figure 3.3 Megaprimer PCR where the megaprimer is produced as the product in the first round of PCR…69
Figure 3.4: Plasmid map of pRSFDuet-1 with the alpha subunit and chaperone gene sequences……73
Figure 3.5: Plasmid map of pET-21a(+) with the beta subunit gene sequence……74
Figure 3.6: pJOE6731.1 vector for recombinant expression of the *R. rhodochrous* L-NHase alpha subunit and chaperone……79
Figure 3.7: pJOE6556.1 vector for recombinant expression of the *R. rhodochrous* L-NHase beta subunit…80
Figure 3.8: Plasmid map showing the position of mutagenic primers and the megaprimer formed in the first PCR step of iterative saturation mutagenesis……85
Figure 3.9: Restriction digests from single-plasmid transformations……87
Figure 3.10: Combined restriction digest of two expression plasmids from *E. coli* BL21(DE3)……..88
Figure 3.11: SDS-PAGE separation of induced and uninduced protein fractions from BL21(DE3) containing pRSFDuet-1-αC and pET21a(+)-β……90
Figure 3.12: Excerpt of the MAFFT alignment of original BAA-870 L-NHase beta subunit gene with the 16 ISM PCR samples……93
Figure 4.1: General amide structure……102
Figure 4.2: The formation of the peptide bond between amino acids……103
Figure 4.3: Reaction scheme……107
Figure 4.4: Thin-layer chromatography plates showing samples taken from fractions during workup using gravity-flow column chromatography purification of substrate 3c, showing elution of an impurity with the product……114
Figure 4.5: Enzyme conversion reactions after 30 min using cells containing either the BL21(DE3) two-construct system or the pJOE system……118

Figure 4.6: Conversion of substrate 3a to 4-methoxybenzamide using the two-construct expression system for NHase expression……119
Figure 4.7: Conversion of substrate 2c to 2d using the two-construct expression system for NHase expression……120
Figure 4.8: Conversion of substrate 3c to 3d using the two-construct expression system for NHase expression……121
Figure 4.9: Raw HPLC Chromatogram of the blank sample……122
Figure 4.10: Raw HPLC Chromatogram of NHase product 2d (3-Amino-3-*p*-tolyl propanamide)……124
Figure 4.11: Standard amide 2d was produced using lyophilised NHase from cells of *R. rhodochrous* ATCC BAA-870……125
Figure 4.12: Schematic of the reaction catalysed by glutamate dehydrogenase……126

List of Tables

Table 1.1: All available full and partial *R. rhodochrous* genome sequences……..2

Table 2.1: Summary of ExPASy ProtParam stability scores for *R. rhodochrous* cobalt type nitrile hydratases ……33

Table 2.2: Linked nitrile hydratase sequence and substrate data available on the BRENDA database……46

Table 2.3: Sequence profile based on phylogenetic analysis of nitrile hydratase beta subunits from BRENDA……51

Table 3.1: Volumes and final concentrations of reagents in a single sample for megaprimer PCR……83

Table 3.2: Cycling conditions for megaprimer PCR (first attempt)……84

Table 4.1: Nitrile substrates used in this study……110

CHAPTER 1: GENERAL INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

1.1.1 The genus Rhodococcus and its biocatalytic potential

The bacterial genus *Rhodococcus* belongs to the phylum *Actinobacteria*; class *Actinobacteria*; family *Nocardiacea*e and suborder *Corynbacterineae*. As of this writing, there are 50 validly published and confirmed species in the genus (according to the online database http://www.bacterio.net/rhodococcus.html). They have been described as forming three subclades, one containing *R. rhodochrous*, *R. erythropolis* and *R. equii,* respectively, although there is uncertainty around the third subclade, in particular the classification of *R. equii* (Mcminn *et al.*, 2000; Goodfellow *et al.*, 2012).

These bacteria are gram-positive, aerobic, non-motile, non-sporulating and contain genomes with a high GC content. As members of the suborder *Corynbacterineae,* they produce mycolic acids, branched lipids which form part of the cell wall in a characteristic complex of mycolyl–arabinogalactan–peptidoglycan. *Rhodococci* tend to change morphology with age, starting as cocci, but then forming rods and branched filaments (Gürtler and Seviour, 2010). Many species produce colonies of a reddish-pink colour, hence the name of the genus.

The *Rhodococci* are found commonly in soils, marine sediments and waste water (Gürtler and Seviour, 2010), in nitrile-utilising plants and in agricultural soil enriched in nitrates (Rapheeha *et al.*, 2017). They have also been isolated from air from a Russian space station and soil from an archaeological grave site. This variety in habitat is possibly due to the ability of *Rhodococci* to utilise a wide range of carbon and nitrile sources, possibly due to the genetic variety conferred by multiple large plasmids (Gürtler and Seviour, 2010). Because of their wide range of enzymes and environmental adaptations, the genomes of *Rhodococci* are a rich source of enzymes with potentially novel biocatalytic properties (McLeod *et al.*, 2006; Pathak *et al.*, 2013).

R. rhodochrous is the type species of its genus, while the type strain of *R. rhodochrous* is ATCC 372 (Zopf 1891) Tsukamura 1974 (https://bacdive.dsmz.de/index.php?search=10908&submit=Search). A search on

the National Centre for Biotechnology Information (NCBI) genome database (https://www.ncbi.nlm.nih.gov/genome/) yields 17 results as of March 2020, however only two of these are a complete genome (for strain NCTC10210). There are four scaffolds and ten contigs, shown in Table 1.1. There are 14 plasmid sequences on the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore).

Table 1.1: All available full and partial *R. rhodochrous* genome sequences*

Strain	Assembly	Size (Mb)	GC%	Replicons	WGS	Genes	Proteins	Release Date	Modify Date	Level
NCTC10210	GCA_900187265.1	5.27476	68.2	Chromosome	-	4839	4683	15 08 2017	21 12 2019	Complete Genome
ATCC BAA870	GCA_009883715.1	5.90382	67.8026	Chromosome and plasmid	-	5315	5119	15 01 2020	17 01 2020	Complete Genome
EP4	GCA_003004765.2	5.71765	67.9	Chromosome	-	5188	4932	18 09 2018	21 09 2019	Chromosome
NBRC 16069	GCA_001047055.1	5.20172	68.2	-	BBXP01	4798	4646	24 06 2015	07 02 2020	Contig
DSM 43241	GCA_001646825.1	5.18405	68.2	-	LRRK01	4805	4639	18 05 2016	08 02 2020	Scaffold
NCTC630	GCA_900455745.1	6.51314	70.6	-	UGVK01	6064	5822	30 07 2018	17 08 2019	Contig
NRRL B-1306	GCA_000716895.1	6.77732	61.7	-	JNWS01	6280	6066	01 07 2014	30 11 2019	Contig
J3	GCA_900177695.1	6.10685	67.9	-	FXAV01	5594	5376	20 04 2017	30 07 2019	Scaffold
SPC17	GCA_009831075.1	4.98578	68.4	-	WUUR01	4536	4398	06 01 2020	26 01 2020	Contig
J38	GCA_007830015.1	6.02043	67.9	-	VLKA01	5531	5278	02 08 2019	07 08 2019	Scaffold
J45	GCA_007829555.1	6.36116	67.8	-	VLJT01	5880	5545	02 08 2019	06 08 2019	Scaffold
ATCC 17895	GCA_000469645.1	6.86723	62.3	-	ASJJ01	6419	6183	23 09 2013	29 11 2019	Contig
ATCC 21198	GCA_000517665.1	6.47755	70.2	-	AZHI01	6003	5749	13 01 2014	10 02 2020	Contig
TRN71	GCA_001511235.1	4.87101	70.2	-	FBUK01	4615	3189	04 01 2016	04 12 2019	Contig
KG-21	GCA_001278665.1	6.09995	69.6	-	AZYO01	5661	5310	03 09 2015	11 02 2020	Contig
BKS6-46	GCA_000239135.3	6.21364	67.4	-	AGVW02	5772	5517	20 12 2011	21 12 2019	Contig
11Y	GCA_002003765.1 (contaminated)	8.58441	70.4	-	JNVR01	8676	7062	17 02 2017	17 02 2017	Contig

*This table is compiled based on the NCBI genome database and is accurate as of March 2020

WGS = Whole genome shotgun project

Brady *et al.* (2004) screened soil samples and identified *Rhodococcus rhodochrous* ATCC BAA-870, also known as “Ismall”, as a potentially useful nitrile-degrading organism with enantioselective capabilities. It was found that the NHase in this organism was able to hydrolyse β-hydroxynitriles (Kinfe *et al.*, 2009) and β-aminonitriles (Chhiba *et al.*, 2012) with varying levels of enantioselectivity. A later study of soil samples from various locations in South Africa found that the ability to hydrolyse β-substituted nitriles was most prevalent in *Rhodococci* isolated from agricultural soils (Rapheeha *et al.*, 2017). Five of these new *Rhodococcal* isolates were examined in the current study for their similarities to *R. rhodochrous* ATCC BAA-870.

1.1.2 Nitrile hydratase enzymology and structure

The first identification of a nitrile hydratase (NHase) was the NHase from *Arthrobacter sp*. J-1. (Asano *et al.* 1980). Asano *et al.* did recover active NHase, although no mention is made of any metal ions being necessary for expression. NHases constitute their own protein family and are metalloproteins containing either a non-heme iron ion (Sugiura *et al.*, 1987) or a non-corrin cobalt ion (Brennan *et al.*, 1996a). *R. rhodochrous* strains contain cobalt, while *R. erythropolis* and *R. equi* strains contain iron.

1.1.3 Central metal ion

Sugiura *et al.* (1987) first identified that the NHases from *Brevibacterium* R312, later renamed to *Rhodococcus* sp. R312 (Brennan *et al.*, 1996b), and *Pseudomonas chlororaphis* B23 both contained a central low-spin non-heme Fe(III) ion. Nelson *et al.* (1991) then identified the active site motif of C-X-Y-C-Z at residues 110 to 155 of the alpha subunit in the Fe-type NHase from *Brevibacterium sp*. Nagasawa *et al.*, (1988) found that *R. rhodochrous* J1 produced an NHase when the medium was supplemented with $CoCl_2$ and that cobalt acted as an inducer of the enzyme (Nagasawa *et al.*, 1988) as well as urea and amides (Nagasawa *et al.*, 1991). Fe-type NHases are inhibited by nitric oxide (NO), which is released by light exposure, a process known as photoactivation (Nakajima *et al.*, 1987; Mascharak, 2001).

NHases consist either of αβ-dimers or $\alpha_2\beta_2$-tetramers with a single ion per dimer (Huang *et al.,* 1997), although they may form larger multimers (Zhou *et al.*, 2010). The conserved active site sequence in all NHases consists of α-subunit residues only and forms the pattern $VC_1(T/S)LC_2SC_3(Y/T)$, where the Co-type contains VCTLCSCY and the Fe-type contains VCSLCSCT (Payne *et al.*, 1997). The metal ion is found at the α-β subunit interface. It is held in a claw-like complex by bonds to the sulphur atoms of the three cysteine residues, where C_2 is modified by oxidation to cysteine sulfinic acid (Cys-SO_2H) and C_3 is modified to cysteine sulfenic acid (Cys-SOH) (Murakami *et al.*, 2000). The other ligands are formed by two main-chain amide nitrogens, one from cysteine C_3 and one from the adjacent serine. This coordinates the metal ion in an octahedral coordination sphere with the sixth position in the Co-type being occupied by a hydroxide ion involved in the catalytic mechanism (Huang et al., 1997) or an NO molecule in the Fe-type (Nagashima *et al.*, 1998) (Figure 1.1). Despite some residues differing, the chemistry of the active site is retained in both Co- and Fe-type NHases, hence the coordination of the ion in both active sites is the same. The

structure of the NHase from *R. rhodochrous* ATCC has been predicted by homology modelling in the present study (see Sections 2.3.3.1 and 2.3.3.2).

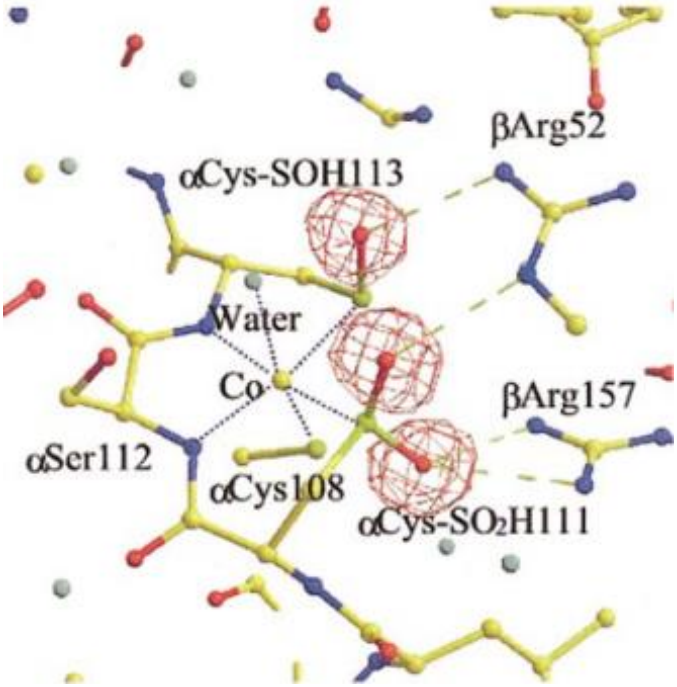


Figure 1.1: The nitrile hydratase active site coordination complex based on the crystal structure of the Co-type NHase from *Pseudonocardia thermophila* JCM 3095. The Co ion is coordinated by three cysteine residues and a serine from the alpha subunit as well as a hydroxide molecule (labelled *water*). βArg52 and βArg157 stabilise the active site through hydrogen bonds to the oxygen atoms (red spheres) of the modified cysteines. Reproduced from Miyanaga *et al.* (2001). PDB ID: 1UGQ. Residue numbers may be different in other NHases.

1.1.4 Gene structure

The operons of both Fe- and Co-type NHases contain corresponding genes, but in a differing order. It was found early on that NHase is co-expressed with an amidase (Asano *et al.*, 1980; Nagasawa *et al.*, 1988) and that they are found on the same operon (Nojiri *et al.*, 1999). In the Fe-type, the amidase gene is located upstream of both the subunit genes while in the Co-type the amidase gene can be downstream of both subunits (Komeda *et al.*, 1996) or upstream of both (Cameron *et al.*, 2005). In the Fe-type, the α-subunit gene is upstream of the β-subunit gene (Nojiri *et al.*, 1999) and in the Co-type this is the opposite

(Kobayashi *et al.*, 1991; Kim and Oriel, 2000). Finally there is an activator or chaperone protein, a 14kDa protein with 25% sequence identity to the beta subunit, which assists in inserting the cobalt ion into the alpha subunit (Cameron *et al.*, 2005) via self-subunit swapping (Liu *et al.*, 2012). It has been suggested that the differing gene organisation may be a result of gene rearrangement (Komeda *et al.*, 1996). A number of regulator genes have also been identified on the operon which control the expression of NHase (Figure 1.2).

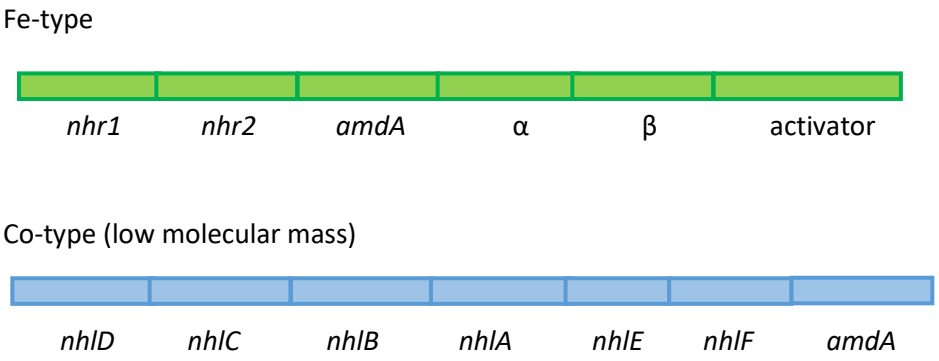


Figure 1.2: Simplified schematic of the Fe-type and Co-type NHase genes. The open reading frames in each gene correspond to each other, although the order of α and β subunits is opposite. *nhr1, nhr2, nhlD, nhlC*: regulator proteins; *amdA*: amidase. *nhlB*: beta subunit ; *nhlA*: alpha subunit; *nhlE*: beta homologue; *nhlF*: Co transporter. Fe-type from Rhodococcus sp. N-771 (Nojiri *et al.*, 1999); Co-type from *Bacillus sp*. BR449 (Kim and Oriel, 2000) and R. rhodochrous J1 (Komeda *et al.*, 1996).

1.1.5 Co-type: high and low molecular mass

Studies on the Co-type NHase from *R. rhodochrous* J1 Kobayashi *et al.* (1991) showed that two types exist– one of low molecular weight at ~130 kDa (L-NHase) and one of high molecular weight at ~530 kDa (H-NHase) – and that these types are coded for by two separate genes. Thus far, *R. rhodochrous* J1 is the only organism which possesses the H-NHase as well as the L-NHase, while *R. rhodochrous* ATCC BAA-870 and other *Rhodococci* possess only the low molecular weight Co-type NHase (Frederick, 2013). It was thought that the H-NHase arose from a unique gene duplication event in *R. rhodochrous* J1 (Komeda *et al.*, 1996) however H-NHase sequences have since been reported in other bacteria. A search for H-NHase sequences on the NCBI website gives 145 results (as of March 2020), although all except the *R. rhodochrous* J1 results appear to be gene sequences while the associated proteins have not yet been studied.

1.1.6 Enzymology

The H-NHase has been reported to exist as a 530 kDa $\alpha_{10}\beta_{10}$ oligomer (Nagasawa *et al.*, 1991; Mizunashi *et al.*, 1998; Lan *et al.*, 2017). The L-NHase and H-NHase have different optimal pH and temperature conditions, as well as different substrate activity profiles (Wieser *et al.*, 1998), as early research found that the H-NHase had a higher or similar affinity for aliphatic nitriles while the L-NHase had a greater affinity for aromatic nitriles (Kobayashi *et al.*, 1992; Wieser *et al.*, 1998). Urea acts as an inducer of H-NHase while amides induce both NHases and amidase. In L-NHase, nhlD represses NHase expression. Amides induce the nhlC gene, the protein product of which represses the nhlD gene, allowing the NHase to be expressed (Komeda *et al.*, 1996). Crotonamide was found to be necessary for the H-NHase to assemble its oligomeric form (Mizunashi *et al.*, 1998).

An alignment of the amino acid sequences of the two NHases from *R. rhodochrous* J1 by Frederick (2013) indicated a 42.89% identity between the two proteins which suggested that they had arisen from a gene-duplication event, and then diverged to act on different nitrile substrates, as originally suggested by Wieser *et al.* (1998). The alpha and beta subunits of each type are in fact of similar length, however the L-NHase alpha subunit has a seven residue insertion at the N-terminus. The H-NHase beta subunit has a three residue insertion at the 64^{th} position (Figure 1.3).

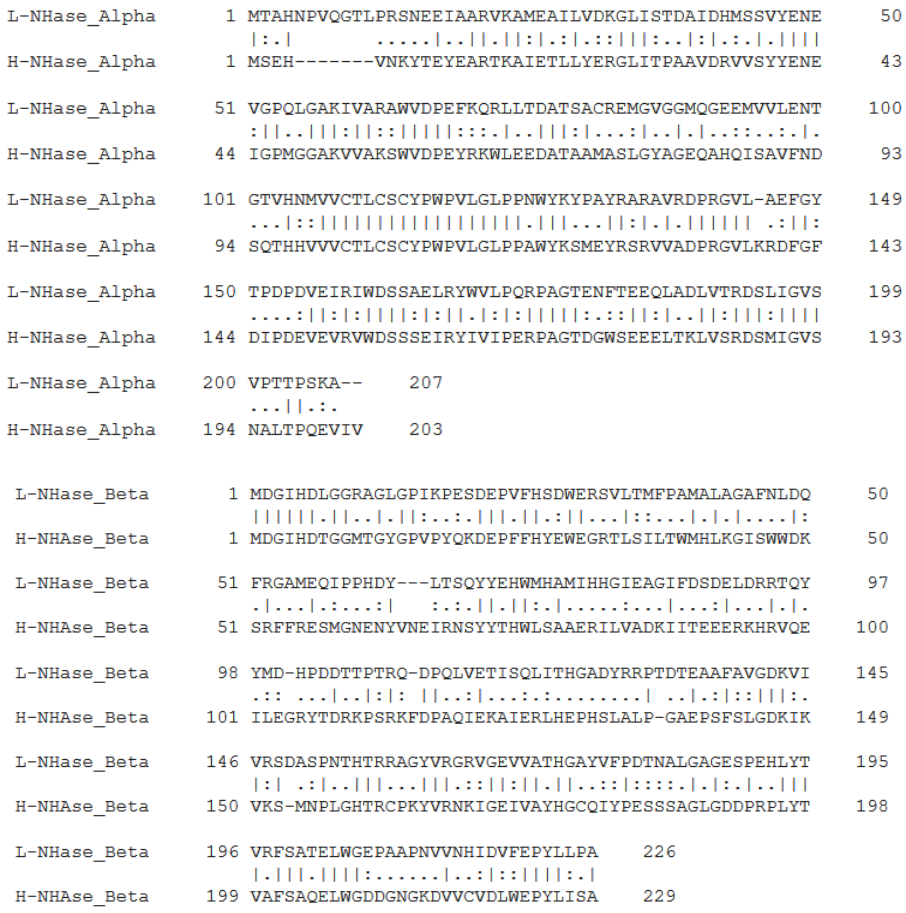


Figure 1.3: Alignment of the alpha and beta subunits from *R. rhodochrous* L-NHase and H-NHase. Both alignments were produced using EMBOSS NEEDLE on the EMBL-EBI website (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) using the BLOSUM62 matrix and default gap opening and extension penalties.

1.1.7 Gene Structure

The gene structure of both L-NHase and H-NHase is similar, with both genes encoding their respective alpha and beta subunits on a single mRNA. In both genes, the beta subunit occurs upstream of the alpha subunit with a short distance of 13 bp between the subunits (Kobayashi *et al.*, 1991). Interestingly, the H-NHase operon does not contain an amidase gene (Komeda *et al.*, 1996). The *nhhF* gene encodes a transposase, an enzyme which can excise a DNA sequence and re-insert it elsewhere in the genome (Komeda *et al.*, 1996). This supports the suggestion by Wieser *et al.* (1998) that the H-NHase arose from a gene duplication event. Figure 1.4 shows compares the Fe- and Co-type gene layouts.

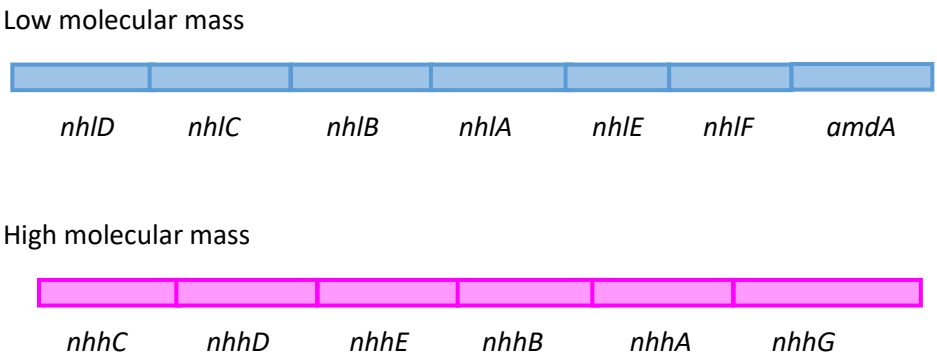


Figure 1.4: Simplified schematic of the Co-type low and high molecular mass NHase genes. *nhlD*: down-regulator of expression; *nhlC*: up-regulator of expression; *nhlB, nhhB*: beta subunit; *nhlA, nhhA*: alpha subunit; *nhlF*: cobalt transporter; *nhlE, nhhG*: beta homologue; *amdA*: amidase. Based on (Komeda *et al.*, 1996, 1997; Cameron *et al.*, 2005).

1.1.8 Nitrile hydratase: function and importance to industry

Some of the most comprehensively studied *Rhodococcal* enzymes are those involved in nitrile degradation, particularly a group of three enzymes with interlinked functions. These are nitrile hydratase, amidase and nitrilase (Bunch, 1998). Nitrile hydratase and amidase are able to accomplish the same hydrolytic pathway as nitrilase alone (Brady *et al.,* 2004). The relationship between reactions catalysed by the three enzymes is shown in Figure 1.5.

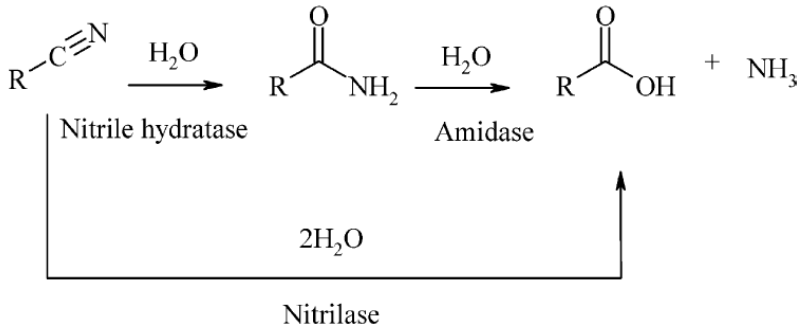


Figure 1.5: Nitrile Hydratase hydrolyses nitriles to amides which are then hydrolysed by amidase to carboxylic acids and ammonia. Nitrilase is able to catalyse hydrolysis of nitriles directly to carboxylic acids. Reproduced from Brady *et al.* (2004).

The various Co- and Fe-type NHases have been shown to have highly varied substrate specificities, although Co-type NHases tend to act preferably on aromatic nitriles while Fe-type NHases hydrolyse mainly aliphatic nitriles (Brandao *et al*., 2003). More recent research by Rzeznicka *et al.* (2010) on the Fe-type NHase from *Rhodococcus equi* TG328-2 found that this Fe-type NHase was able to hydrolyse aromatic and heterocyclic nitriles, but it had a greater preference for aliphatic nitriles, which is in line with the previously reported results. Work by van Pelt *et al.* (2011) showed that an Fe-type NHase from *R. erythropolis* AJ270 showed preference for hexanenitrile and the arylaliphatic substrate phenylacetonitrile, while the Co-type NHases used in the study showed a preference for the aromatic substrate benzonitrile. Using the crystal structure of the Co-type NHase from *Pseudonocardia thermophila* JCM 3095, Miyanaga *et al.* (2001) predicted that the conserved βTrp72 residue creates a larger space in the active site compared to the corresponding βTyr76 residue in the Fe-type, thereby contributing to the ability of the Co-type to accept larger, aromatic substrates. Mashweu *et al.* (2020) profiled the aromatic substrate scope of the L-NHase from *R. rhodochrous* ATCC BAA-870 and found that the NHase was able to hydrolyse many smaller aromatic nitriles, with increased size, position of ring substituents and steric hindrance around the cyano group decreasing the enzyme's affinity. It is possible that the increased size of the substrate pocket causes reduced affinity for less bulky aliphatic nitriles by preventing or decreasing the strength of interactions between the substrate and residues within the active site.

The Co-type H-NHase from *R. rhodhchrous* J1 is used as a third generation biocatalyst, replacing earlier Fe-type NHases (Kobayashi *et al.*, 1992) to produce acrylamide from acrylonitrile by the Mitsubishi Rayon Co. (Kobayashi and Shimizu, 1998) and nicotinamide by Lonza AG (Wieser and Nagasawa 2000). An NHase has also been used to produce 5-cyanovaleramide, the starting material for the herbicide azafenidin by DuPont (Thomas *et al.*, 2002). The Co-type H-NHase was found to tolerate a wider range of temperatures and to have a higher affinity for acrylonitrile than the older Fe-type NHase biocatalysts (Kobayashi *et al.*, 1992), as well as catalysing an almost 100% conversion of nitrile to amide (Yamada and Kobayashi, 1996). Other potential uses include producing intermediates in the synthesis of antiretrovirals (Priefert *et al.*, 2004) and beta-adrenergic agents (Rapheeha *et al.*, 2017), as well as bioremediation in polluted environments (Kuyukina and Ivshina, 2010).

It is likely that the wide variety of NHases could be applied to generating many chemicals, for both industry and pharmaceutical use. There are few commercially available NHases, hence the discovery and activity profiling of new NHases may be a profitable area. There are currently no NHase enzymes available on the Sigma-Aldrich website (www.sigmaaldrich.com). Creative Enzymes (New York, USA) provides seven nitrile hydratases, three of which are identified as originating from *R. erythropolis, Rhodopseudomonas palustris* and *Sinorhizobium meliloti* respectively. One Co-type NHase is available from Prozomix (Haltwhistle, England), although disclosure of the source organism requires an enzyme toolkit agreement to be signed first. Alternatively, NHase-producing organisms can be obtained from the ATCC (American Type Culture Collection) and DSMZ (German Collection of Microorganisms and Cell Cultures GmbH) culture collections. As of March 2020, there are 435 *Rhodococcus* results and 95 *R. rhodochrous* isolates in the ATCC collection. The DSMZ collection yields 320 results for *Rhodococcus* and 24 *R. rhodochrous* results.

The use of enzymes allows for the use of cleaner “green” chemical processes. Traditional methods produce large amounts of toxic by-products or waste and require large inputs of resources and energy (Thomas *et al.*, 2002). Biological molecules have the advantage of being stable and active at milder temperatures, pressures and conditions of acidity (Sheldon and Rantwijk, 2004), and so less energy is spent in maintaining the reaction at extreme conditions (Thomas *et al.*, 2002). Because enzymes catalyse reactions under milder conditions, protecting groups do not have to be added to labile groups on the substrate, which decreases the number of steps in the process (Pollard and Woodley, 2006). As well as the environmental benefits, the discovery of new biocatalysts may allow the innovation of new industrial processes and products (Wong and Whitesides, 1994).

1.2 AIMS AND OBJECTIVES

1.2.1 Aim

The overall aim is to characterise the enantioselectivity of the nitrile hydratase enzyme of *Rhodococcus rhodochrous* towards β-substituted aminonitriles first based on genetic variability and then to determine if this can be improved through directed evolution targeting the nitrile hydratase gene. This is in order to produce enantiomerically pure β-substituted amides as intermediates for pharmaceutical and industrial applications.

1.2.2 Specific objectives

I. To evaluate the nitrile-hydratase activity of the new *Rhodococcus Rhodochrous* soil isolates as well as that of *R. rhodochrous* ATCC BAA-870 ("Ismall") towards beta-substituted aminonitriles.
   - How does the activity of the isolates towards different substrates compare?
   - Do they show enantioselectivity?
   - Is the difference significant?
   - Minor aim: To synthesis the β-substituted nitrile substrates in the chemical laboratory.

II. To sequence the nitrile hydratase gene(s) of each isolate.
   - Are both the high and low molecular weight genes present and can they be successfully isolated and sequenced?
   - How similar are the genes from different isolates?
   - Are any patterns evident between the sequences and the activity towards different substrates, e.g. SNPs
   - Can a particular gene feature be associated with greater activity?
   - Should the L-NHase or the H-NHase be the focus of directed evolution?

III. To determine the most active isolate and carry out directed evolution on its nitrile hydratase gene to see if enantioselectivity towards beta-substituted aminonitriles can be improved.
    - Use either the gene from Ismall or from another isolate. Multiple genes may be used if time allows.
    - Select a suitable method of directed evolution.
    - Find a suitable high throughput screening assay to analyse the large number of variant enzymes which will be produced through directed evolution.
    - It may be necessary to develop an enantioselective screening method.
    - Does enantioselectivity improve with successive rounds of evolution?
    - Can activity of the nitrile hydratase enzyme be maintained at a sufficient level with changes in enantioselectivity? Are there any other changes in the behaviour or stability with directed evolution?

1.3 CHAPTER SUMMARY

Chapter 2

This chapter focuses on data mining and categorising the sequence differences between the various L-NHase genes as well as examining the reported substrate profiles in the literature to determine if there are any similarities or differences which can be associated with substrate preference. Alignments of the alpha and beta subunits to the alpha and beta subunits from other Co- and Fe-type NHases will be discussed to see if there are any patterns of residue characteristics or noticeable differences which could be linked to substrate preference. This may allow the residues involved in enantioselectivity to be identified. A three dimensional homology model of the L-NHase from *R. rhodochrous* ATCC BAA-870 will be produced to study the three dimensional position of its residues and to see if there are any patterns in the residues that can be associated with reported activity and enantioselectivity. Finally residues will be selected which can be mutated using iterative saturation mutagenesis and the combinatorial active site saturation test (CASTing) to improve enantioselectivity for either the R- or S-enantiomer of a chiral substrate.

Chapter 3

This chapter discusses the attempt to produce recombinant active L-NHase using a two-construct plasmid system in *E. coli* BL21(DE3) and then documents the process of introducing mutations to the selected residues in the beta subunit via iterative site saturation mutagenesis PCR and iterative cycles of CASTing. It also covers the investigation of activity in the samples prior to mutation and during storage and discusses the results of the mutagenic PCR.

Chapter 4

This chapter covers the synthesis of chiral β-substituted nitrile substrates for the L-NHase from *R. rhodochrous* ATCC BAA-870 followed by activity tests of the wild type enzyme. It discusses the development of a high-performance liquid chromatography (HPLC) method to quantify enantioselectivity as well as discussing the difficulty and future possibility of developing a high-throughput screening assay for enantioselectivity.

1.4 REFERENCES

Asano, Y., Tani, Y. and Yamada, H., 1980. A new enzyme “nitrile hydratase” which degrades acetonitrile in combination with amidase. *Agricultural and Biological Chemistry*, 44(9), pp.2251–2252.

Brady, D., Beeton, A., Zeevaart, J., Kgaje, C., van Rantwijk, F. and Sheldon, R.A., 2004. Characterisation of nitrilase and nitrile hydratase biocatalytic systems. *Applied Microbiology and Biotechnology*, 64(1), pp.76–85.

Brennan, B.A., Alms, G., Nelson, M.J., Durney, L.T. and Scarrow, R.C., 1996a. Nitrile Hydratase from Rhodococcus rhodochrous J1 Contains a Non-Corrin Cobalt Ion with Two Sulfur Ligands. *Journal of the American Chemical Society*, 118(38), pp.9194–9195.

Brennan, B.A., Cummings, J.G., Chase, D.B., Turner, I.M. and Nelson, M.J., 1996b. Resonance Raman spectroscopy of nitrile hydratase, a novel iron-sulfur enzyme. *Biochemistry*, 35(31), pp.10068–10077.

Bunch, A.W., 1998. Biotransformation of nitriles by rhodococci. *Antonie van leeuwenhoek*, 74(1–3), pp.89–97.

Cameron, R.A., Sayed, M. and Cowan, D.A., 2005. Molecular analysis of the nitrile catabolism operon of the thermophile Bacillus pallidus RAPc8. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1725(1), pp.35–46.

Chhiba, V., Bode, M.L., Mathiba, K., Kwezi, W. and Brady, D., 2012. Enantioselective biocatalytic hydrolysis of β-aminonitriles to β-amino-amides using Rhodococcus rhodochrous ATCC BAA-870. *Journal of Molecular Catalysis B: Enzymatic*, 76, pp.68–74.

Frederick, J., 2013. *Genetic characterization of Rhodococcus rhodochrous ATCC BAA-870 with emphasis on nitrile hydrolysing enzymes*. [online] University of Cape Town. Available at: <http://open.uct.ac.za/handle/11427/4262> [Accessed 11 Feb. 2016].

Goodfellow, M., Kämpfer, P., Busse, H.-J., Trujillo, M.E., Suzuki, K., Ludwig, W. and Whitman, W.B. eds., 2012. *Bergey’s Manual® of Systematic Bacteriology*. [online] New York, NY: Springer New York. Available at: <http://link.springer.com/10.1007/978-0-387-68233-4> [Accessed 5 Mar. 2016].

Gürtler, V. and Seviour, R.J., 2010. Systematics of Members of the Genus Rhodococcus (Zopf 1891) Emend Goodfellow et al. 1998. In: H.M. Alvarez, ed. *Biology of Rhodococcus*, Microbiology Monographs. [online] Berlin, Heidelberg: Springer.pp.1–28. Available at: <https://doi.org/10.1007/978-3-642-12937-7_1> [Accessed 25 Mar. 2020].

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

Huang, Y., Zhong, C., Lin, H.X., Wang, J. and Peng, Y., 2018. Reconstructing Phylogeny by Aligning Multiple Metabolic Pathways Using Functional Module Mapping. *Molecules*, 23(2), p.486.

Kim, S.-H. and Oriel, P., 2000. Cloning and expression of the nitrile hydratase and amidase genes from Bacillus sp. BR449 into Escherichia coli. *Enzyme and Microbial Technology*, 27, pp.492–501.

Kinfe, H.H., Chhiba, V., Frederick, J., Bode, M.L., Mathiba, K., Steenkamp, P.A. and Brady, D., 2009. Enantioselective hydrolysis of β-hydroxy nitriles using the whole cell biocatalyst Rhodococcus rhodochrous ATCC BAA-870. *Journal of Molecular Catalysis B: Enzymatic*, 59(4), pp.231–236.

Kobayashi, M., Nagasawa, T. and Yamada, H., 1992. Enzymatic synthesis of acrylamide: a success story not yet over. *Trends in Biotechnology*, 10, pp.402–408.

Kobayashi, M., Nishiyama, M., Nagasawa, T., Horinouchi, S., Beppu, T. and Yamada, H., 1991. Cloning, nucleotide sequence and expression in Escherichia coli of two cobalt-containing nitrile hydratase genes from Rhodococcus rhodochrous J1. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1129(1), pp.23–33.

Kobayashi, M. and Shimizu, S., 1998. Metalloenzyme nitrile hydratase: Structure, regulation, and application to biotechnology. *Nature Biotechnology*, 16(8), pp.733–736.

Komeda, H., Kobayashi, M. and Shimizu, S., 1996. A novel gene cluster including the Rhodococcus rhodochrous J1 nhlBA genes encoding a low molecular mass nitrile hydratase (L-NHase) induced by its reaction product. *Journal of Biological Chemistry*, 271(26), pp.15796–15802.

Komeda, H., Kobayashi, M. and Shimizu, S., 1997. A novel transporter involved in cobalt uptake. *Proceedings of the National Academy of Sciences*, 94(1), pp.36–41.

Kuyukina, M.S. and Ivshina, I.B., 2010. Application of Rhodococcus in Bioremediation of Contaminated Environments. In: *Biology of Rhodococcus*, Microbiology Monographs. [online] Springer, Berlin, Heidelberg.pp.231–262. Available at: <https://link.springer.com/chapter/10.1007/978-3-642-12937-7_9> [Accessed 29 Mar. 2018].

Lan, Y., Zhang, X., Liu, Z., Zhou, L., Shen, R., Zhong, X., Cui, W. and Zhou, Z., 2017. Overexpression and characterization of two types of nitrile hydratases from Rhodococcus rhodochrous J1. *PLOS ONE*, 12(6), p.e0179833.

Liu, Y., Cui, W., Xia, Y., Cui, Y., Kobayashi, M. and Zhou, Z., 2012. Self-Subunit Swapping Occurs in Another Gene Type of Cobalt Nitrile Hydratase. *PLoS ONE*, 7(11), p.e50829.

Mascharak, P.K., 2001. Structural and functional models of nitrile hydratase. *Coordination chemistry reviews*, 225(1), pp.201–214.

Mashweu, A.R., Chhiba-Govindjee, V.P., Bode, M.L. and Brady, D., 2020. Substrate Profiling of the Cobalt Nitrile Hydratase from Rhodococcus rhodochrous ATCC BAA 870. *Molecules*, 25(1), p.238.

McLeod, M.P., Warren, R.L., Hsiao, W.W., Araki, N., Myhre, M., Fernandes, C., Miyazawa, D., Wong, W., Lillquist, A.L., Wang, D. and others, 2006. The complete genome of Rhodococcus sp. RHA1 provides insights into a catabolic powerhouse. *Proceedings of the National Academy of Sciences*, 103(42), pp.15582–15587.

Mcminn, E.J., Alderson, G., Dodson, H.I., Goodfellow, M. and Ward, A.C., 2000. Genomic and phenomic differentiation of Rhodococcus equi and related strains. *Antonie van Leeuwenhoek*, 78(3–4), pp.331–40.

Miyanaga, A., Fushinobu, S., Ito, K. and Wakagi, T., 2001. Crystal Structure of Cobalt-Containing Nitrile Hydratase. *Biochemical and Biophysical Research Communications*, 288(5), pp.1169–1174.

Mizunashi, W., Nishiyama, M., Horinouchi, S. and Beppu, T., 1998. Overexpression of high-molecular-mass nitrile hydratase from Rhodococcus rhodochrous J1 in recombinant Rhodococcus cells. *Applied Microbiology and Biotechnology*, 49(5), pp.568–572.

Murakami, T., Nojiri, M., Nakayama, H., Odaka, M., Yohda, M., Dohmae, N., Takio, K., Nagamune, T. and Endo, I., 2000. Post-translational modification is essential for catalytic activity of nitrile hydratase. *Protein Science*, 9(05), pp.1024–1030.

Nagasawa, T., Mathew, C.D., Mauger, J. and Yamada, H., 1988. Nitrile hydratase-catalyzed production of nicotinamide from 3-cyanopyridine in Rhodococcus rhodochrous J1. *Applied and environmental microbiology*, 54(7), pp.1766–1769.

Nagasawa, T., Takeuchi, K. and Yamada, H., 1988. Occurrence of a cobalt-induced and cobalt-containing nitrile hydratase in Rhodococcus Rhodochrous J1. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS*, 155(2), p.9.

Nagasawa, T., Takeuchi, K. and Yamada, H., 1991. Characterization of a new cobalt-containing nitrile hydratase purified from urea-induced cells of Rhodococcus rhodochrous J1. *European journal of biochemistry*, 196(3), pp.581–589.

Nagashima, S., Nakasako, M., Dohmae, N., Tsujimura, M., Takio, K., Odaka, M., Yohda, M., Kamiya, N. and Endo, I., 1998. Novel non-heme iron center of nitrile hydratase with a claw setting of oxygen atoms. *Nature Structural Biology*, 5(5), pp.347–351.

Nakajima, Y., Doi, T., Satoh, Y., Fujiwara, A. and Watanabe, I., 1987. Photoactivation of nitrile hydratase in Corynebacterium sp. N-774. *Chemistry Letters*, 9, pp.1767–1770.

Nelson, M.J., Jin, H., Turner, I.M., Grove, G., Scarrow, R.C., Brennan, B.A. and Que, L., 1991. A novel iron-sulfur center in nitrile hydratase from Brevibacterium sp. *Journal of the American Chemical Society*, 113(18), pp.7072–7073.

Nojiri, M., Yohda, M., Odaka, M., Matsushita, Y., Tsujimura, M., Yoshida, T., Dohmae, N., Takio, K. and Endo, I., 1999. Functional expression of nitrile hydratase in Escherichia coli: requirement of a nitrile hydratase activator and post-translational modification of a ligand cysteine. *Journal of biochemistry*, 125(4), pp.696–704.

Pathak, A., Green, S.J., Ogram, A. and Chauhan, A., 2013. Draft Genome Sequence of Rhodococcus opacus Strain M213 Shows a Diverse Catabolic Potential. *Genome Announcements*, [online] 1(1). Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3569331/> [Accessed 29 Mar. 2018].

Payne, M.S., Wu, S., Fallon, R.D., Tudor, G., Stieglitz, B., Turner, I.M. and Nelson, M.J., 1997. A stereoselective cobalt-containing nitrile hydratase. *Biochemistry*, 36(18), pp.5447–5454.

van Pelt, S., Zhang, M., Otten, L.G., Holt, J., Sorokin, D.Y., van Rantwijk, F., Black, G.W., Perry, J.J. and Sheldon, R.A., 2011. Probing the enantioselectivity of a diverse group of purified cobalt-centred nitrile hydratases. *Organic & Biomolecular Chemistry*, 9(8), p.3011.

Pollard, D.J. and Woodley, J.M., 2006. Biocatalysis for pharmaceutical intermediates: the future is now. *Trends in Biotechnology*, 25(2), pp.66–73.

Priefert, H., O'Brien, X.M., Lessard, P.A., Dexter, A.F., Choi, E.E., Tomic, S., Nagpal, G., Cho, J.J., Agosto, M., Yang, L., Treadway, S.L., Tamashiro, L., Wallace, M. and Sinskey, A.J., 2004. Indene bioconversion by a toluene inducible dioxygenase of Rhodococcus sp. I24. *Applied Microbiology and Biotechnology*, 65(2), pp.168–176.

Rapheeha, O.K.L., Roux-van der Merwe, M.P., Badenhorst, J., Chhiba, V., Bode, M.L., Mathiba, K. and Brady, D., 2017. Hydrolysis of nitriles by soil bacteria: variation with soil origin. *Journal of Applied Microbiology*, 122(3), pp.686–697.

Rzeznicka, K., Schätzle, S., Böttcher, D., Klein, J. and Bornscheuer, U.T., 2010. Cloning and functional expression of a nitrile hydratase (NHase) from Rhodococcus equi TG328-2 in Escherichia coli, its purification and biochemical characterisation. *Applied Microbiology and Biotechnology*, 85(5), pp.1417–1425.

Sheldon, R.A. and Rantwijk, F. van, 2004. Biocatalysis for Sustainable Organic Synthesis. *Australian Journal of Chemistry*, 57(4), pp.281–289.

Sugiura, Y., Kuwahara, J., Nagasawa, T. and Yamada, H., 1987. Nitrile hydratase. The first non-heme iron enzyme with a typical low-spin iron(III)-active center. *Journal of the American Chemical Society*, 109(19), pp.5848–5850.

Thomas, S.M., DiCosimo, R. and Nagarajan, V., 2002. Biocatalysis: applications and potentials for the chemical industry. *TRENDS in Biotechnology*, 20(6), pp.238–242.

Wieser, M., Takeuchi, K., Wada, Y., Yamada, H. and Nagasawa, T., 1998. Low-molecular-mass nitrile hydratase from Rhodococcus rhodochrous J1: purification, substrate specificity and comparison with the analogous high-molecular-mass enzyme. *FEMS Microbiology Letters*, 169(1), pp.17–22.

Wong, C.-H. and Whitesides, G.M., 1994. *Enzymes in Synthetic Organic Chemistry*. Academic Press.

Yamada, H. and Kobayashi, M., 1996. Nitrile Hydratase and Its Application to Industrial Production of Acrylamide. *Bioscience, Biotechnology, and Biochemistry*, 60(9), pp.1391–1400.

Zhou, Z., Hashimoto, Y., Cui, T., Washizawa, Y., Mino, H. and Kobayashi, M., 2010. Unique Biogenesis of High-Molecular Mass Multimeric Metalloenzyme Nitrile Hydratase: Intermediates and a Proposed Mechanism for Self-Subunit Swapping Maturation. *Biochemistry*, 49(44), pp.9638–9648.

Chapter 2: COMPARING THE SEQUENCE VARIATION AND SUBSTRATE SPECIFICITY OF NITRILE HYDRATASES WITHIN AND BEYOND THE *RHODOCOCCUS* GENUS

2.1 INTRODUCTION

2.1.1 Directed evolution as a means to improve industrial enzymes

Although enzymes are viable targets as novel catalysts, they may have characteristics which are not conducive to large scale industrial processes. For example, in some cases the enzyme may not be stable under the conditions required to produce the greatest reaction yield (Burton *et al.*, 2002), it may not be selective or specific enough for the substrate or catalyse at a fast enough rate (Burton *et al.,* 2002), or it may be necessary to widen its catalytic activity towards unnatural substrates or to reverse its enantioselectivity (Nair *et al.*, 2010). When engineering an enzyme for improved enantioselectivity, according to Carrea and Riva (2008), the eventual effect of the introduced mutations is to make the activation energy for one enantiomeric transformation much lower than for the other enantiomeric transformation. If the enzyme is to be used to produce pharmaceuticals, enantiopurity of the product is often required as only one enantiomer is active, or in some cases, one enantiomer may cause negative side effects (Caner *et al.*, 2004). In order to change the selectivity of an industrially useful enzyme, it is useful to determine which residues in the amino acid sequence should be targeted to produce the desired effect, instead of taking a completely random approach.

Guidelines for the selection of residues for directed evolution have previously been described by Acevedo-Rocha *et al.* (2014). The following analyses in this chapter were motivated by these guidelines.
Briefly, the process involves:

1) Visualising the structure either using a crystal structure or a homology model.
2) Identifying residues necessary to the function you want to improve (activity, enantioselectivity, stability). This can be done using docking studies and by reviewing previous studies. The residues are then grouped into clusters based on their position in the gene sequence. A cluster of residues is then mutated at the same time.

3) Deciding what technique to use to introduce mutations – often this involves error prone PCR (ePCR) or another variation on PCR. In this step, mutagenic primers are designed, often using a reduced amino acid alphabet to decrease the screening effort when selecting recombinant colonies.
4) Choosing an assay to test for the desired characteristic (preferably high or medium throughout) as well as doing quick quality control (QQC) on the library to ensure the expected level of diversity has been achieved.

In order to choose residues as targets for directed evolution, a homology model was created for this study using the amino acid sequence of the low molecular mass nitrile hydratase (L-NHase) from *Rhodococcus rhodochrous* ATCC BAA-870, characterised by Frederick (2013). L-NHase structures available on the Protein Data Bank were used as templates for the modelling. The literature was consulted to understand patterns of activity and enantioselectivity found in other nitrile hydratases as well as to determine if any residues could be associated with enzyme behaviour.

2.1.2 Substrate specificity and geographic origin

Brandao *et al.* (2003) conducted a study of various nitrile hydratases extracted from *R. erythropolis* isolates from distinct geographical regions, concluding that the variation in the gene sequences and thus the amino acid sequences was influenced by the different environments, conferring unique substrate specificity based on geographic source. They further found that the amino acid variations were located near the active site, contributing to the stabilisation of the active site. Hence, different residues in the vicinity of the active site would confer slight conformational differences, resulting in differing substrate specificities (and potentially enantioselectivity). This idea of nearby residues influencing the active site conformation is supported by the findings of Morley and Kazlauskas (2005), who analysed multiple enzymes and concluded that the residues within a 10Å radius of the active site residues tend to have the greatest effect on enantioselectivity and selectivity.

A study by van Pelt *et al.* (2011) found that Co-type NHases were more active towards aromatic substrates, and the particular Co-type NHases utilised were selective towards *(S)*-2-phenylpropionitrile. They also found that enantioselectivity was reduced when the aromatic nitrile substrate contained a hydroxyl group on the alpha carbon, and was reduced even further by the presence of an amine group in the same

position. Although they do not discuss which residues may interact with the atoms of a bound substrate molecule, it may be possible to analyse such interactions using molecular docking and hence determine which residues could be altered to improve or change enantioselectivity.

Another geographic survey of Co-type NHases was conducted by Rapheeha *et al.* (2017), who studied the substrate selectivity of both Fe- and Co-type NHases from various South African agricultural soil isolates, with a focus on the Co-types. The isolates which produced Co-type NHases were identified by 16s rRNA analysis as *R. rhodochrous,* except one (designated A99), which was identified as *Pimelobacter sp.* It was further found that both the Fe- and Co-type NHases specifically from *Rhodococci* were able to hydrolyse β-substituted aromatic nitriles (Chhiba *et al.*, 2012). It is the NHases from these isolates which were used in the present study. In the current study, the gene sequences of the L-NHases from these isolates were compared to the activity data to determine if any activity patterns could be correlated to genetic variation.

The closely related South African soil isolate, *R. rhodochrous* ATCC BAA-870, was studied to determine its enantioselectivity to various β-substituted hydroxynitriles and aminonitriles (Chhiba *et al.*, 2012). Despite the previous finding that an amino group decreased enantioselectivity (van Pelt *et al.,* 2011), enantioselectivity was still observed, in particular towards the *(S)*-enantiomer. However, it is not certain which residues in the protein are responsible for this characteristic.

2.1.3 Substrate channel amino acids may be a target for directed evolution

It is known that the conserved residues in the active site are those which are important to enzymatic capability and structural integrity and therefore remain unchanged (Huang *et al.*, 1997). However, crystal structures of NHase indicate that the enzyme contains a channel for substrate molecules at the interface of the alpha and beta subunit. Some of the residues in this channel are not conserved, suggesting a role in substrate specificity. These residues were identified by Huang *et al.* (1997) on a crystal structure of the Fe-NHase from *R. erythropolis* R312. The alignments constructed in this study were analysed to identify the Co-type residues which corresponded with the substrate channel residues in the Fe-type enzymes, in order to identify residues which may play a role in substrate specificity and possibly enantioselectivity. The non-conserved residues identified in the *R. erythropolis* R312 enzyme were αGln91, αTrp118, αSer114, βTyr37, βVal52, βArg56, βTyr67, βTyr72, βArg75, βTyr76 and βArg141. They concluded that the

majority of these residues were variable, with the only strictly conserved residues being βArg56, βTyr72 and βArg141, with the variable residues contributing to selectivity differences between NHases.

2.2 MATERIALS AND METHODS

2.2.1 Materials

2.2.1.1 Chemicals

Chemicals were obtained from Sigma-Aldrich (Now Merck; Modderfontein, South Africa) unless otherwise stated. Reagents for microbiology and protein biochemistry were obtained from BioRad (California, USA) and Lasec (Thermo Fisher Scientific, Massachusetts, USA).

2.2.1.2 Organisms and media

The South African agricultural soil isolates of *R. rhodochrous* (A12, A19, A29, A73, A98, A99) as well as ATCC BAA-870 were obtained from Dr Varsha Chhiba at the CSIR, Pretoria, South Africa. Cells containing recombinant nitrile hydratase from BAA-870 were provided by Dr Brandon Weber at the University of Cape Town.

2.2.1.3 Primers and sequencing services

Forward and reverse primer pairs for the low molecular mass and high molecular mass NHases of *R. rhodochrous* were previously designed by Frederick (2013) based on the low and high molecular weight cobalt NHases from *R. rhodochrous* J1. The primer sequences were designed to amplify the gene cassette consisting of the alpha subunit, beta subunit and chaperone and were as follows:

L-NHase forward primer: 5'-CATATGGATGGAATACACGACCTCGGT-3'
L-NHase reverse primer: 5'-AAGCTTCTACCCGTCGGAGTCAGTGGT-3'

H-NHase forward primer: 5'-CATATGGATGGTATACACGACACAGGC -3'
H-NHase reverse primer: 5'-AAGCTTTCAGTCGATGATGGCCAT -3'

PCR Primers were ordered from Inqaba Biotec (Pretoria, South Africa). Inqaba Biotec was also employed to sequence the identified NHases from the isolates and BAA-870. The sequencing method involved amplifying the gene from extracted gDNA in a PCR reaction using EconoTaq 2x GreenMasterMix (Lucigen, USA) and then sequencing using the BigDye version 3.1 (Thermo Fisher Scientific, Massachusetts, USA) kit.

2.2.2 Methods

2.2.2.1 Glycerol stocks

Overnight cultures of 100 ml of tryptone soya broth (Merck, New Jersey, USA) were inoculated with a single isolate colony (one colony per culture per isolate). The culture was grown overnight for approximately 16 h at 30°C with continuous shaking at 200 rpm. Following growth, 500 µL of cells were mixed with 500 µL of sterile 80% glycerol in a Cryo. S^{TM} cryogenic vial (Greiner Bio One, Frickenhausen, Germany). The vials were snap frozen in liquid nitrogen and stored at -80°C.

2.2.2.2 Extraction and identification of nitrile hydratase genes from South African agricultural soil isolates via polymerase chain reaction

For each isolate, a culture of 100 ml of tryptone soya broth was inoculated with 100 µL from a thawed glycerol stock. The culture was grown for approximately 16 h at 30°C and 200 rpm. Tryptone soya agar plates were then streaked with the culture and grown at 30°C for a maximum of 72 h. A single colony was selected from the plate and inoculated into a new 100 ml tryptone soya broth culture and grown for approximately 12 h.

Bacterial cells were pelleted in a MiniSpin Plus Microfuge (Eppendorf, Hamburg, Germany) and then resuspended in nuclease-free water. The cells were then either ground in liquid nitrogen or lysed with 2.5 $mg.ml^{-1}$ lysozyme (Sigma-Aldrich) in order to weaken the cell wall. Genomic DNA was then extracted using the ZR Fungal/Bacterial DNA MiniPrep™ kit (Zymo Research, California, USA).

The polymerase chain reaction method to amplify the low and high molecular mass NHases (if present) was modified from that of Frederick (2013). Taq polymerase (NEB, Massachusetts, USA) and a Bio-Rad C1000 Touch Cycler (Bio-Rad,California, USA) were used. Cycling conditions for the L-NHase PCR were as follows: initial denaturation at 94°C for 1 min; followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at (Tm - 5°C (62.7°C)) for 30 sec and extension at 68°C for 1 min/kb. In the case of L-NHase this was 1 min 30 sec. Cycling conditions for the H-NHase PCR were: initial denaturation at 94°C for 2 min; followed by 30 cycles of denaturation at 94°C for 30 s, annealing at (Tm - 5°C) for 40 s and extension at 68°C for 1 min/kb. In the case of H-NHase this was 1 min 40 s. BAA-870 was used as a positive control. A negative control containing no template DNA was used.
The H-NHase protocol was repeated using different PCR stabilisers: BioStab (Sigma-Aldrich) and Betaine (Sigma-Aldrich), as well as replacing the Taq buffer with GC Buffer and GC Enhancer (NEB, Massachusetts, USA).

2.2.2.3 Database mining for nitrile hydratase protein sequences

The BRENDA database is a curated database which aims to summarise functional data for enzymes, which includes solved structures, amino acid sequences, substrate profiles and kinetic parameters as reported by individual studies (Placzek *et al.*, 2017). It was chosen as the preferred source for nitrile hydratase sequences as all data is manually curated and redundancy is avoided.

The sequences for each set of BRENDA entries were downloaded from the UniProt database at http://www.uniprot.org/ by utilising the Retrieve/ID Mapping function. Microsoft Excel 2013 was then used to sort and filter the sequences and then edit them into a FASTA style (utilising “>” style headers) before each sequence list was converted into a FASTA file by copying into a Microsoft Notepad document with a “.txt” or “.fas” file extension.

2.2.2.4 Sequence alignments and analyses

The multiple sequence alignment programmes available on the EMBL-EBI website (https://www.ebi.ac.uk/Tools/msa/) were used to create multiple sequence alignments during this study (Chojnacki *et al.*, 2017). MAFFT was used to separately align the amino acid sequences of the alpha and beta subunits belonging to the South African isolates, as well as the alpha subunits when comparing the full *R.* genus. In all cases, default settings were used: BLOSUM62 matrix; 1.53 gap open penalty; 0.123 gap extension penalty; max iterations: 2. When aligning the beta subunits belonging to the *Rhodococcus* genus, Clustal Omega was used as it produced a better scoring alignment with fewer small gaps. The default settings for Clustal Omega were used.

2.2.2.5 Phylogenetic analysis

MEGA7 (Kumar, Stecher and Tamura, 2016) was used to create a neighbour-joining, unrooted phylogenetic tree of each subunit for both the alpha and beta subunits of all available nitrile hydratases. Neighbour joining was selected due to the large number of sequences. Bootstrapping was carried out (1000 replicates) to yield a bootstrap consensus tree. The evolutionary distances were computed using the Poisson method. The rate variation among sites was modelled with a gamma distribution (shape parameter = 1).

The initial trees were too extensive and thus a tree based on substrate profiles was created instead.

2.2.2.6 Molecular modelling

A homology model of the *R. rhodochrous* ATCC BAA-870 was constructed using Schrödinger Maestro v. 11.2 (Schrödinger, New York, USA) using the Structure Prediction Wizard and using PDB ID 4ob0 (Martinez *et al.*, 2014) as a template for both alpha and beta subunit. The alpha and beta subunits were prepared for homology modelling using the Maestro Protein Preparation Wizard (Sastry *et al.*, 2013).

Protein structural and biochemical data were predicted using the ExPASy ProtParam tool (Gasteiger *et al.*, 2005).

2.3 RESULTS

2.3.1 Identification of nitrile hydratase genes in South African agricultural soil isolates by polymerase chain reaction

Extracted genomic DNA was visualised on an agarose gel to confirm its quality prior to PCR analysis (Figure 2.1). The strong band at a high molecular weight (larger than 20000 bp) indicates that the majority of DNA was in the form of long strands. Smears indicate smaller DNA fragments and the small amount of smearing further confirmed the DNA quality.

Following this, separate polymerase chain reaction (PCR) reactions were prepared for each primer set and using the extracted gDNA as the template to determine which, if any, nitrile hydratase genes were present in each *R. rhodochrous* isolate. The isolates A12, A19, A29, A73 and A99 were found to contain L-NHases (Figure 2.2). None of the isolates appeared to contain an H-NHase. Attempts to stabilise the PCR reaction with additives due to the high-GC content of the *Rhodococcus* genome yielded only non-specific binding. It was assumed that the agricultural isolates did not contain an H-NHase gene, or if they did, its sequence was sufficiently dissimilar to that of *R. rhodochrous* J1 that the primers did not bind to the gene. The isolate A98 did not produce results for either an L-NHase or an H-NHase. Its growth pattern and colony morphology was sufficiently different from the other isolates that it was judged to be contamination and excluded from further studies.

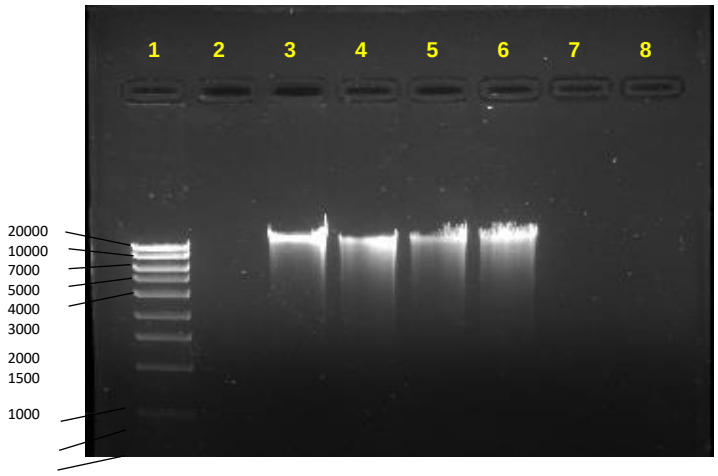


Figure 2.1: Confirmation of gDNA quality from *R. rhodochrous* soil isolates. gDNA was run on a 0.8% agarose gel at 90V for 1.5 hrs. The gel was stained with ethidium bromide. Lane 1, GeneRuler 1kb Plus DNA Ladder; Lane 3, A19; Lane 4, A29; Lane 5, A73; Lane 6, BAA-870.

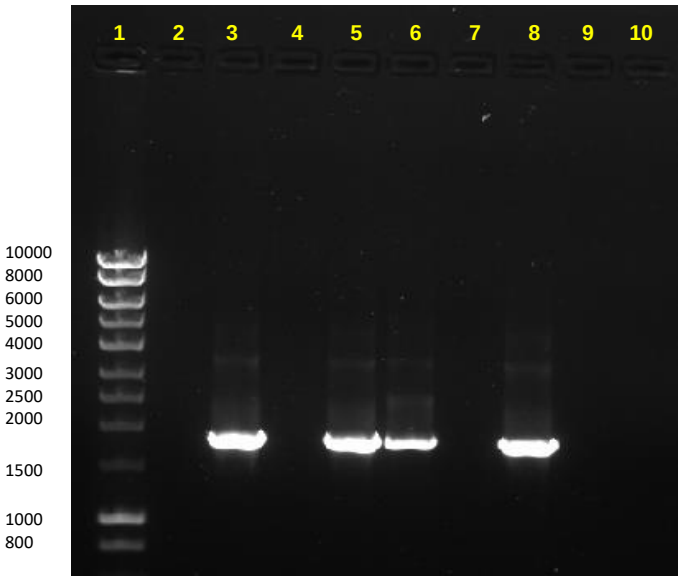


Figure 2.2: Example of PCR results showing the presence of L-NHase genes in some of the agricultural isolates as well as the absence of an H-NHase gene. Results were run on a 0.8% agarose gel at 90V for 1.5 hrs. The gel was stained with ethidium bromide. L: result of L-NHase PCR. H: result of H-NHase PCR. Lane 1, DNA Hyperladder I (bp); Lane 3, A73(L); Lane 4, A98(L) Lane 5, A99(L); Lane 6, A19(L); Lane 7, A19(H); Lane 8, Positive control - BAA-870(L); Lane 9, Negative control (L); Lane 10, Negative control (H).

2.3.2 Database mining overview

As of March 2018, a search for nitrile hydratases (EC 4.2.1.84) on the BRENDA database returned 1315 results for amino acid sequences. As most of these are separate alpha and beta subunit sequences belonging to the same isolate, the actual number of curated nitrile hydratase sequences available was initially assumed to be about half, or 657 sequences. Seventeen of the results had been declared redundant on UniProtKB (The UniProt Consortium, 2017), thus bringing the results to 640 nitrile hydratases.

Of the 1315 results, 32 were specifically identified as binding a Co^{2+} or Co^{3+} metal ion and 33 were reported to bind an Fe^{2+} or Fe^{3+} metal ion. However, only five of the sequences identified on BRENDA as containing a cobalt ion and eight of those containing an Fe ion have accession numbers on UniProtKB. It appears that not all identified nitrile hydratases have been satisfactorily classified as either Co- or Fe- type, while many of those identified do not have sequences submitted to UniProt KB. It was theorised that a full alignment of all of the available nitrile hydratase amino acid sequences would shed light on the classification of those not yet identified as Co-type or Fe-type, as well as which may be more closely related to the high molecular mass Co-type of *R. rhodochrous* J1. When extracting the sequences from UniProtKB, 140 sequences had been tagged as cobalt type on this database.

BRENDA reports 45 available crystal structures available on the Protein Data Bank (www.rcsb.org; Berman *et al.,* 2000), of which 28 of these are from *R. erythropolis* and four from *Comamonas testosteroni* (both Fe type); ten from *Pseudonocardia thermophila* and three from *Bacillus sp.* RAPc8 (*Geobacillus pallidus*), both cobalt type. Although there are no available crystal structures of cobalt type nitrile hydratases from *R. rhodochrous*, the literature reports other attempts to crystallise the cobalt type nitrile hydratases from other species, including *Pseudomonas putida* (Brodkin *et al.*, 2011).

2.3.3 The low molecular mass nitrile hydratases from various South African *Rhodococcus rhodochrous* isolates form a cluster with that of *Rhodococcus rhodochrous* J1

Separate alignments of both the alpha and beta subunits showed that overall the sequences are very similar, but appear to cluster into three groups based on a small number of residue differences. These clusters were apparent in both alignments.

2.3.3.1 Alpha subunits

The three clusters can be seen on an alignment of the sequences (Figure 2.3). One cluster (hereafter referred to as "Cluster 1") is formed by A12, A19 and A99. A second cluster ("Cluster 2") is formed by A73, BAA-870 and J1. A29 is distinct from the other two groups (hereafter referred to as "Cluster 3"). This is interesting, as J1 originates from Japan, which is a distant geographic location, yet it is more similar to the South African isolates, while A29 diverges the most from the other South African isolates.

At position α101, Clusters 1 and 3 have an aspartate, white Cluster 2 has a glycine, a non-conservative mutation. This position has not previously been reported as being important for catalysis. This residue is in fact on a loop on the outside of the protein, where it would interact with solvent (Figure 2.4). This loop is near to, but not within, the residues which form the substrate channel and thus it more than likely does not play a role in substrate acceptance. Cluster 3 also diverges from Clusters 1 and 2 at 5 positions: α176, α180, α187, α200 and α203. Positions α176 and α187, Cluster 3 has a valine at each position whereas Clusters 1 and 2 have an alanine at each position. At position α200, Cluster 3 has an alanine, while Clusters 1 and 2 have a valine. These are all conservative mutations. Furthermore the residues are all small and non-polar and are also found towards the outside of the alpha subunit, usually in coil regions and distant from the active site. Hence they are likely not vital for catalysis or substrate selectivity.

At position α203, Cluster 3 has an alanine while Clusters 1 and 2 have a threonine, a non-conservative mutation. The threonine is bulkier, as well as having a polar –OH group. Although not likely important for catalysis or selectivity, a mutation in a surface residue to a more polar or charged residue may cause a slight improvement in the solubility or stability of the protein. It was hypothesised that the seemingly interchangeable conservative glycine and alanine substitutions may make a small difference to the protein's aliphatic index, where a slightly larger non-polar aliphatic side chain may contribute to increasing the aliphatic index (and thus stability) of the protein. Aliphatic index and instability index were computed using ExPASy ProtParam (Gasteiger *et al.*, 2005). Cluster 1 and cluster 2 were found to have the same aliphatic index of 83.82 while cluster 3 had a higher aliphatic index at 85.22, which is likely due to the valines at positions α176 and α187. Cluster 3 did however have a slightly higher instability index at 37.41, whereas clusters 1 and 2 had an instability index of 35.92 and 35.82 respectively, indicating that cluster 3 was slightly less stable. The ProtParam values are summarised in Table 2.1.

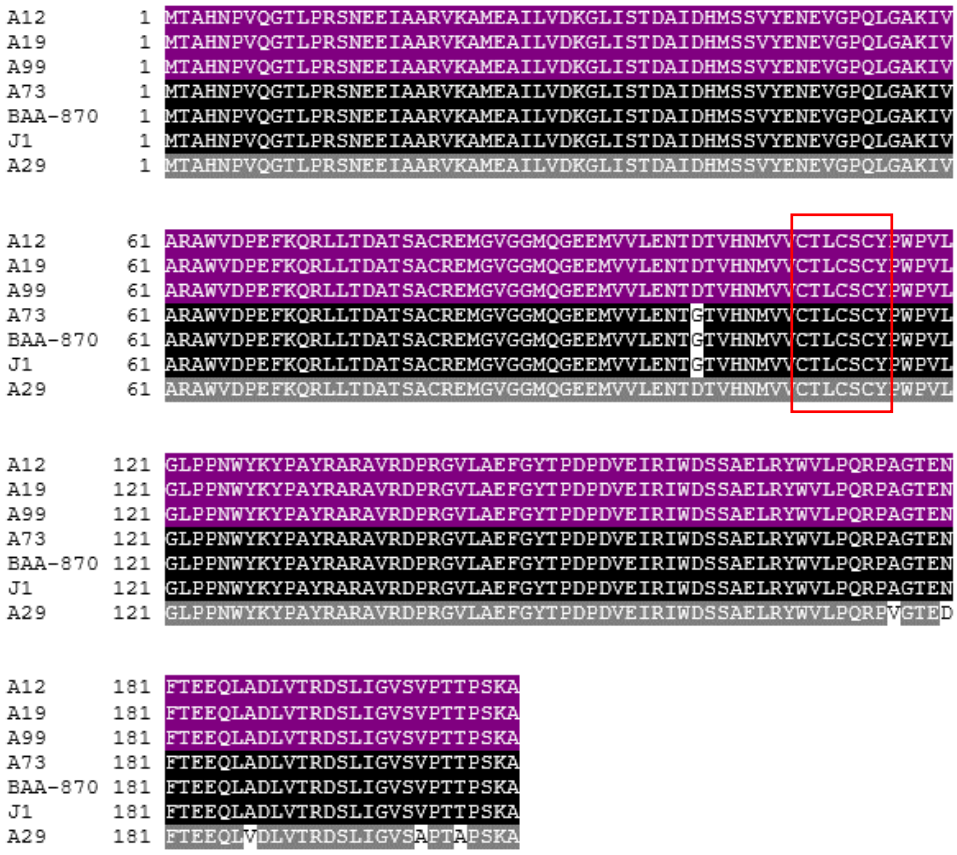


Figure 2.3: Alignment of nitrile hydratase alpha subunits. MAFFT on the EMBL-EBI server (Katoh and Standley, 2013) was used to align the amino acid sequences of the NHase alpha subunits from South African *R. rhodochrous* soil isolates and that from *R. rhodochrous* J1. Three clusters are evident: Purple, Cluster 1; Black, Cluster 2; Grey, Cluster 3. A12-A99: South African agricultural isolates. BAA-870: *R. rhodochrous* ATCC BAA-870, original South African *R. rhodochrous* soil isolate, characterised by Frederick (2013). White: non-conservative mutations. Red box: active site residues.

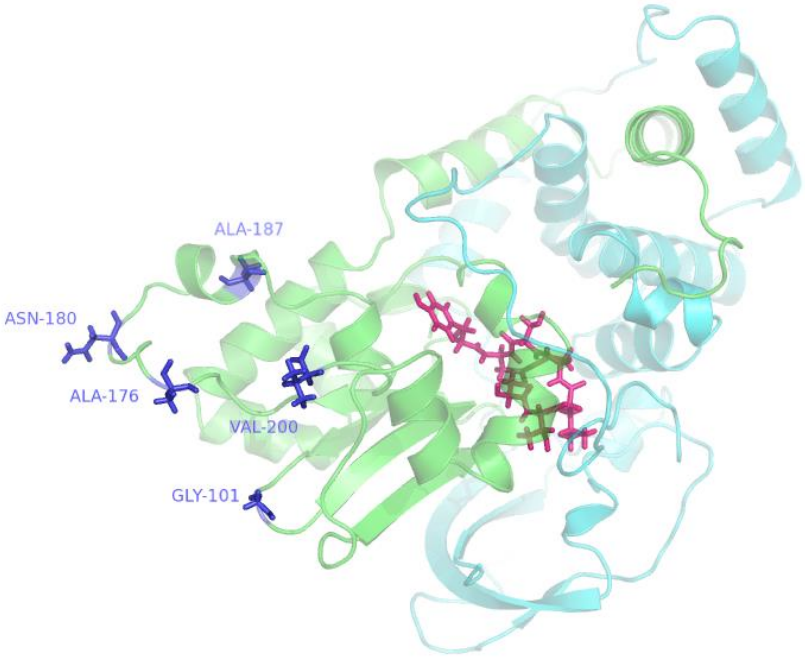


Figure 2.4: Alpha subunit residues which show variation between the three identified *R. rhodochrous* clusters. A homology model of *R. rhodochrous* ATCC BAA-870 was made based on PDB ID 4OB0 (Martinez *et al.*, 2014) using Schrödinger Maestro v. 11.2 (Schrödinger, New York, USA). Pymol v. 0.99 was used to visualise the model. Pink: active site residues. Green: alpha subunit. Light blue: Beta subunit. Dark blue: Cluster-specific residue variations. Residue α203 is not shown as it was not produced in the homology model.

2.3.3.2 Beta subunits

An alignment of the amino acid sequences of the beta subunit shows the same three phenotypic clusters as seen from the alignment of the alpha subunits, although the beta subunits have a greater number of variable residues than the alpha subunits. Cluster 1 contains eight residue variations, at positions β16, β17, β105, β113, β133, β136, β148 and β175. Cluster 2 has three unique residue mutations, at positions β88, β104 and β189. Cluster 3 differs from the other two clusters at position β143 only (Figure 2.5). Like the variations in the alpha subunits, all of these variations are found on the outside of the protein, in random coil regions and are distant from the active site and substrate channel. Thus it is likely that these amino acid differences do not play a role in substrate selectivity or enantioselectivity. These residues are visualised on a molecular model of the BAA-870 NHase in Figure 2.6.

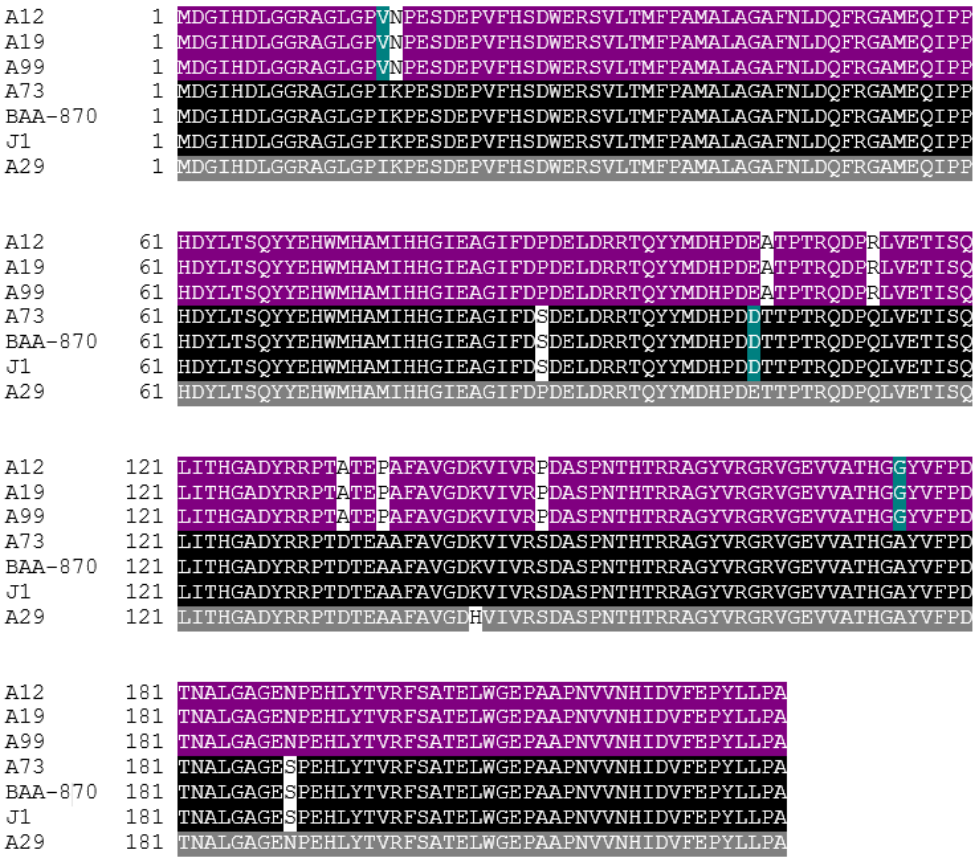


Figure 2.5: Alignment of nitrile hydratase beta subunits. MAFFT on the EMBL-EBI server (Katoh and Standley, 2013) was used to align the amino acid sequences of the NHase beta subunits from South African *R. rhodochrous* soil isolates and that from *R. rhodochrous* J1. A12-A99: South African agricultural isolates. BAA-870: *R. rhodochrous* ATCC BAA-870, original South African *R. rhodochrous* soil isolate, characterised by Frederick (2013). Three clusters are evident: Purple, Cluster 1; Black, Cluster 2; Grey, Cluster 3. Blue: Conservative mutations. White: Non-conservative mutations.

An ExPASy ProtParam analysis of each cluster indicated that the beta subunits all had lower aliphatic indices and higher instability indices than the alpha subunits (summarised in Table 2.1) and were therefore less stable than the alpha subunits. If the beta subunit is more prone to mutation than the alpha subunit this may result in a larger number of "silent" mutations which only affect stability. It has indeed previously been suggested that the source of instability is the beta subunit as it was mostly expressed in the insoluble fraction (Frederick, 2013). It may be that the beta subunit is also the source of substrate selectivity and enantioselectivity differences between NHases, but this remains to be determined.

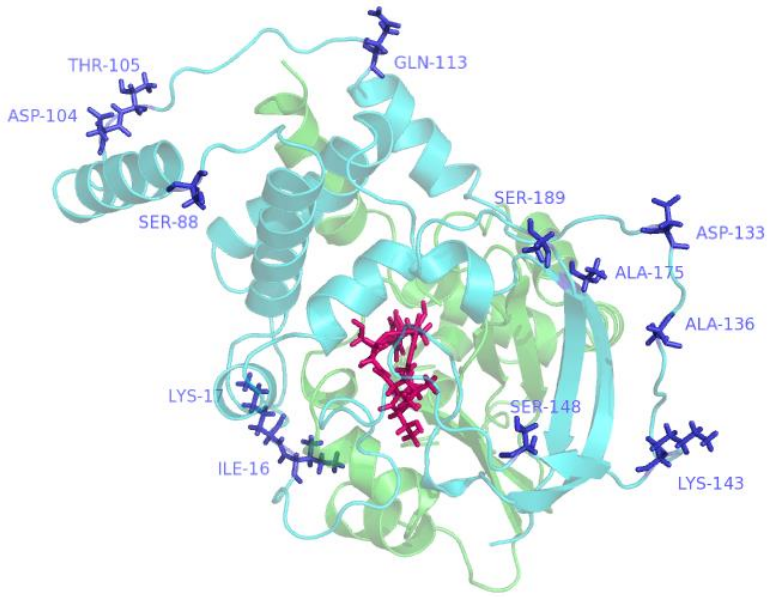


Figure 2.6: Beta subunit residues which show variation between the three identified *R. rhodochrous* clusters. A homology model of *R. rhodochrous* ATCC BAA-870 was made based on PDB ID 4OB0 (Martinez *et al.*, 2014) using Schrödinger Maestro v. 11.2 (Schrödinger, LLC, 2018). Pymol v. 0.99 was used to visualise the model. Pink: active site residues. Green: alpha subunit. Light blue: Beta subunit. Dark blue: Cluster-specific residue variations.

Table 2.1: Summary of ExPASy ProtParam stability scores for *R. rhodochrous* cobalt type nitrile hydratases

Isolate	Instability Index		Aliphatic Index	
	Alpha Subunit	**Beta Subunit**	**Alpha Subunit**	**Beta Subunit**
Cluster 1 (A12, A19, A99)	35.92	41.76	83.82	69.96
Cluster 2(A73, J1, BAA-870)	35.82	43.47	83.82	70.40
Cluster 3 (A29)	37.41	43.47	85.22	70.40

These results were then compared to the activity of the agricultural soil isolates towards various substrates (aliphatic, aromatic and aryl-aliphatic), as determined by (Rapheeha *et al.*, 2017). There was no clear pattern of substrate usage which corresponded to the amino acid sequence patterns identified here. However, the study by Rapheeha *et al.* may have given variable results, as whole cells were grown on solid minimal media containing substrates. It is possible that targeted expression of the NHase protein and subsequent reactions carried out using the cell lysate may have yielded different results. Furthermore, the kinetic parameters such as K_m, V_{max} and K_{cat} of each NHase towards these substrates may show variations which would not be picked up using a rough screen.

As a further analysis, the substrate channel residues identified by Huang *et al.* (1997) were visualised on the homology model of *R. rhodochrous* ATCC BAA-870 to determine where these residues were located on a Co-type NHase (Figure 2.7). The residue numbers which corresponded in the Co-type sequence to those identified in the Fe-type *R. erythropolis* R312 were determined by Chhiba (2017; unpublished data). The residues in the Co-type NHase from *Rhodococcus rhodochrous* ATCC BAA-870 are αGln90, αVal108-αCys114, αTrp117, αArg167, These residues were also located in the substrate channel on the Co-type NHase and are therefore candidates for directed evolution, except for αSer113, αTrp117, βArg52 and βTyr68, all of which have previously been identified as necessary for correct enzyme function (Mitra and Holz, 2006), as mutation of these residues lead to drastically decreased enzymatic activity.

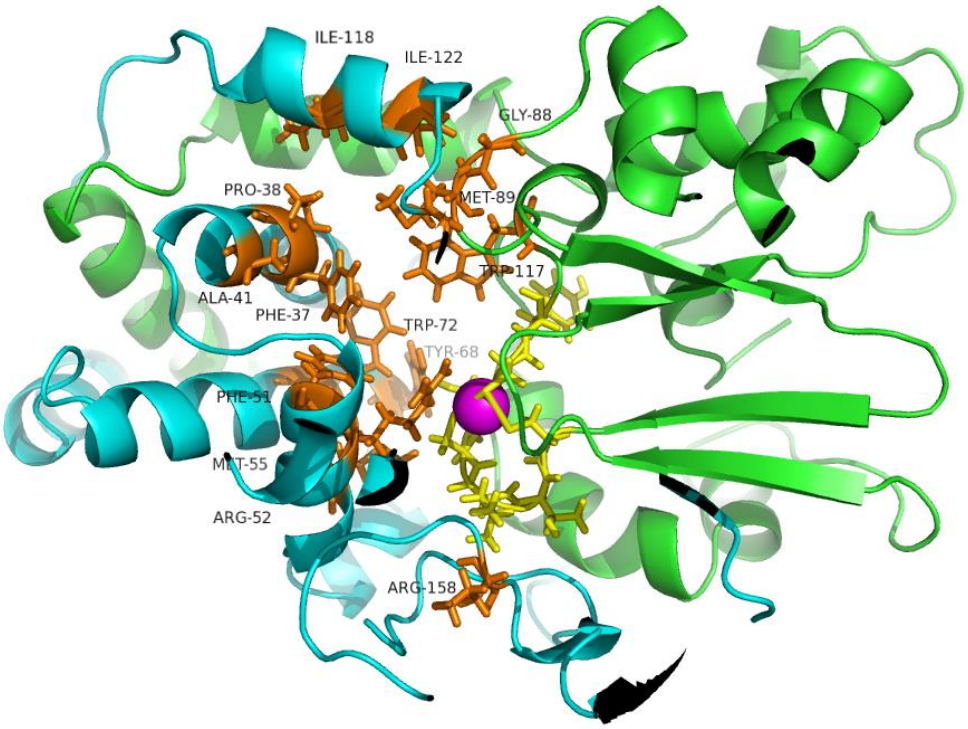


Figure 2.7: Visualisation of substrate channel amino acids in the Co-type low molecular mass NHase from *R. rhodochrous* ATCC-BAA 870. Alpha subunit: green; Beta subunit: blue; active site residues: yellow; Co ion: magenta; substrate channel residues: orange. The substrate channel residues were identified by Chhiba (2017; unpublished data) and from an alignment between the Fe-type NHase from *R. erythropolis*, PDB ID: 1AHJ (Huang et al., 1997) and the Co-type NHase from *R. rhodochrous* J1 (Kobayashi *et al.*, 1991). Homology model of *R. rhodochrous* ATCC BAA-870 made based on PDB ID 4OB0 (Martinez *et al.*, 2014) using Schrödinger Maestro v. 11.2. Pymol v. 0.99 was used to visualise the model.

2.3.4 Comparison of nitrile hydratase sequences within the *Rhodococcus* genus

The BRENDA results were filtered to identify the low molecular mass nitrile hydratases from the *Rhodococcus* genus only. The results were disproportionate, as there were 12 results for alpha subunits and 60 results for beta subunits. The excess of beta sequences were mainly from *R. erythropolis* and various unidentified *Rhodococcus* species.

2.3.4.1 *Rhodococcus* alpha subunits

MAFFT (EMBL-EBI) was used to align all available alpha subunit sequences from *Rhodococci,* including the South African isolates used in this study. Similarity and identity matrices were constructed using the Sequence Identity and Similarity (SIAS) tool at http://imed.med.ucm.es/Tools/sias.html (Reche, 2017). Identity refers to a residue at a specific position being identical in different sequences, while similarity refers to a position which is conservatively substituted by a residue with close physicochemical properties. The scores indicated that the mean pairwise identity between the alpha subunits was 62.85%. The mean pairwise similarity between the alpha subunits was 71.52%. The lowest pairwise identity score of 42.66% and the lowest similarity score of 55.50% both occurred between *R. sp.* EsD8 (N1M888) and *R. jostii* (strain RHA1) (Q0SJU0). The pairwise identity is visualised in Figure 2.8. The isolates in this study showed not more than 52% identity and not more than 63% similarity to any other available NHase alpha subunit sequence.

It was theorised that residues which played a role in substrate specificity and enantioselectivity would potentially be the less conserved residues which were near but not directly next to the active site residues. This was based on the findings by Brodkin *et al.*, (2011, 2015) that distal residues are necessary for substrate specificity in the Co-type NHase of *Pseudomonas putida*, as well as on the findings by Morley and Kazlauskas (2005) that residues within 10 Å tend to influence enantioselectivity. Based on this alignment, residues α101-α105, α126, α130, α131, α132 and α139 may be potential targets for mutagenesis as they have lower conservation scores. This may be confirmed using molecular modelling by docking potential substrates and identifying if they interact with these residues. Although residue α117 is very close to the active site residues, it is interesting to note that in the Fe-type alpha subunits, it is an alanine, while in the Co-types it is a proline, which may have some implications for substrate selectivity by causing increased steric hindrance with the presence of the proline in the Co-types. Swiss PDB Viewer (Guex and Peitsch, 1997) was used to visualise the amino acids within 10Å of the active site Co ion in *R. rhodochrous* ATCC Baa-870 and it was confirmed that some of the predicted residues (or their near neighbours) fell within this region, supporting the idea that they influenced substrate specificity and were potential mutagenesis targets (Figure 2.10).

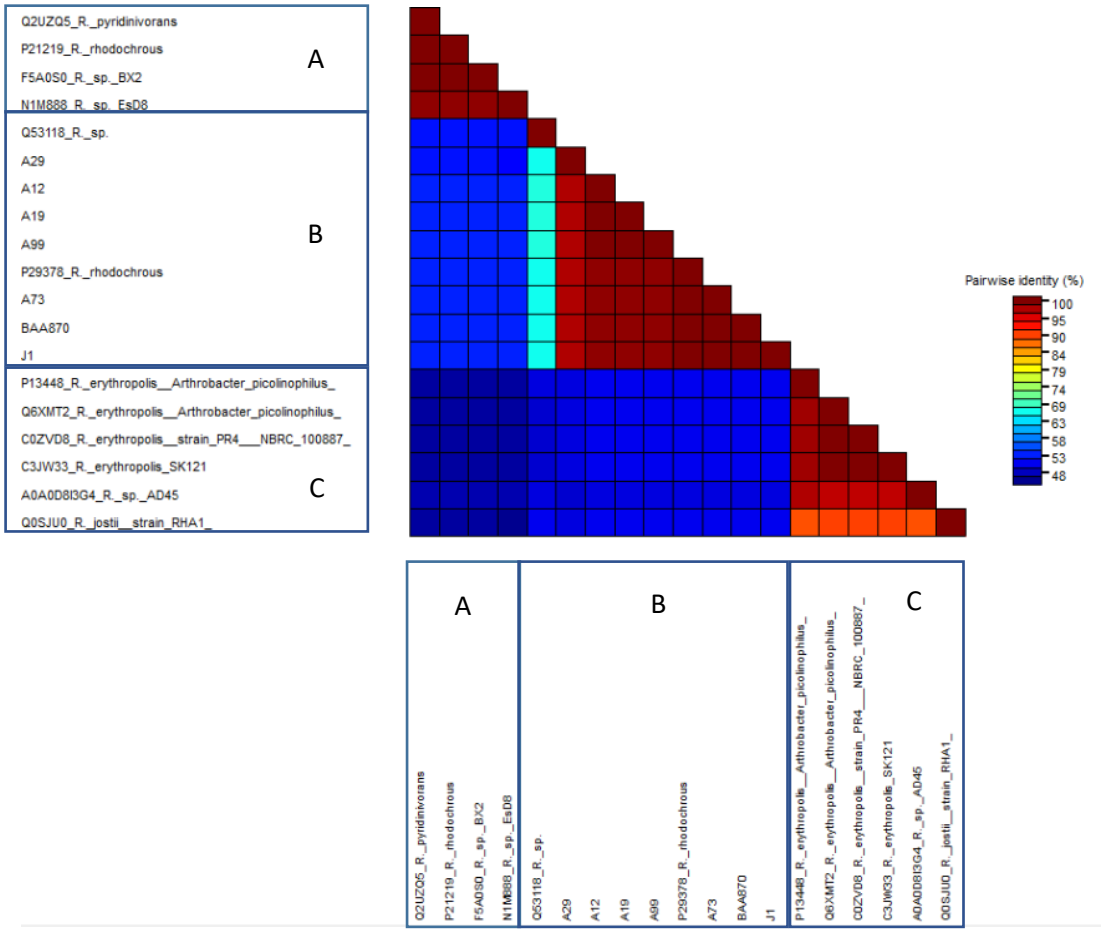


Figure 2.8: Pairwise identity matrix constructed from a MAFFT sequence alignment of the alpha subunits from various *Rhodococci.* SDTv1.2 (Muhire et al., 2014) was used to produce the matrix. A, sequences similar to the Co-type H-NHase from *R. rhodochrous* J1; B, Co-type L-NHases; C, Fe-type NHases.

The subunits were then aligned using the PRALINE multiple sequence alignment server (http://www.ibi.vu.nl/programs/pralinewww/), in order to examine the extent of conservation and other structural characteristics of the *Rhodococcal* NHase alpha subunits. Figure 2.9 shows the conservation of residues for an excerpt of the alignment, centred on the active site. It can be seen that the active site sequence (CT/SLCSCY/T) is conserved and the Fe-type and Co-types are identifiable based on this. The residues which surround this site appear to be quite highly conserved as well, with identity and similarity decreasing from approximately eight residues on either side.

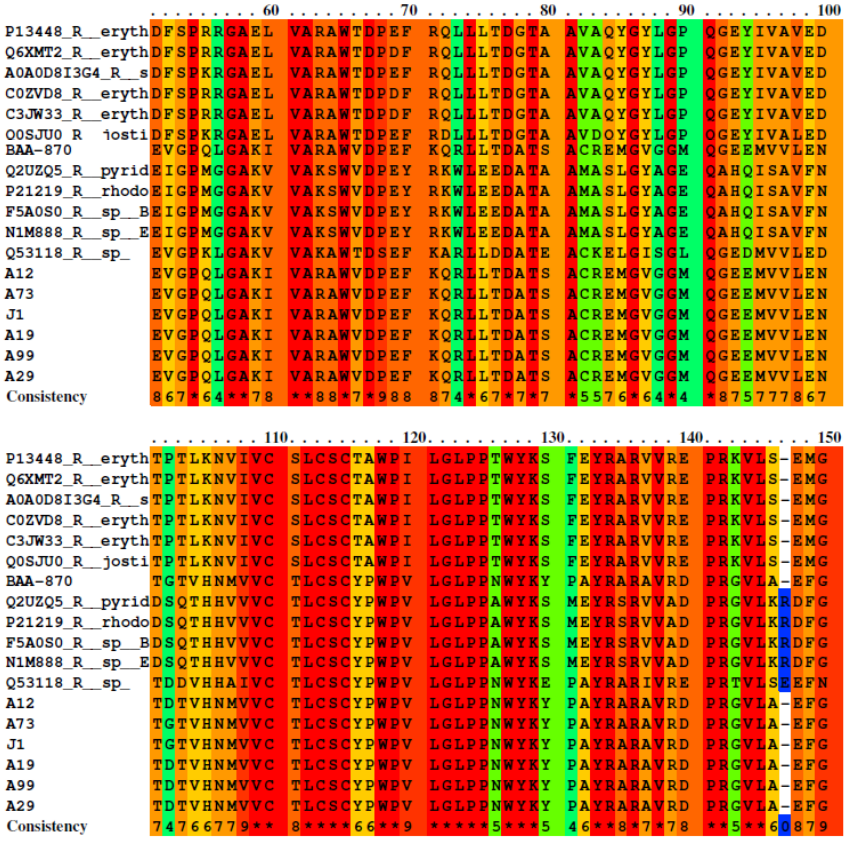


Figure 2.9: Alignment of *Rhodococcus* alpha subunits showing conservation of residues for an excerpt of the alignment, centred on the active site. Red: most conserved. Dark blue: least conserved. The PRALINE server at http://www.ibi.vu.nl/programs/pralinewww/ was used to create the alignment (Simossis and Heringa, 2003). Sequences possessing an Arg (R) or Glu (E) residue at position 147 are similar to the Co-type high molecular mass NHase from *R. rhodochrous* J1 (P21219).

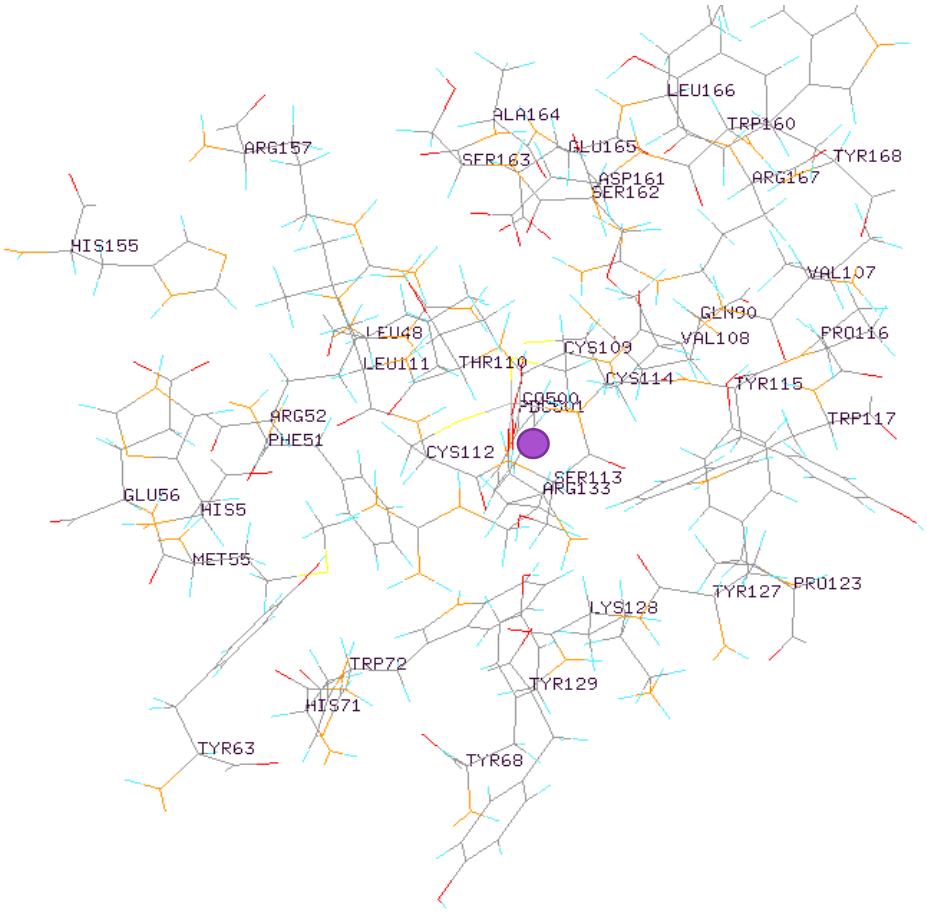


Figure 2.10: Schematic diagram showing the Co ion in the active site of *R. rhodochrous* ATCC BAA-870 and the residues within 10 Å of the ion. Homology model of *R. rhodochrous* ATCC BAA-870 made based on PDB ID 4OB0 (Martinez *et al.*, 2014) using Schrödinger Maestro v. 11.2. Swiss PDB Viewer (Guex and Peitsch, 1997) was used to visualise the model.

The hydrophobicity of the identified residues was also analysed based on the PRALINE alignment (Figure 2.11). Of these residues, α104, α126, α131 and α132 showed that as well as being variable sites, the variations also led to a reversal of hydrophobicity, which could alter the interactions occurring with substrates at those points, hence changing the substrate specificity of the enzyme. Once again, this remains to be tested.

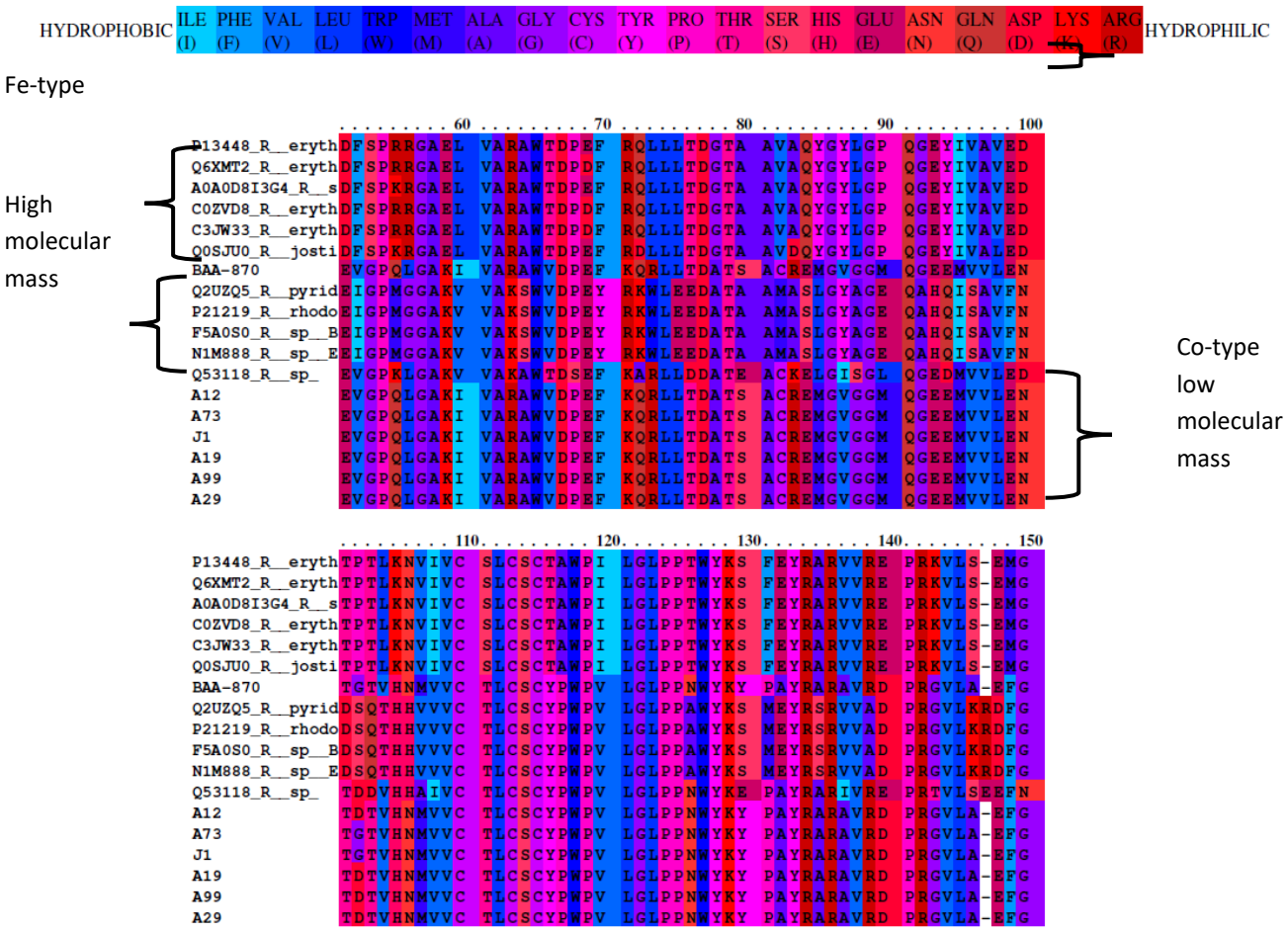


Figure 2.11: Alignment of *Rhodococcus* alpha subunits showing hydrophobicity of residues for an excerpt of the alignment, centred on the active site. Hydrophobicity for each residue is indicated on the key. The PRALINE server at http://www.ibi.vu.nl/programs/pralinewww/ was used to create the alignment (Simossis and Heringa, 2003). Sequences possessing an Arg (R) or Glu (E) residue at position 147 are similar to the Co-type high molecular mass NHase from *R. rhodochrous* J1 (P21219).

2.3.4.2 *Rhodococcus* beta subunits

All available *Rhodococcus* NHase beta subunit sequences were then compared in the same way as the alpha subunits (Section 2.4.1). ClustalOmega (EMBL-EBI) was used for the initial alignment. The alignment was visualised using the Sequence Demarcation Tool (Muhire et al., 2014), presented in Figure 2.12. Pairwise identity and similarity scores were again calculated using SIAS to determine the percentage of conserved residues and residues which were conservatively substituted.

The mean pairwise identity was 51.55% and the mean pairwise similarity was 65.63%, both scores lower than the corresponding scores for the alpha subunits, indicating that the beta subunits have greater variation. The minimum pairwise similarity was 27.19%, between *R. pyridinivorans* (Q2UZQ6) and *R. erythropolis* (*Arthrobacter picolinophilus*) (P13449). The minimum pairwise identity score of 45.6% occurred between *R. pyridinivorans* (Q2UZQ6) and a different *R. erythropolis* strain, strain PR4 / NBRC 100887 (C0ZVD9). The *R. rhodochrous* sequence (accession number P21220) with low pairwise identity to the South African isolates was the high molecular mass nitrile hydratase from *R. rhodochrous* J1 (Kobayashi *et al.,* 1991; Nagasawa, *et al.*, 1991).

Cluster 1 of the South African isolates in this study showed a maximum of 79.07% similarity to *R. opacus* M213, while clusters 2 and 3 showed approximately 78% similarity. The maximum identity was also linked to *R. opacus* M123, with cluster 1 showing 71.12% identity and clusters 2 and 3 showing 69.87% identity. It was interesting to note that this sequence is the only other *Rhodococcus* beta subunit sequence with such high scores compared to the isolates in this study. The similarity and identity of the South African isolates to all other available *Rhodococcus* beta subunit sequences was much lower, averaging around 35% identity and 55% similarity. There is unfortunately little other data about the *R. opacus* M123 nitrile hydratase, as it was identified during a whole genome sequencing study of the organism (Pathak *et al.,* 2013). Hence the protein has not been physically expressed or characterised and its substrate specificity is unknown. The alpha subunit was also not reported.

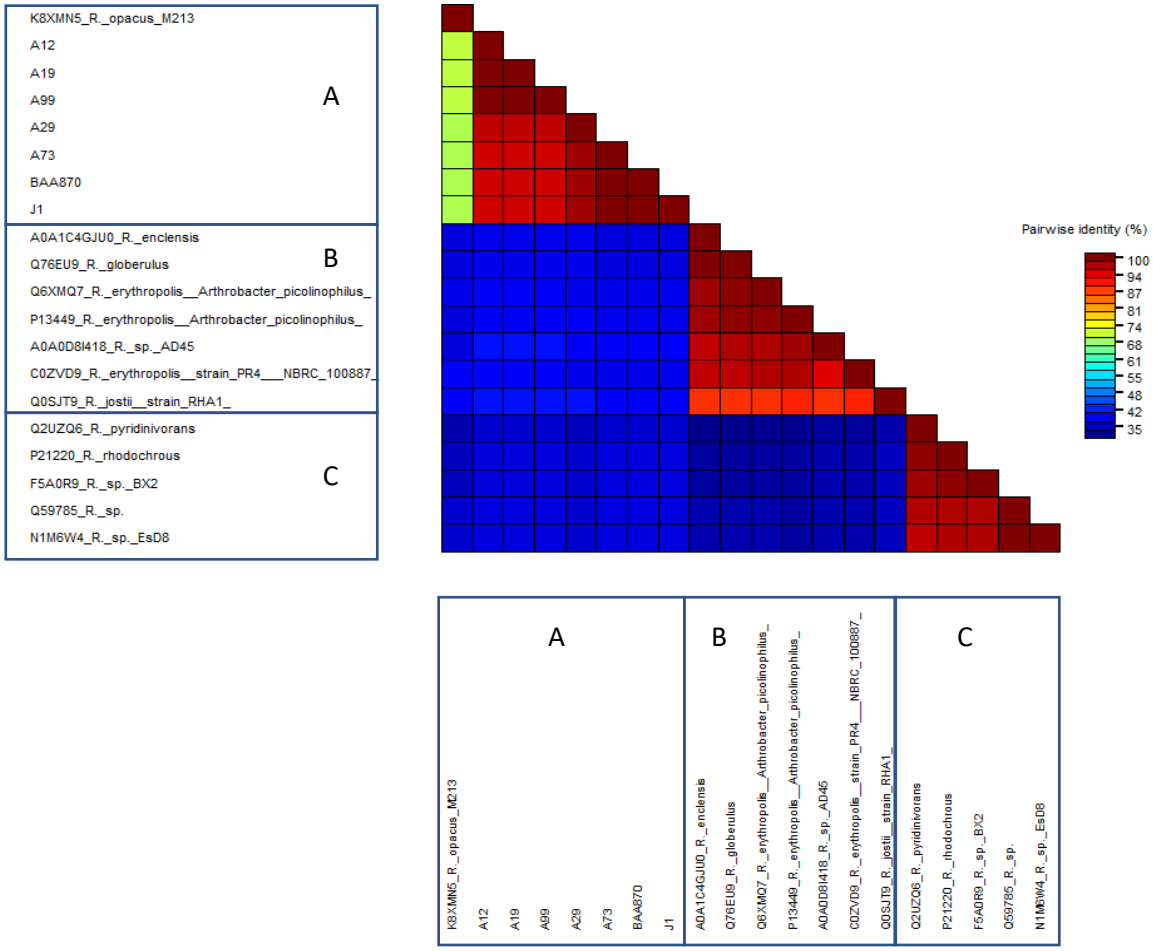


Figure 2.12: Pairwise identity matrix constructed from a MAFFT sequence alignment of the beta subunits from various *Rhodococci.* SDTv1.2 (Muhire et al., 2014) was used to produce the matrix. A, Co-type L-NHases; B, Fe-type NHases; C, sequences similar to the Co-type H-NHase from *R. rhodochrous* J1.

As with the alpha subunits, an alignment of the beta subunits was created using the PRALINE alignment program (Simossis and Heringa, 2003). These alignments showed that the overall conservation was lower than that of the alpha subunits. There were also more gaps in the beta subunits, with both single-residue gaps and longer gaps present. From the patterns in the alignment, the Fe-type, Co-type and high molecular mass NHases could be identified (Figure 2.13). However, due to the large number of less conserved residues, it was not possible to determine those which could be used as targets for directed evolution.

When comparing the hydrophobicity of the *Rhodococcus* beta subunits, it was also possible to identify the type of NHase based on the patterns of hydrophobicity (Figure 2.14). It was theorised that residues which caused a reversal in hydrophobicity were potential targets for mutagenesis.

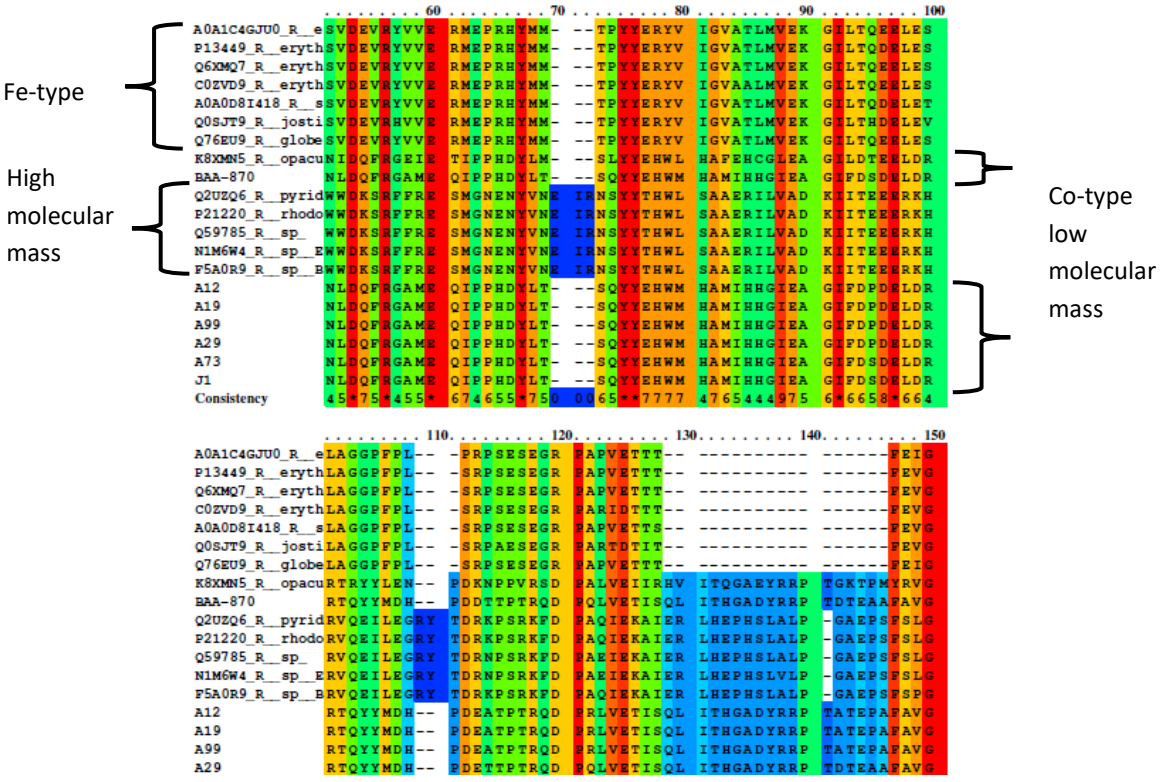


Figure 2.13: Alignment of *Rhodococcus* beta subunits showing conservation of residues for an excerpt of the alignment, excluding the N- and C-termini. Red: most conserved. Dark blue: least conserved. The IBIVU PRALINE server at http://www.ibi.vu.nl/programs/pralinewww/ was used to create the alignment (Simossis and Heringa, 2003).

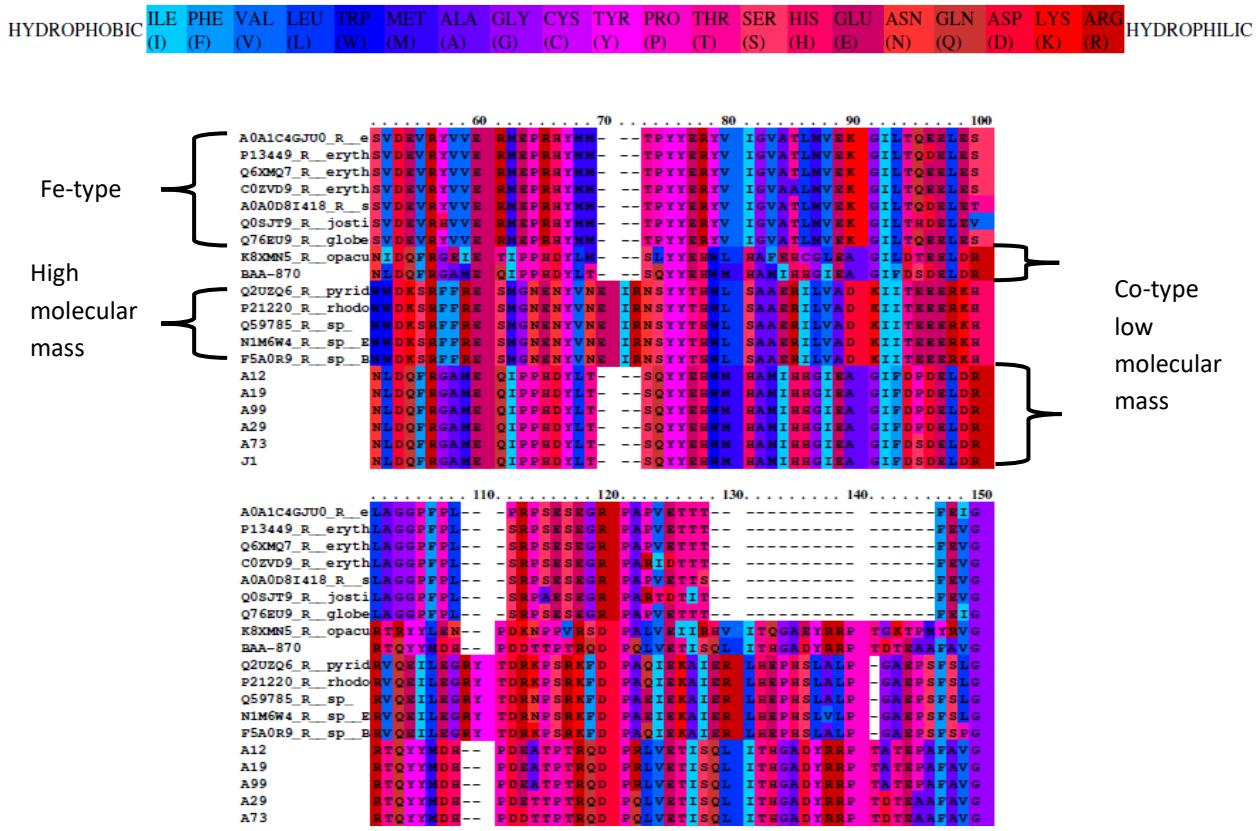


Figure 2.14: Alignment of *Rhodococcus* beta subunits showing hydrophobicity of residues for an excerpt of the alignment, excluding the N- and C-termini. Hydrophobicity for each residue is indicated on the key. The PRALINE server at http://www.ibi.vu.nl/programs/pralinewww/ was used to create the alignment (Simossis and Heringa, 2003).

2.3.5 Commentary on the origins of the high molecular mass cobalt type nitrile hydratase

The comparisons of both the beta subunits and the alpha subunits showed that the *R. pyridinivorans* sequences (Q2UZQ5 and Q2UZQ6) are very similar to the alpha and beta subunit sequences for the H-NHase from *R. rhodochrous* J1 (P21219 and P21220). It is possible that the H-NHase in one of these species was acquired via horizontal gene transfer (HGT) from the other. It was previously suggested that the H-

NHase arose from a duplication of the L-NHase gene within *R. rhodochrous* J1 *(*Wieser *et al.,* 1998*;* Frederick*,* 2013*)* and the H-NHase operon does contain a transposase, an enzyme which allows for gene rearrangement *(*Komeda *et al.,* 1996*)*. Horizontal transfer of virulence genes located on a virulence plasmid has previously been reported between Rhodococci *(*Savory *et al.,* 2017*)*. A gene to degrade organic sulphur-containing compounds was found to have been acquired by *Rhodococcus sp.* (Eu-32) via HGT from another Rhodococcus *(*Akhtar *et al.,* 2015*)*. Furthermore, HGT from other bacterial species to Rhodococci has also been reported (O'Mahony *et al.,* 2005*)*. The results in the present study show a greater similarity between the J1 H-NHase and other H-NHase genes, and a lower similarity between the L-NHase and H-NHase both from J1, therefore it is more likely that *R. rhodochrous* J1 acquired its H-NHase gene via HGT from another *Rhodococcus* or another bacterial species.

2.3.6 Phylogenetic analysis of nitrile hydratases based on substrate usage

The final analysis involved the construction of a phylogenetic tree, which was then annotated based on available activity data to determine if any patterns in substrate selectivity and activity (or enantioselectivity, if available) correlated with phylogenetic dispersion. A tree was constructed only for the beta subunits, as these show the greatest variation which is believed to contribute to substrate selectivity. All nitrile hydratase sequences curated on BRENDA and available from UniProt were included, but high molecular mass nitrile hydratases were excluded based on their low similarity to the low molecular mass nitrile hydratase. Due to the large number of results, in some cases representatives of a species or genus were selected in order to keep the sequence alignments at a maximum of 150 sequences per alignment. All sequences specifically identified as cobalt-containing in the BRENDA entry name were included in the phylogeny, as well as all *Rhodococci*, except *R. erythropolis*, where representative sequences were chosen due to the large number of *R. erythropolis* sequences available. There were 140 BRENDA results for this search. Of these 140 results, 20 were identified as high molecular mass NHases and were removed. Of the remaining 120, 90 were identified specifically as alpha subunits and 30 as beta subunits. A further three sequences were removed as they were incorrectly named as cobalt type in the UniProt database.

Alignments were done using Expresso on the T-Coffee alignment server (http://tcoffee.crg.cat/), which uses an algorithm to identify conserved domains, which is useful for sequences with low identity and similarity and where large gaps occur (Di Tommaso *et al.*, 2011). The alignment of the alpha subunits

showed the greatest variation at the C and N termini, which was expected as these regions formed random coils and are not known to be vital to activity, hence the high variation. An excerpt of the alpha subunit alignment is shown in Figure A1 (see Appendix). The beta subunits showed a similar pattern of high variation and poor consensus at the C and N termini, and greater similarity towards the centre of the sequences. However overall there were many more small gaps within the body of the alignment, with numerous conserved regions which were smaller than the conserved regions in the alpha subunits. An excerpt of the beta subunit alignment is shown in Figure A2 (see Appendix). A neighbour joining phylogenetic tree was then constructed for each subunit using Mega7: Molecular Evolutionary Genetics Analysis version 7.0 for Bigger Datasets (Kumar *et al.*, 2016).

The tree showed an unexpected distribution, as the *Pseudomonas* and *Pseudonocardia* species did not cluster together exclusively, but formed multiple dispersed clusters, with some clusters containing both *Pseudomonas* and *Pseudonocardia* (Figure A3 in Appendix). This could be due to a number of possibilities: 1) There are more species of *Pseudomonas* and *Pseudonocardia* which have been characterised than species in other genera, thus showing greater variability in NHase sequences; 2) The organisms themselves were incorrectly classified and are in fact other species and/or genera; or 3) the NHases in these and possibly other species arose from horizontal gene transfer, resulting in similar NHases while the rest of the genome, including the 16s rRNA genes are less divergent – hence a 16S rRNA phylogeny would potentially yield two clusters, one of *Pseudomonas* and one of *Pseudonocardia.*

It was then decided to construct a tree focused only on those sequences with reported activity data and then to attempt to find patterns of amino acid residues linked to substrate usage. The BRENDA database has a comprehensive list of substrates linked to the species they were tested on. There are 729 unique entries in the database. However, few of these entries had associated sequence data. When looking at the available sequences on UniProt, only 20 of these were linked to substrate analyses and most of these were simple substrates, not enantiomers, so no comparisons could be made based on enantioselectivity (Table 2.2). These BRENDA results included many *Rhodococcus* results, including those for some enantiomers. The reader is referred to BRENDA for further details on these substrates. The *Rhodococci* had no associated sequence data. Therefore the *Rhodococcus* sequences used in Section 2.3.4 were included and aligned to the sequences with known substrate data.

Table 2.2: Linked nitrile hydratase sequence and substrate data available on the BRENDA database

UniProt Entry	Organism	Protein names	Substrates	Products
Q8GJG6, Q8GJG7	*Nocardia sp.* YS-2002	Nitrile hydratase alpha subunit, Nitrile hydratase subunit beta (NHase) (EC 4.2.1.84)	acrylonitrile	acrylamide
Q7SID2, Q7SID3	*Pseudonocardia thermophila*	Cobalt-containing nitrile hydratase subunit alpha (L-NHase) (L-nitrilase) (EC 4.2.1.84), Cobalt-containing nitrile hydratase subunit beta (L-NHase) (L-nitrilase) (EC 4.2.1.84)	3-cyanopyridine	nicotinamide
			4-cyanopyridine	isonicotinamide
Q5XPL5, Q5XPL4	*Comamonas testosteroni* (*Pseudomonas testosteroni*)	Nitrile hydratase alpha subunit (EC 4.2.1.84), Nitrile hydratase subunit beta (NHase) (EC 4.2.1.84)	3-cyanopyridine	pyridine-3-carbamide
			3-hydroxybutyronitrile	3-hydroxybutyramide
			3-hydroxypropionitrile	3-hydroxypropionamide
			3-hydroxyvaleronitrile	3-hydroxyvaleramide
			acrylonitrile	acrylamide
			adiponitrile	adipamide
			butyronitrile	butyramide
			glycolonitrile	glycolamide
			methacrylonitrile	methacrylamide

Table 2.2 cont.: Linked nitrile hydratase sequence and substrate data available on the BRENDA database

UniProt Entry	Organism	Protein names	Substrates	Products
Q2UZQ5, Q2UZQ6	*Rhodococcus pyridinivorans*	Nitrile hydratase alpha subunit (EC 4.2.1.84), Nitrile hydratase subunit beta (NHase) (EC 4.2.1.84)	acrylonitrile	acrylamide
P13448	*Rhodococcus erythropolis* (*Arthrobacter picolinophilus*)	Nitrile hydratase subunit alpha (NHase) (Nitrilase) (EC 4.2.1.84)	(R)-2-chloromandelonitrile	(R)-2-chloromandelamide
			2-hydroxy-4-phenylbutyronitrile	2-hydroxy-4-phenylbutyramide
			3-cyanopyridine	pyridine-3-carbamide
			3-phenoxymandelonitrile	3-phenoxymandelamine
			4-methylmandelonitrile	4-methylmandelamine
			acetonitrile	acetamide
			acrylonitrile	acrylamide
			benzonitrile	benzamide
			methacrylonitrile	methacrylamide
			propionitrile	propionamide
K9NKC3, K9NLN7	*Pseudomonas* sp. UW4	Nitrile hydratase alpha subunit (EC 4.2.1.84), Nitrile hydratase subunit beta (NHase) (EC 4.2.1.84)	indole-3-nitrile	indole-3-acetamide
B6CWJ5, B6CWJ4, B6CWJ3	*Streptomyces rimosus*	ToyL, ToyK, ToyJ	3-cyanopyridine	pyridine-3-carbamide
			toyocamycin	toyocamycin acid amide
A5W402	*Pseudomonas putida* (strain ATCC 700007 / DSM 6899 / BCRC 17059 / F1)	Nitrile hydratase, alpha subunit (EC 4.2.1.84)	3-cyanopyridine	pyridine-3-carbamide
			3-methylbenzonitrile	3-methylbenzamide
			4-chlorobutyronitrile	4-chlorobutyramide
			4-cyanopyridine	pyridine-4-carbamide
			acrylonitrile	acrylamide
			benzonitrile	benzamide
			isobutyronitrile	isobutyramide

A phylogenetic analysis of the alpha subunits showed, as expected, that the South African agricultural soil isolates clustered together (Figure 2.15). Their nearest neighbour based on the alignment was *Pseudinocardia thermophila*. Substrate data for *P. thermophila* on BRENDA indicated that it was able to hydrolyse many simpler aliphatic and aromatic substrates, as seen for the *R. rhodochrous* isolates in this study.

A phylogeny of the beta subunits also showed a close relationship between the South African agricultural soil isolates and *Pseudonocardia thermophila*, but *R. opacus* M213 was clustered closer to the isolates (Figure 2.16). It could be assumed that these isolates have similar activity, but such sequence activity is not available. Substrate data was then linked to this alignment, by selecting a subset of aromatic or arylaromatic substrates (those most similar to beta-aminonitriles). Where available, kinetic data was included (K_m values). However, this data was very sparse, thus it was difficult to identify possible patterns of activity (Table 2.3).

Due to the small amount of overlap between the data sets, any analysis of substrate usage based on sequence will be highly incomplete, if not flawed. A suggested method to analyse sequence data with associated substrate data would be to use functional module mapping which aligns metabolic pathways to identify similarities (see Huang *et al.* (2018) for a detailed description). Unfortunately this method was beyond the scope of this study and so manual analysis was undertaken.

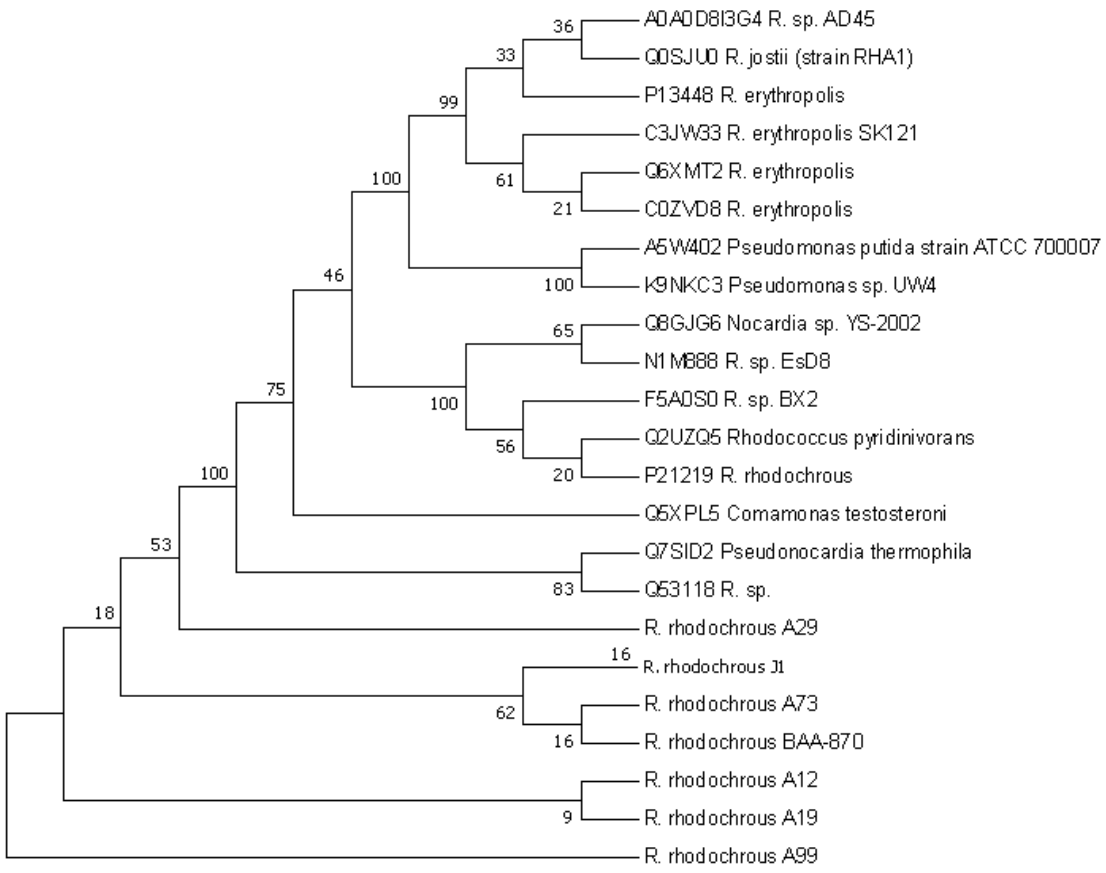


Figure 2.15: Alignment of alpha subunits of all NHase sequences for which activity data was available. The evolutionary history was inferred using the Neighbour-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 3.27035227 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). 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 JTT matrix-based method (Jones *et al.,* 1992) and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 1). The analysis involved 24 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 194 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.,* 2016).

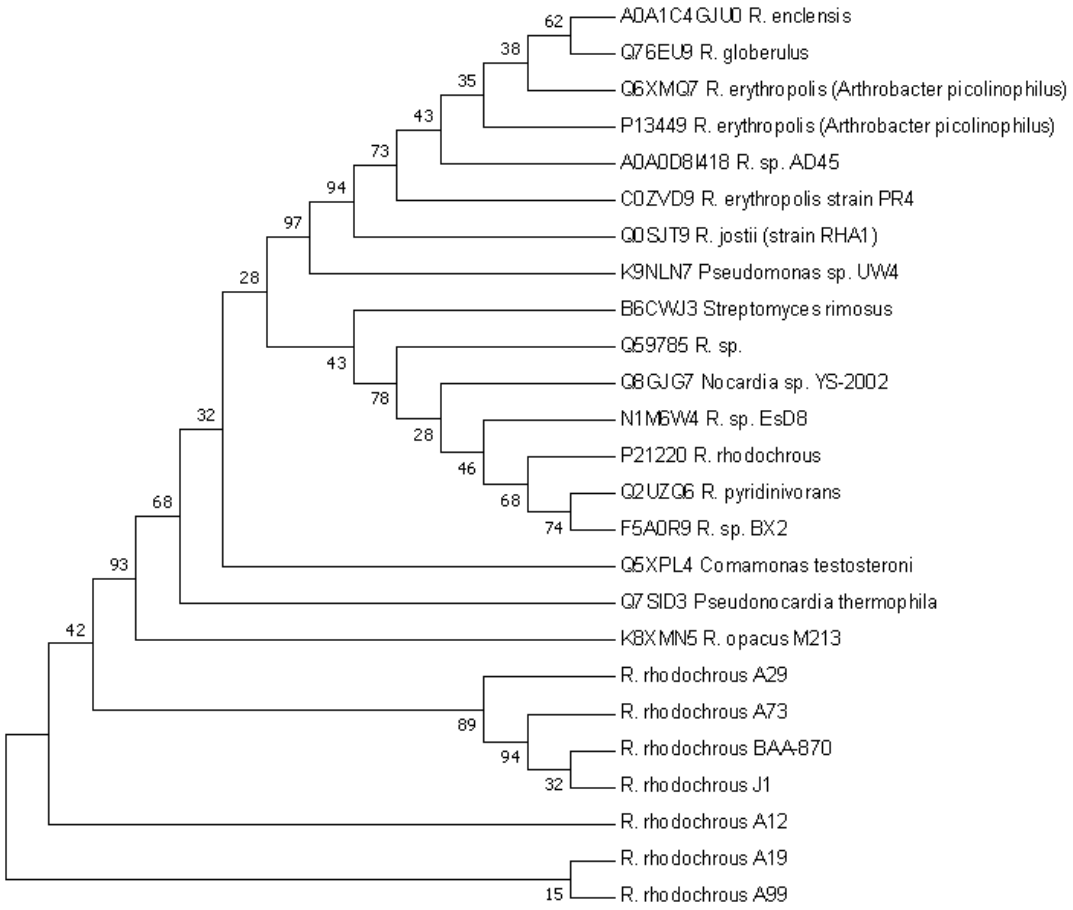


Figure 2.16: Alignment of beta subunits of all NHase sequences for which activity data was available. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 12.74738201 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). 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 JTT matrix-based method (Jones *et al.,* 1992) and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 1). The analysis involved 25 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 201 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.,* 2016).

Table 2.3: Sequence profile based on phylogenetic analysis of nitrile hydratase beta subunits from BRENDA

Isolate	Substrate*											
	Benzonitrile	3-Cyano-pyridine	3-amino-3-phenylpropionitrile	3-methyl-benzonitrile	2-furonitrile	2-phenyl-propionitrile	4-nitro-benzaldehyde	3-hydroxy-3-phenylpropionitrile	(R)-2-chloromandelonitril	3,4-dimethoxybenzonitr	3-methylbenzonitrile	4-aminobenzonitrile
R. enclensis												
R. globerulus												
R. erythropolis (Arthrobacter_picolinophilus)												
P13449 *R. erythropolis (Arthrobacter_picolinophilus)*												
A0A0D8I418 *R. sp.* AD45												
C0ZVD9 *R. erythropolis* strain PR4												
Q0SJT9 *R. jostii* (strain RHA1)												
K9NLN7 *Pseudomonas sp.* UW4												
B6CWJ3 *Streptomyces rimosus*												
Q59785 *R. sp.*	33.3											
Q8GJG7 *Nocardia sp.* YS-2002												
N1M6W4 *R. sp.* EsD8												
P21220 *R. rhodochrous* (H-NHase)		200 **										
Q2UZQ6 *R. pyridinivorans*												
F5A0R9 *R. sp.* BX2												
Q5XPL4 *Comamonas testosteroni*												
Q7SID3 *Pseudonocardia thermophila*	0.02	0.12										
K8XMN5 *R. opacus* M213												
R. rhodochrous A29												
R. rhodochrous A73												
R. rhodochrous BAA-870												
R. rhodochrous J1	0.05 **											
R. rhodochrous A12												
R. rhodochrous A19												
R. rhodochrous A99												

*Includes K_m at 25°C if available (mM). **K_m at 20°C (mM).
Clear cells indicate data unavailable in the BRENDA database.

2.3.7 Selection of amino acid residues for mutation

Although many aspects of the alpha and beta subunits were analysed, it was very difficult to determine without more complex computational methods which residues should be targeted for mutation. Alignments allowed similarities and differences between the NHases of different species to be considered, but without additional data, it was not possible to conclusively determine which residues were functionally important. In analysing the substrate scope of the enzymes, it was hoped that patterns of substrate usage may have been linked to amino acid variants, but the information for individual enzyme substrates was too incomplete to come to a conclusion from this data. As more studies are done on the different NHases, patterns of substrate usage may become more apparent. The most useful data was obtained by identifying the residues in the substrate channel or within 10Å. Based on this data, sites were chosen for directed evolution. The first site included two residues, βPhe41, as identified by Huang *et al.* (1997) and a close neighbour, βAla37, chosen at random but still located within the channel. These will be mutated using a single primer, and thus the two residues are collectively referred to as a 'multisite' (Acevedo-Rocha *et al.*, 2014).

Hashimoto *et al.* (2008) studied the binding of *tert*-Butylisonitrile (tBuNC) to the Fe-type NHase from *R. erythropolis* N771 and although their research mainly discussed the interaction of the substrate with the catalytic residues, they did note a conformation change in βMet40 when the substrate bound. This suggests that altering the residue at this position may alter the substrate binding properties or shape of the substrate channel and it is a potential target for future mutagenic studies. However, because the work of Hashimoto *et al.* focused on the Fe-type and its interaction with a compact yet branched aliphatic substrate, it is not certain how this residue may affect the Co-type NHase and its binding to aromatic beta-amino nitriles. During the course of the present study, similar research was published by Cheng *et al.* (2018) in which they also attempted to modify and succeeded in improving the enantioselectivity of the L-NHase from *R. rhodochrous* J1 by semi-rational engineering of the beta subunit. Cheng *et al.* carried out more complex docking studies using steered molecular dynamics to predict which residues in the substrate channel interacted with the substrate *rac*-mandelonitrile. They identified residues β37, β41, β48 and β51 as the residues involved in stereoselectivity of the L-NHase for *rac*-mandelonitrile. Site saturation mutagenesis via whole-plasmid PCR was used to introduce mutations at the selected residues and following two rounds of mutagenic PCR, a number of enzyme variants were identified which showed improved enantioselectivity towards (*S*)-mandelonitrile. The best variant, F37H displayed an excess of

product (ee_p) of greater than 96%, compared to the wild type ee_p of 52.6%. Interestingly, a second round of mutagenic PCR using sites β41, β48 and β51 on this variant did not improve the ee_p, although the other sites when mutated alone did produce weaker improvements to enantioselectivity.

2.4 CONCLUSIONS

Phylogenetic and sequence analysis of various groups of Co-type low molecular mass nitrile hydratases yielded data on the sequence similarities and differences on various levels. The alpha subunits consistently showed more similarity to other alpha subunits than the beta subunits did to other beta subunits. Due to the fact that the alpha subunit contains residues vital for catalysis, it was hypothesised that the beta subunit is the source of substrate selectivity between NHases, as it is more likely to undergo mutations that do not compromise activity.

Due to the complexity of the data, it was not possible to predict with certainty which residues were vital to substrate selectivity or enantioselectivity, although the analysis did provide possible residues and residue groups which could be targets for directed evolution, in particular those residues which were nearby the active site residues but which show decreased levels of conservation as well as substrate channel residues, both of which are within 10Å of the active site residues. The sites chosen for directed evolution were βPhe41 and βAla37, based on the results of previous studies and the proximity of the selected residues to the substrate channel. βMet40 is suggested as a future target for mutation.

Future work could focus on metabolic phylogenetics, or on further docking studies of substrates which show enantioselectivity, in order to determine which residues in the enzyme interact (or potentially interact) with the chemical regions of the substrate known to influence enantioselectivity. In this manner the protein active site could be theoretically moulded using molecular modelling to fit the shape of the substrate (either R- or S-enantiomer) and then direct mutations could be introduced to those residues. The interspecies NHase sequence comparisons indicated a high level of similarity between the high molecular mass NHase of *R. rhodochrous* J1 and some other species of *Rhodococci* including *R. pyridinivorans.* Further research can be done to express and characterise these other H-NHases, and attempts may be made to trace the lineage of H-NHase gene in *Rhodococci*, as the H-NHase gene in *R. rhodochrous* J1 appears to be unique amongst *R. rhodochrous* strains. The 16s rRNA sequence similarity should also be compared to the NHase sequence similarity. If the 16s rRNA sequences are very divergent,

but the NHase sequences are similar, it may support the suggestion that the H-NHase in R. rhodochrous J1 was acquired by horizontal gene transfer.

2.5 REFERENCES

Acevedo-Rocha, C.G., Hoebenreich, S. and Reetz, M.T., 2014. Iterative Saturation Mutagenesis: A Powerful Approach to Engineer Proteins by Systematically Simulating Darwinian Evolution. In: E.M.J. Gillam, J.N. Copp and D. Ackerley, eds. *Directed Evolution Library Creation*. [online] New York, NY: Springer New York.pp.103–128. Available at: <http://link.springer.com/10.1007/978-1-4939-1053-3_7> [Accessed 11 Apr. 2016].

Akhtar, N., Ghauri, M.A., Anwar, M.A. and Heaphy, S., 2015. Phylogenetic characterization and novelty of organic sulphur metabolizing genes of Rhodococcus spp. (Eu-32). *Biotechnology Letters; Dordrecht*, 37(4), pp.837–847.

Brandao, P.F.B., Clapp, J.P. and Bull, A.T., 2003. Diversity of Nitrile Hydratase and Amidase Enzyme Genes in Rhodococcus erythropolis Recovered from Geographically Distinct Habitats. *Applied and Environmental Microbiology*, 69(10), pp.5754–5766.

Brodkin, H.R., DeLateur, N.A., Somarowthu, S., Mills, C.L., Novak, W.R., Beuning, P.J., Ringe, D. and Ondrechen, M.J., 2015. Prediction of distal residue participation in enzyme catalysis. *Protein Science*, 24(5), pp.762–778.

Brodkin, H.R., Novak, W.R.P., Milne, A.C., D'Aquino, J.A., Karabacak, N.M., Goldberg, I.G., Agar, J.N., Payne, M.S., Petsko, G.A., Ondrechen, M.J. and Ringe, D., 2011. Evidence of the Participation of Remote Residues in the Catalytic Activity of Co-Type Nitrile Hydratase from *Pseudomonas putida*. *Biochemistry*, 50(22), pp.4923–4935.

Burton, S.G., Cowan, D.A. and Woodley, J.M., 2002. The search for the ideal biocatalyst. *Nature biotechnology*, 20(1), pp.37–45.

Caner, H., Groner, E., Levy, L. and Agranat, I., 2004. Trends in the development of chiral drugs. *Drug discovery today*, 9(3), pp.105–110.

Carrea, G. and Riva, S., 2008. Medium Engineering. In: V. Gotor, I. Alfonso and E. Garcia-Urdiales, eds. *Asymmetric organic synthesis with enzymes*. Weinheim: Wiley-VCH.

Cheng, Z., Peplowski, L., Cui, W., Xia, Y., Liu, Z., Zhang, J., Kobayashi, M. and Zhou, Z., 2018. Identification of key residues modulating the stereoselectivity of nitrile hydratase toward rac-mandelonitrile by semi-rational engineering. *Biotechnology and Bioengineering*, 115(3), pp.524–535.

Chhiba, V., Bode, M.L., Mathiba, K., Kwezi, W. and Brady, D., 2012. Enantioselective biocatalytic hydrolysis of β-aminonitriles to β-amino-amides using Rhodococcus rhodochrous ATCC BAA-870. *Journal of Molecular Catalysis B: Enzymatic*, 76, pp.68–74.

Chojnacki, S., Cowley, A., Lee, J., Foix, A. and Lopez, R., 2017. Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic acids research*, 45(W1), pp.W550–W553.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F. and Notredame, C., 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA

sequences using structural information and homology extension. *Nucleic Acids Research*, 39(Web Server issue), pp.W13–W17.

Frederick, J., 2013. *Genetic characterization of Rhodococcus rhodochrous ATCC BAA-870 with emphasis on nitrile hydrolysing enzymes*. [online] University of Cape Town. Available at: <http://open.uct.ac.za/handle/11427/4262> [Accessed 11 Feb. 2016].

Gasteiger, E., Gattiker, R., Duvaud, S., Wilkins, M.R. and Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: J.M. Walker, ed. *The Proteomics Protocols Handbook*. Humana Press.pp.571–607.

Guex, N. and Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, 18, pp.2714–2723.

Huang, W., Jia, J., Cummings, J., Nelson, M., Schneider, G. and Lindqvist, Y., 1997. Crystal structure of

Huang, Y., Zhong, C., Lin, H.X., Wang, J. and Peng, Y., 2018. Reconstructing Phylogeny by Aligning Multiple Metabolic Pathways Using Functional Module Mapping. *Molecules*, 23(2), p.486.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), pp.772–780.

Kobayashi, M., Nishiyama, M., Nagasawa, T., Horinouchi, S., Beppu, T. and Yamada, H., 1991. Cloning, nucleotide sequence and expression in Escherichia coli of two cobalt-containing nitrile hydratase genes from Rhodococcus rhodochrous J1. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1129(1), pp.23–33.

Komeda, H., Kobayashi, M. and Shimizu, S., 1996. A novel gene cluster including the Rhodococcus rhodochrous J1 nhlBA genes encoding a low molecular mass nitrile hydratase (L-NHase) induced by its reaction product. *Journal of Biological Chemistry*, 271(26), pp.15796–15802.

Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), pp.1870–1874.

Martinez, S., Wu, R., Sanishvili, R., Dali, L. and Holz, R., 2014. The active site sulfenic acid ligand in nitrile hydratases can function as a nucleophile. *Journal of the American Chemical Society*, 136(4), pp.1186–1189.

Mitra, S. and Holz, R.C., 2006. Unraveling the Catalytic Mechanism of Nitrile Hydratases. *Journal of Biological Chemistry*, 282(10), pp.7397–7404.

Morley, K.L. and Kazlauskas, R.J., 2005. Improving enzyme properties: when are closer mutations better? *Trends in Biotechnology*, 23(5), pp.231–237.

Muhire, B.M., Varsani, A. and Martin, D.P., 2014. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE*, 9(9), p.e108277.

Nagasawa, T., Takeuchi, K. and Yamada, H., 1991. Characterization of a new cobalt-containing nitrile hydratase purified from urea-induced cells of Rhodococcus rhodochrous J1. *European journal of biochemistry*, 196(3), pp.581–589.

Nair, N., U., Denard, C., A. and Zhao, H., 2010. Engineering of enzymes for selective catalysis. *Current Organic Chemistry*, 14(17), pp.1870–1882.

O'Mahony, R., Doran, J., Coffey, L., Cahill, O.J., Black, G.W. and O'reilly, C., 2005. Characterisation of the nitrile hydratase gene clusters of Rhodococcus erythropolis strains AJ270 and AJ300 and Microbacterium sp. AJ115 indicates horizontal gene transfer and reveals an insertion of IS1166. *Antonie Van Leeuwenhoek*, 87(3), pp.221–32.

Pathak, A., Green, S.J., Ogram, A. and Chauhan, A., 2013. Draft Genome Sequence of Rhodococcus opacus Strain M213 Shows a Diverse Catabolic Potential. *Genome Announcements*, [online] 1(1). Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3569331/> [Accessed 29 Mar. 2018].

van Pelt, S., Zhang, M., Otten, L.G., Holt, J., Sorokin, D.Y., van Rantwijk, F., Black, G.W., Perry, J.J. and Sheldon, R.A., 2011. Probing the enantioselectivity of a diverse group of purified cobalt-centred nitrile hydratases. *Organic & Biomolecular Chemistry*, 9(8), p.3011.

Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J. and Schomburg, D., 2017. BRENDA in 2017: new perspectives and new tools in BRENDA. *Nucleic Acids Research*, 45(D1), pp.D380–D388.

Rapheea, O.K.L., 2013. *The enzymatic conversion of nitriles by bacterial strains isolated from soil*. Tshwane University of Technology.

Rapheeha, O.K.L., Roux-van der Merwe, M.P., Badenhorst, J., Chhiba, V., Bode, M.L., Mathiba, K. and Brady, D., 2017. Hydrolysis of nitriles by soil bacteria: variation with soil origin. *Journal of Applied Microbiology*, 122(3), pp.686–697.

Sastry, G.M., Adzhigirey, M., Day, T., Annabhimoju, R. and Sherman, W., 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, 27(3), pp.221–234.

Savory, E.A., Fuller, S.L., Weisberg, A.J., Thomas, W.J., Gordon, M.I., Stevens, D.M., Creason, A.L., Belcher, M.S., Serdani, M., Wiseman, M.S., Grünwald, N.J., Putnam, M.L. and Chang, J.H., 2017. Evolutionary transitions between beneficial and phytopathogenic Rhodococcus challenge disease management. *eLife*, [online] 6. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5726852/> [Accessed 18 Jun. 2018].

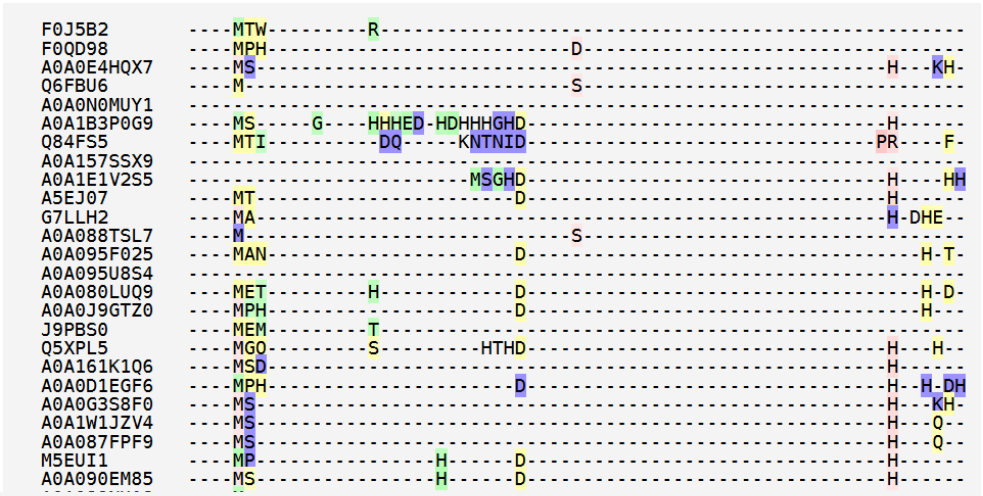
Simossis, V.A. and Heringa, J., 2003. The PRALINE online server: optimising progressive multiple alignment on the web. *Computational Biology and Chemistry*, 27(4–5), pp.511–519.

Wieser, M., Takeuchi, K., Wada, Y., Yamada, H. and Nagasawa, T., 1998. Low-molecular-mass nitrile hydratase from Rhodococcus rhodochrous J1: purification, substrate specificity and comparison with the analogous high-molecular-mass enzyme. *FEMS Microbiology Letters*, 169(1), pp.17–22.

2.6 APPENDIX

The figures included here are those which were too long to include in the main body of this thesis, including sequence alignments and phylogenetic trees; as well as figures which are of further interest but were not examined in detail in the text. In many cases, it was more appropriate to show excerpts of more complex bioinformatic analyses. Also included are any large figures reproduced from previous works which were too extensive to include in the main text.

A1.1



A1.2

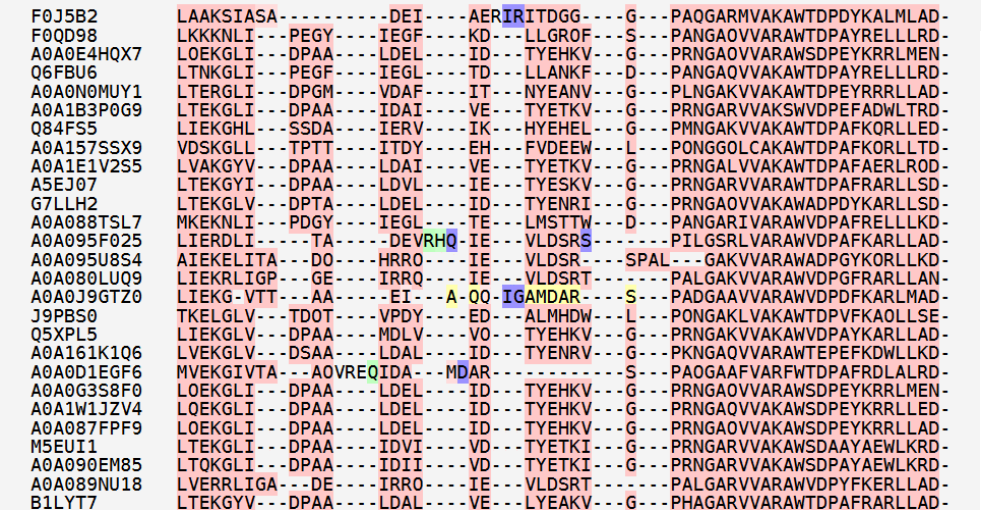
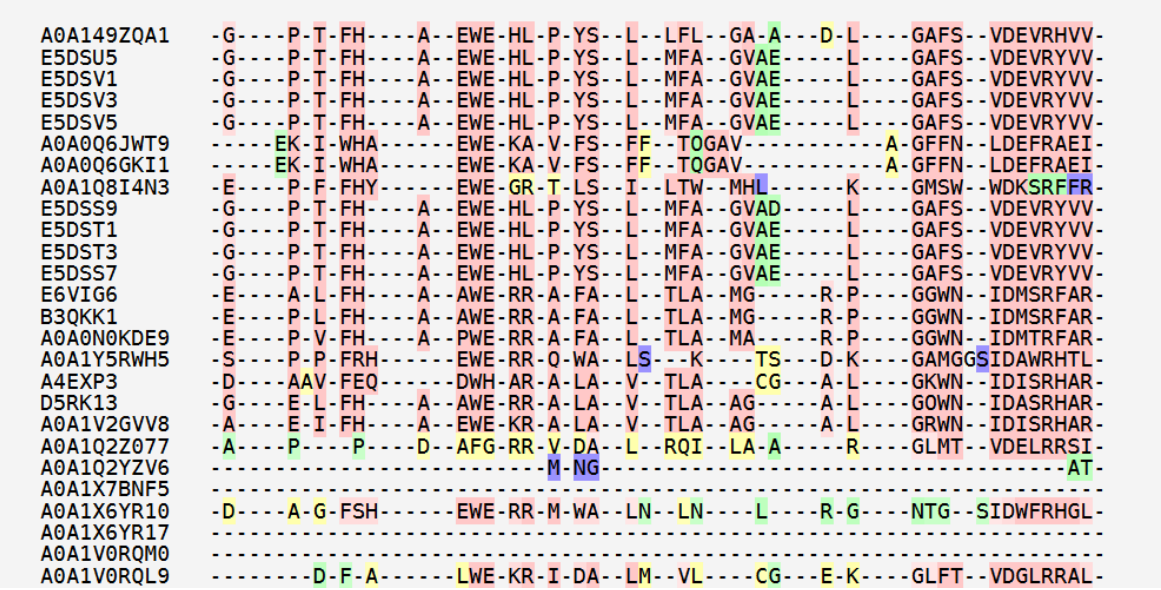


Figure A1: Excerpt from the alignment of the nitrile hydratase alpha subunits from representative organisms available on the BRENDA database. Expresso on the T-Coffee database was used (Di Tommaso *et al.*, 2011). A1.1: There are many gaps and poor similarity at the beginning of the alignment. A1.2: Conserved domains are shown by the longer regions of similarity towards the centre of the alpha subunit sequences.

A2.1



A2.2

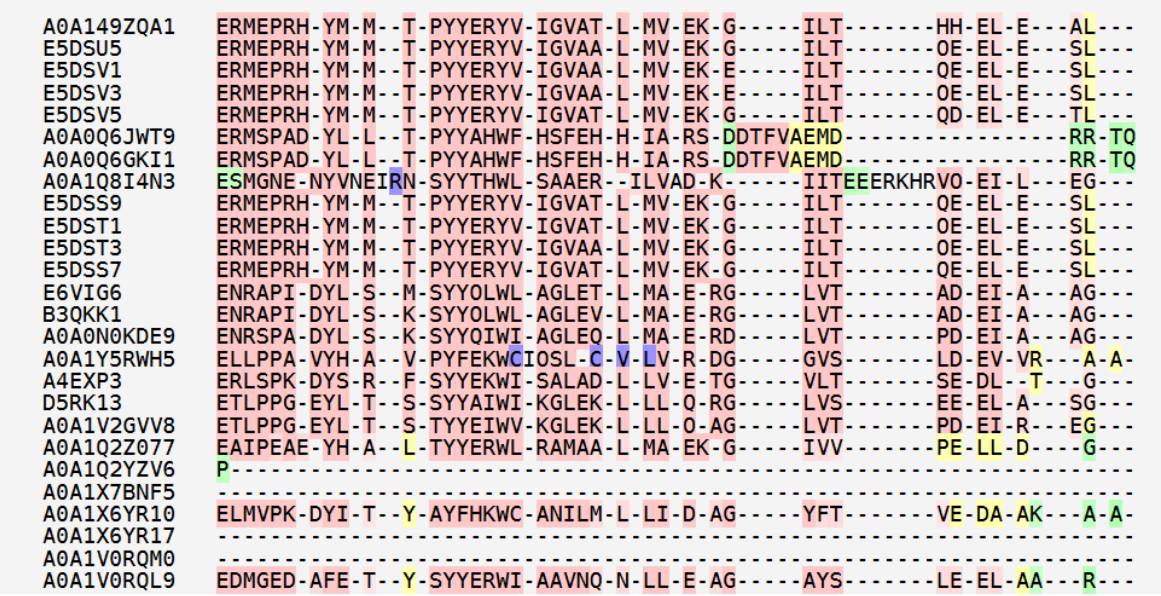


Figure A2: Excerpt from the alignment of the nitrile hydratase beta subunits from representative organisms available on the BRENDA database. Expresso on the T-Coffee database was used (Di Tommaso *et al.*, 2011). A2.1: There are many gaps and poor similarity at the beginning of the alignment. A2.2: Conserved domains are shown by the longer regions of similarity towards the centre of the alpha subunit sequences.

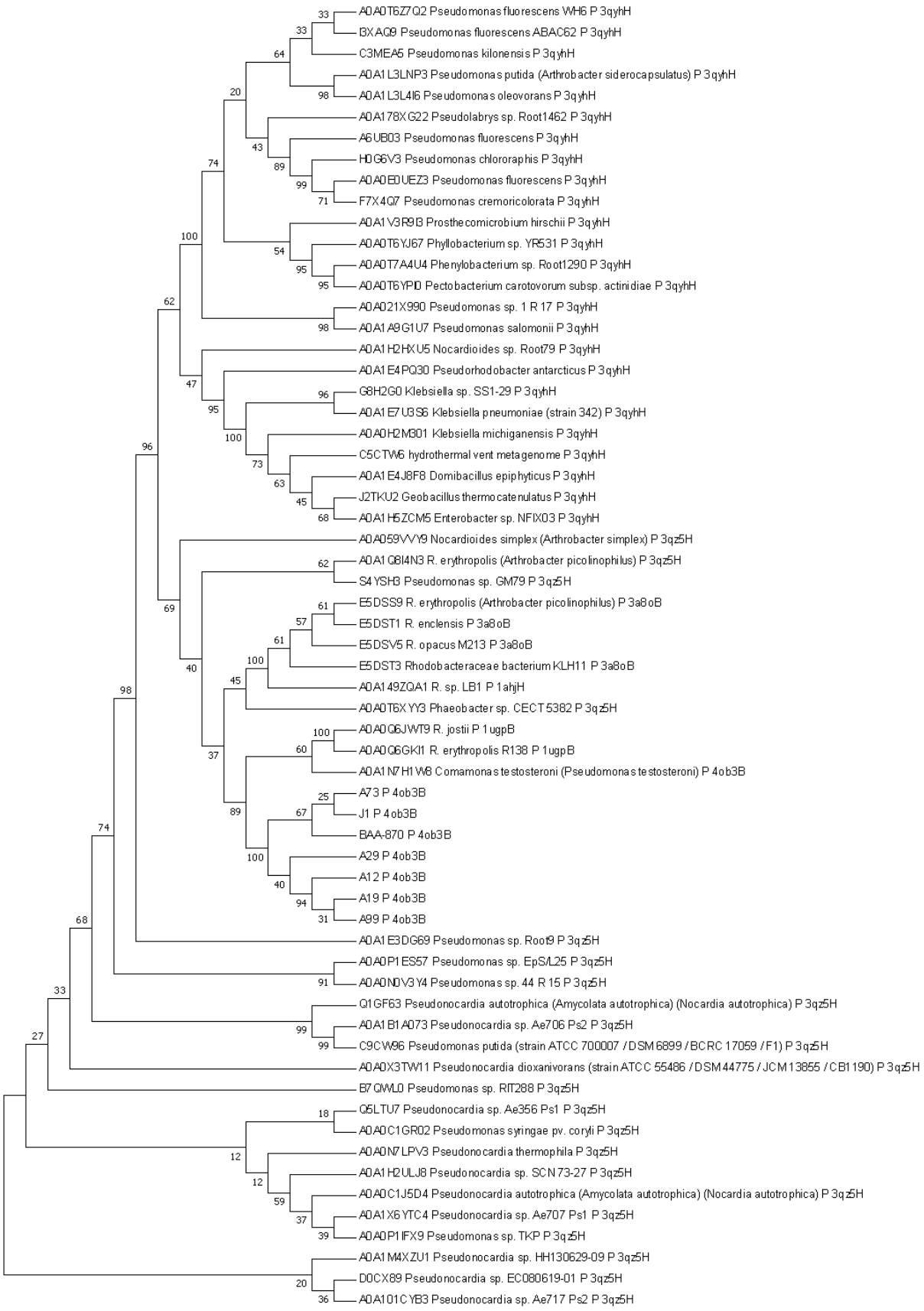

A0A0T6Z7Q2 Pseudomonas fluorescens WH6 P 3qyhH
B3XAQ9 Pseudomonas fluorescens ABAC62 P 3qyhH
C3MEA5 Pseudomonas kilonensis P 3qyhH
A0A1L3LNP3 Pseudomonas putida (Arthrobacter siderocapsulatus) P 3qyhH
A0A1L3L4I6 Pseudomonas oleovorans P 3qyhH
A0A178XG22 Pseudolabrys sp. Root1462 P 3qyhH
A6UB03 Pseudomonas fluorescens P 3qyhH
H0G6V3 Pseudomonas chlororaphis P 3qyhH
A0A0E0UEZ3 Pseudomonas fluorescens P 3qyhH
F7X4Q7 Pseudomonas cremoricolorata P 3qyhH
A0A1V3R9I3 Prosthecomicrobium hirschii P 3qyhH
A0A0T6YJ67 Phyllobacterium sp. YR531 P 3qyhH
A0A0T7A4U4 Phenylobacterium sp. Root1290 P 3qyhH
A0A0T6YPI0 Pectobacterium carotovorum subsp. actinidiae P 3qyhH
A0A021X990 Pseudomonas sp. 1 R 17 P 3qyhH
A0A1A9G1U7 Pseudomonas salomonii P 3qyhH
A0A1H2HXU5 Nocardioides sp. Root79 P 3qyhH
A0A1E4PQ30 Pseudorhodobacter antarcticus P 3qyhH
G8H2G0 Klebsiella sp. SS1-29 P 3qyhH
A0A1E7U3S6 Klebsiella pneumoniae (strain 342) P 3qyhH
A0A0H2M301 Klebsiella michiganensis P 3qyhH
C5CTW6 hydrothermal vent metagenome P 3qyhH
A0A1E4J8F8 Domibacillus epiphyticus P 3qyhH
J2TKU2 Geobacillus thermocatenulatus P 3qyhH
A0A1H5ZCM5 Enterobacter sp. NFIX03 P 3qyhH
A0A059VVY9 Nocardioides simplex (Arthrobacter simplex) P 3qz5H
A0A1Q8I4N3 R. erythropolis (Arthrobacter picolinophilus) P 3qz5H
S4YSH3 Pseudomonas sp. GM79 P 3qz5H
E5DSS9 R. erythropolis (Arthrobacter picolinophilus) P 3a8oB
E5DST1 R. enclensis P 3a8oB
E5DSV5 R. opacus M213 P 3a8oB
E5DST3 Rhodobacteraceae bacterium KLH11 P 3a8oB
A0A149ZQA1 R. sp. LB1 P 1ahjH
A0A0T6XYY3 Phaeobacter sp. CECT 5382 P 3qz5H
A0A0Q6JWT9 R. jostii P 1ugpB
A0A0Q6GKI1 R. erythropolis R138 P 1ugpB
A0A1N7H1W8 Comamonas testosteroni (Pseudomonas testosteroni) P 4ob3B
A73 P 4ob3B
J1 P 4ob3B
BAA-870 P 4ob3B
A29 P 4ob3B
A12 P 4ob3B
A19 P 4ob3B
A99 P 4ob3B
A0A1E3DG69 Pseudomonas sp. Root9 P 3qz5H
A0A0P1ES57 Pseudomonas sp. EpSAL25 P 3qz5H
A0A0N0V3Y4 Pseudomonas sp. 44 R 15 P 3qz5H
Q1GF63 Pseudonocardia autotrophica (Amycolata autotrophica) (Nocardia autotrophica) P 3qz5H
A0A1B1A073 Pseudonocardia sp. Ae706 Ps2 P 3qz5H
C9CW96 Pseudomonas putida (strain ATCC 700007 / DSM 6899 / BCRC 17059 / F1) P 3qz5H
A0A0X3TW11 Pseudonocardia dioxanivorans (strain ATCC 55486 / DSM 44775 / JCM 13855 / CB1190) P 3qz5H
B7QWL0 Pseudomonas sp. RIT288 P 3qz5H
Q5LTU7 Pseudonocardia sp. Ae356 Ps1 P 3qz5H
A0A0C1GR02 Pseudomonas syringae pv. coryli P 3qz5H
A0A0N7LPV3 Pseudonocardia thermophila P 3qz5H
A0A1H2ULJ8 Pseudonocardia sp. SCN 73-27 P 3qz5H
A0A0C1J5D4 Pseudonocardia autotrophica (Amycolata autotrophica) (Nocardia autotrophica) P 3qz5H
A0A1X6YTC4 Pseudonocardia sp. Ae707 Ps1 P 3qz5H
A0A0P1IFX9 Pseudomonas sp. TKP P 3qz5H
A0A1M4XZU1 Pseudonocardia sp. HH130629-09 P 3qz5H
D0CX89 Pseudonocardia sp. EC080619-01 P 3qz5H
A0A101CYB3 Pseudonocardia sp. Ae717 Ps2 P 3qz5H

Figure A3: Evolutionary relationships of nitrile hydratase beta subunits. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 13.79287892 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (Felsenstein, 1985). 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 JTT matrix-based method (Jones *et al.,* 1992) and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The analysis involved 62 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 171 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.,* 2016).

Table A1: Activity of South African *R. rhodochrous* agricultural soil isolates. Adapted from Rapheea (2013).

Isolation substrate	Area of sampling	CSIR code	Identification	Propionitrile	Adiponitrile	Benzonitrile	4-chlorobenzonitrile	Cyanopyridine	Hydrocinnamonitrile	3-hydroxy-4-phenoxybutyronitrile	3-amino-3-phenylpropionitrile	Benzylidenemalononitrile
Benzylidenemalononitrile	Cabbage 2	A 12	*Rhodococcus* group II	1	1	1	1	0	1	1	1	0
Benzonitrile	Cabbage 1	A 19	*Rhodococcus* group II	0	1	1	1	0	1	1	1	0
Benzonitrile	Radish	A 29	*Rhodococcus* group II	0	1	1	1	1	1	1	0	0
Adiponitrile	Cabbage1	A 73	*Rhodococcus* group II	0	1	1	1	0	1	0	1	0
4-Chlorobenzonitrile	Cabbage 1	A 98	*Rhodococcus* group II	0	1	1	1	0	0	1	0	1
Benzylidenemalononitrile	Radish	A 85	*Agrobacterium tumefaciens*	0	1	1	0	0	0	0	0	0
4-Chlorobenzonitrile	Cabbage 1	A 99	*Pimelobacter*	0	1	0	1	1	1	0	1	0

*A85 not included in the present study as it was not identified as *R. rhodochrous.*

CHAPTER 3: GENETIC ENGINEERING AND DIRECTED EVOLUTION VIA ITERATIVE SATURATION MUTAGENESIS

3.1 INTRODUCTION

3.1.1 Recombinant protein expression and its contribution to modern biocatalysis

In order to use proteins as biocatalysts for industrial and other applications, being able to alter the structure of the protein or its expression is often necessary. Recombinant expression often allows for more convenient or faster protein expression. It also allows the enzyme to be expressed or studied separately from other host cell metabolic processes which may interfere with the enzyme reaction under study (Rzeznicka *et al.*, 2010; Rosano and Ceccarelli, 2014). Recombinant expression is useful not only for isolated study and manipulation of a potential industrial biocatalyst, but it may also allow an ideal expression system to be designed for the enzyme catalyst. Cells and vectors can be chosen to maximise expression, solubility and stability of the enzyme, or to design cells which are tolerant to high concentrations of substrate or product (Fei *et al.*, 2015; Kurosawa *et al.*, 2015). All of this contributes to cost and effort reduction in the overall industrial process.

3.1.2 Industrially useful *Rhodococci*

Due to their large genome size (Vandergeize and Dijkhuizen, 2004; Nguyen *et al.*, 2016) and ability to grow in toxic environments and to degrade a wide variety of substrates including cholesterol (Yam *et al.*, 2010), cyanide (Lin *et al.*, 2012) and phenol (Arif *et al.*, 2012), the genomes of *Rhodococci* have been explored to identify genes which could be recombinantly expressed and adapted to industrial or other uses. The studies reviewed in this chapter are of two main types: those which have developed expression systems which use the Rhodococcal cells as hosts, and those which use genetic elements from Rhodococci (in the form of enzyme-encoding genes) to create engineered vectors which confer desired characteristics to the host cell.

3.1.3 Rhodococcal hosts and expression vectors

Initial research on Rhodococcal genes and plasmids focused on the use of the plasmid as a vector, or using parts of the plasmid to confer new traits on host cells. For example the construction of a shuttle vector containing *E. coli* genes for induction and selection and *Rhodococcus* genes for arsenic resistance and phage immunity was constructed by Dabbs *et al.* (1990). This vector could be transformed and replicated in both *E. coli* and *Rhodococcus*, allowing it to act as a shuttle for genes between the two species, but the arsenic resistance gene did not confer arsenic resistance to *E. coli* cells containing the plasmid. The shuttle vector was later improved to reduce its size, increase copy number and replace the arsenic resistance selection marker with a chloramphenicol resistance selection marker (Quan and Dabbs, 1993). The shuttle plasmid host range was later increased by Mot *et al.* (1997) to include four Rhodococcus species. Nitrile hydratase from *Rhodococcus sp.* N-774 was first recombinantly expressed by Hashimoto *et al.* (1992) using an *E.coli-Rhodococcus* shuttle plasmid in an *R. rhodochrous* ATCC 12674. Although recombinant gene expression has advanced greatly, there is still a need for shuttle plasmids, as they can allow for the expression of proteins in other hosts where *E. coli* is unsuitable (Stanislauskiene *et al.*, 2012). Improvement to the original *Rhodococcus* shuttle vector by Dabbs *et al.* (1990) was achieved by identifying the genetic elements responsible for characteristics such as arsenic resistance and copy number (Quan and Dabbs, 1993). This seems to have contributed to later research, which focused on the genes present on *Rhodococcus* plasmids (Kesseler *et al.*, 1996) and eventually on recombinantly expressing *Rhodococcus* genes using other plasmids, as described in Section 3.1.3.

3.1.4 Recombinant expression of nitrile hydratase from *Rhodococcus rhodochrous*

It has previously been found that, in order to express functional NHase, the growth media for recombinant *Rhodococci* must be supplemented with either Co or Fe**,** depending on the type of NHase being expressed (Nojiri et al., 2000). Pei *et al.* (2017) were able to express the entire nitrile hydratase and amidase operon in *E. coli* cells in order to degrade benzonitrile-based herbicides to their amides and acids. Naturally occurring environmental microbes were reported to be able to degrade these herbicides to amides, but the amides contributed to groundwater contamination, hence the amidase was included so that the amides could be further degraded to non-toxic carboxylic acids. Experiments with *Rhodococcus erythropolis* A4 and *Aminobacter sp.* MSH1 whole cells harbouring native enzymes were not successful, therefore a recombinant approach was attempted. Both dual construct and single plasmid systems were

tested using Plasmids pET24a and pCDFDuet-1 (Figure 3.1). The genes from three *Rhodococcus* species were tested: the NHases from *R. rhodochrous* J1 and *R. opacus* M213 and the amidase from *R. erythropolis*. Although both proteins were efficiently expressed using the two-construct system, the authors reported that *E. coli* growth was slower in cells containing two plasmids.

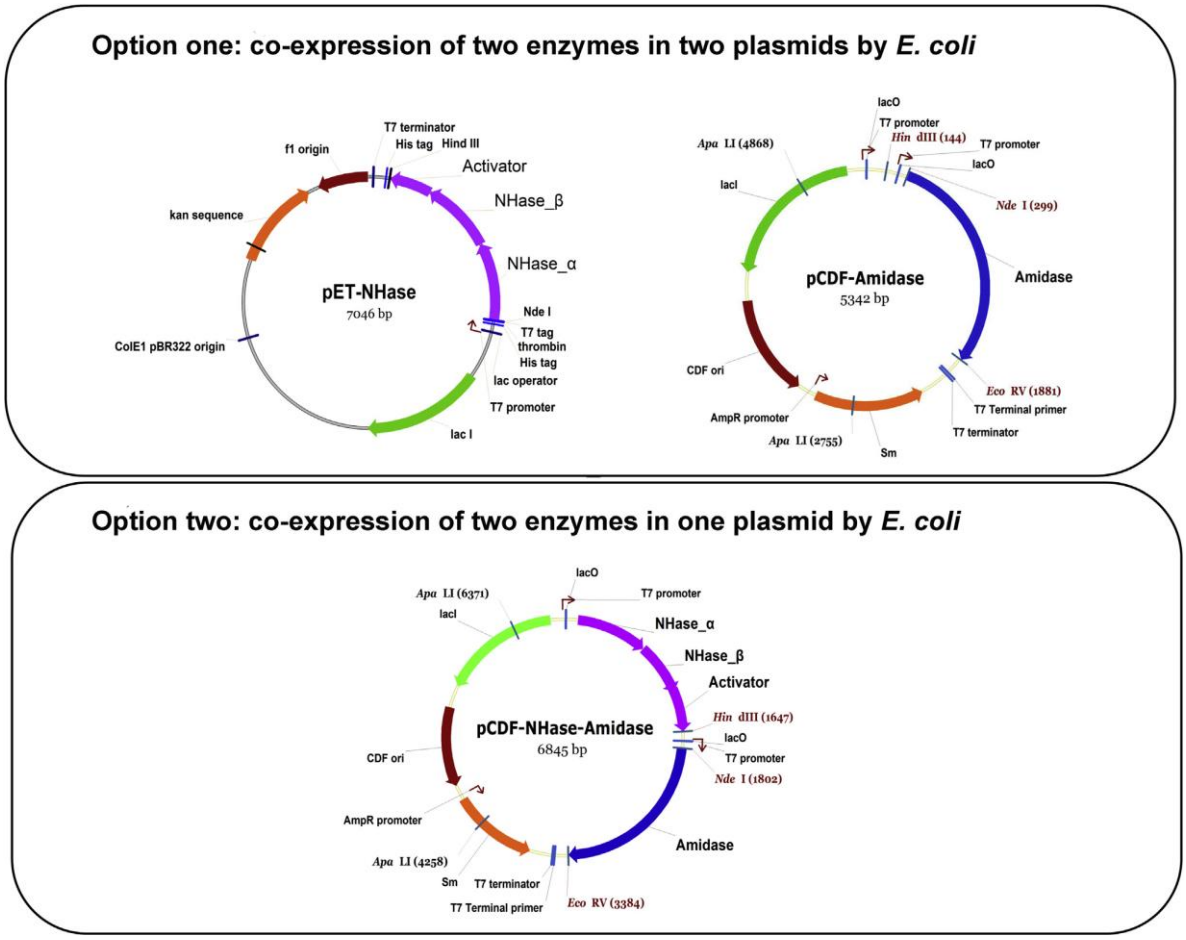


Figure 3.1: Schematic of the two plasmid systems designed to express the Rhodococcal nitrile hydratase and amidase genes for herbicide degradation. Two enzymes can be expressed on separate plasmids, or together on a single plasmid. Reproduced from Pei *et al.* (2017)**.**

Rzeznicka *et al.* (2010) expressed the Fe-type NHase from *R. equi* TG328-2 in *E. coli* BL21(DE3) using plasmids pET21a and pET28a. Interestingly, they report that rhamnose-inducible promoters were less effective than IPTG-induced T7-RNA Polymerase promotors. Furthermore, there was protein expression

when the alpha and beta subunits were expressed without the activator protein. When the activator protein sequence was cloned in the same plasmid as the subunit sequences, no activator protein was detected. It was then determined that cloning the activator into a separate plasmid and co-transforming a single host cell with two plasmids led to ideal NHase expression and activity. Soluble protein expression was improved by expressing at 20°C when the cells reached an OD_{600} of 1.0, which is unusually high for induction.

A similar two-construct expression system was designed by Frederick (2013) to express the low molecular mass cobalt-type nitrile hydratase and its chaperone protein from *R. rhodochrous* ATCC BAA-870. The alpha subunit and chaperone were cloned into pRSFDuet-1 and the beta subunit was cloned into pET21a(+). Although pRSFDuet-1 has two multiple cloning sites (MCS), the alpha-chaperone sequence was cloned as a single sequence into only one MCS of the plasmid. It is possible that the original intention was to separate the two genes and clone one into each MCS. However, the ribosome binding site for the chaperone begins before the end of the alpha subunit gene and they are in fact in two different reading frames, with the alpha subunit gene in Frame 3 and the chaperone gene in Frame 1. The open reading frames (ORFs) were identified in the present study using the ORF Finder at https://www.ncbi.nlm.nih.gov/orffinder/. Frederick (2013) designed the co-expression system to address the problematic alpha and beta subunit expression ratios when expressing them as a single sequence in one plasmid. Single-plasmid expression resulted in lower alpha subunit levels and by separating the subunit genes into separate plasmids, their expression would then be under control of the plasmid promoters and ribosome binding sites. The work by Frederick supported earlier findings by Tsekoa *et al. (*2004) that the chaperone was expressed efficiently despite being in a different reading frame to the alpha subunit. The present study utilised the co-expression system designed by Frederick (methods detailed in Section 3.2.1.1).

3.1.5 Directed evolution for the improvement of industrially useful enzymes

As well as expressing an enzyme of interest in a functional state and in a desired quantity, there may be other characteristics of the enzyme which reduce its functionality in an industrial setting (Arnold and Volkov, 1999; Hammer *et al.*, 2017). Changes then need to be made to the enzyme structure, but it is not always possible to determine which residues affect the characteristic in question; and even when these residues can be identified with some certainty, it is also difficult to predict what mutations would be most

effective (Arnold and Volkov, 1999; Sankararaman *et al.*, 2009). Directed evolution is a powerful method used to evolve enzymes without requiring detailed structural knowledge. Mutations can be introduced in a completely random manner, or a semi-rational approach can be taken where specific residues are chosen based on knowledge of the enzyme's structure and then randomised (Reetz *et al.*, 2010). There are many techniques for carrying out directed evolution, at all steps of the process: from the selection of residues and sites to the introduction of mutations and then the replication of the gene of interest, to the screening of mutants for improvements. The techniques pertinent to the present study are discussed in more detail in Sections 3.1.5 to 3.1.6.

3.1.6 The megaprimer method of PCR for whole-plasmid replication

Directed evolution involves manipulation of the gene sequence of a protein and therefore requires in many cases that the gene be extracted and replaced into the host multiple times. This increases the workload when developing a new biocatalyst or other altered protein and so a method which requires the fewest number of steps is preferable (Pai *et al.*, 2012). Whole plasmid PCR allows for the replication of a full plasmid including a recombinant gene without the need to ligate the gene into the plasmid after each mutation step (Miyazaki and Takenouchi, 2002; Pai *et al.*, 2012).

Megaprimer PCR was developed as a simple way to replicate a section of DNA and introduce mutations at desired regions in as few steps as possible (Sarkar and Sommer, 1990; Barettino *et al.,* 1994). MEGAWHOP PCR (megaprimer PCR of whole plasmid) is an adaptation of megaprimer PCR which involves an initial PCR step using a whole plasmid as the template (Miyazaki and Takenouchi, 2002; Miyazaki, 2011). As well as time and effort reduction, this method ensures that recombinant plasmids contain a single insert by avoiding multiple digestion and ligation steps. Thus, it is not necessary to screen for plasmids lacking the insert or containing multiple inserts (Pai *et al.*, 2012). It is necessary to note that the term megaprimer is used differently in different publications, however it always refers to a large primer which is several hundred nucleotides long (Sarkar and Sommer, 1990). In Miyazaki and Takenouchi (2002), the term refers to the set of primers used in a single PCR reaction to replicate a whole plasmid. The primers are complementary and bind to either strand of the plasmid at the same point, allowing both strands to be replicated (Figure 3.2). In other studies (Sarkar and Sommer, 1990; Barettino *et al.*, 1994; Sanchis *et al.*, 2008), a mutagenic primer and a silent reverse primer, both of typical primer length (approximately 20 bp) anneal to the plasmid and the product of this PCR is a large fragment of the plasmid, termed the

megaprimer. A second PCR step is then done where the megaprimer is used to reproduce the entire plasmid as a product (Figure 3.3).

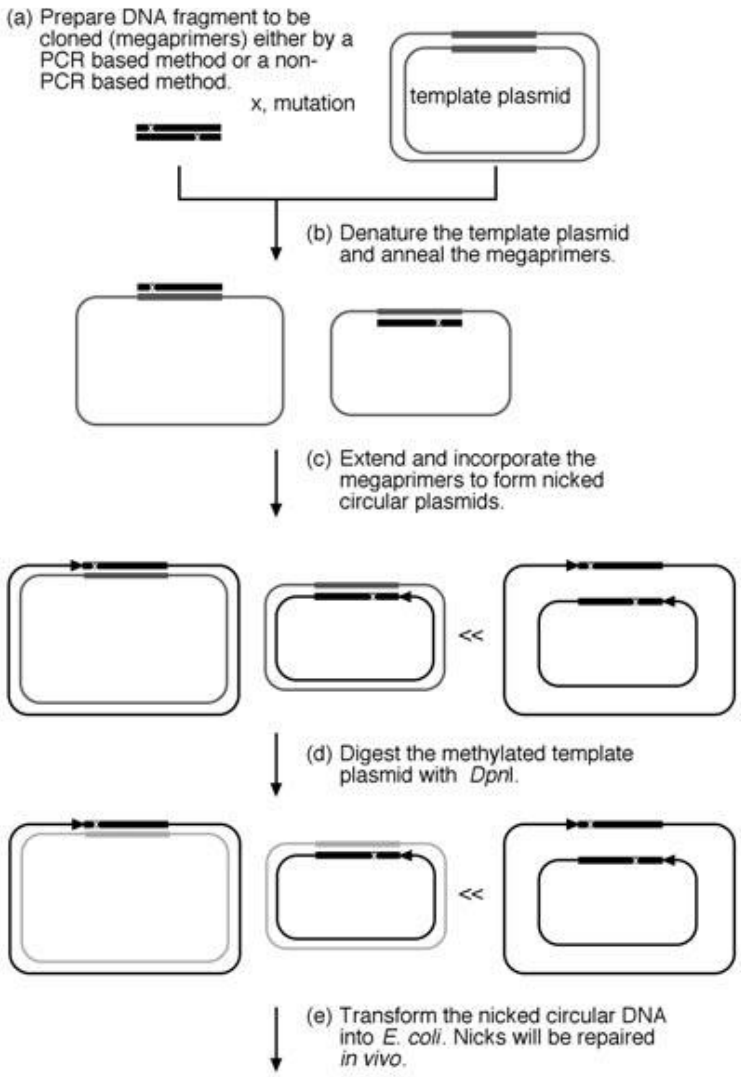


Figure 3.2: The MEGAWHOP (megaprimer PCR of whole plasmid) method. The megaprimers are complementary and are used in a single PCR step to replicate the whole plasmid. The template plasmid is methylated while the PCR product is not, and so the template can then be digested with *DpnI,* which recognises only methylated DNA. Reproduced from Miyazaki and Takenouchi (2002).

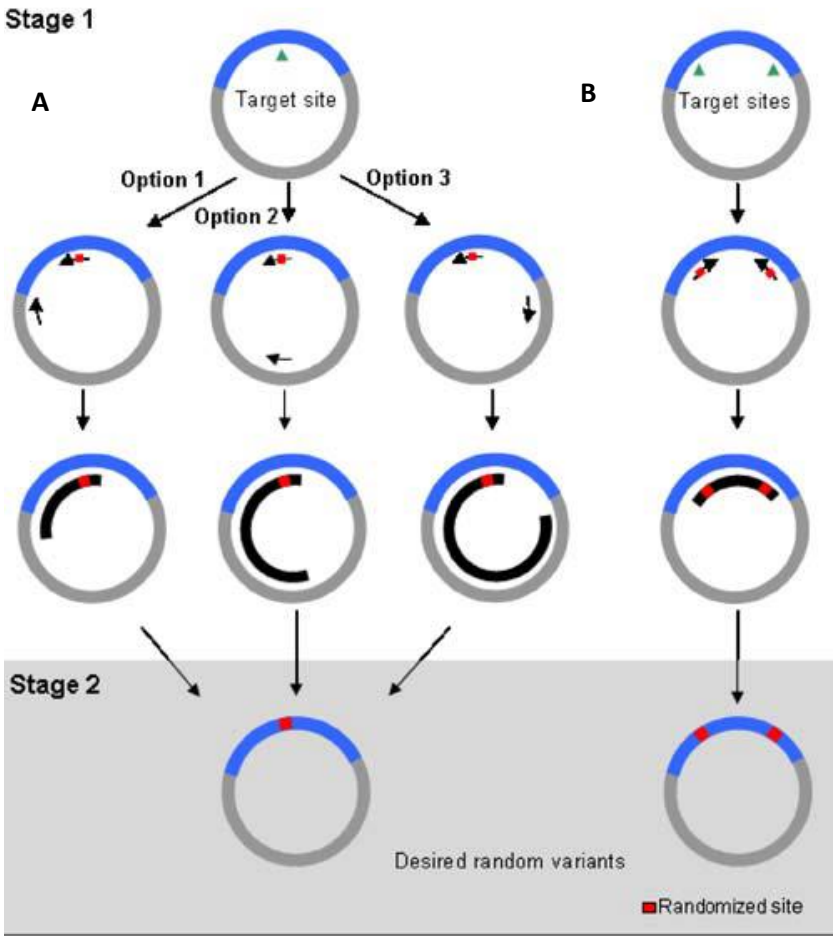


Figure 3.3 Megaprimer PCR where the megaprimer is produced as the product in the first round of PCR. A: The primers in the first stage are of typical size (18-22 bp). One primer is mutagenic, while the second primer is a silent primer, usually a standard primer for the plasmid in question, which binds outside the insert region. B: The primers used in the first PCR are both mutagenic and bind within the insert region. Black: The megaprimer produced as a result of the first PCR, which is approximately a quarter or greater of the total plasmid size. Red spots indicate introduced mutations. A second PCR step is done using template plasmid and the new megaprimer, resulting in the final product shown in Stage 2, a whole plasmid with new mutations. Adapted from Sanchis *et al.* (2008).

3.1.7 Iterative saturation mutagenesis and the Combinatorial Active-site Saturation Test (CAST)

Site directed mutagenesis is the alteration of a specific site in a gene or protein (Ho *et al.*, 1989). Some reasons for its use include the alteration of a nucleotide within a codon which then changes the amino acid encoded by that codon (Ho *et al*., 1989); to alter a ribosome binding site (Frederick, 2013); and to add or remove restriction sites. Site saturation mutagenesis (SSM) is the replacement of a single amino acid with all other possible amino acids, each time observing how each amino acid mutation affects protein structure and function (DeSantis *et al.*, 2003; Li *et al*., 2018). Site saturation mutagenesis is a valuable tool in directed evolution, but in order to correctly mimic natural evolutionary progress, the process must be carried out multiple times with mutations being introduced at multiple sites (Reetz *et al.*, 2006). This is referred to as iterative saturation mutagenesis (ISM). This technique involves carrying out a mutagenesis step, selecting mutants that show an improvement in the desired protein characteristic (screening for 'hits') and then submitting those selected mutants to another round of mutagenesis at a different site (Reetz *et al.*, 2010), an example of semi-rational design.

A variation on ISM is called the Combinatorial Active-site Saturation Test (CAST), in which residues surrounding the active site are targeted specifically for mutation. The residues are grouped into sets of two or three residues which are situated near each other in the gene sequence and can thus be mutated at the same time using a single primer per group of residues (Reetz *et al.*, 2006). This allows smaller libraries to be created, reducing screening time; while also allowing for potential cooperative effects between mutations (Reetz *et al.*, 2006). This was originally designed to alter substrate scope, but has also proven to be effective for improving enantioselectivity (Reetz *et al.*, 2006; Reetz and Zheng, 2011). There have been many studies where enantioselectivity of an enzyme has been changed or improved. Reetz and Zheng (2011) improved the enantioselectivity of *Aspergillus niger* epoxide hydrolase for the (*S*)-glycidyl phenyl ether from E=4.6 to E=160, with improved expression of the enzyme using ISM. The enantioselectivity of *Candida antarctica* lipase A towards ibuprofen ester was improved from an E of 3.4 to an E of 100, with a concurrent improvement in activity using CASTing (Sandström *et al.*, 2012). DeSantis *et al.* (2003) improved the enantioselectivity of a nitrilase used to produce (*R*)-4-cyano-3-hydroxybutyric acid, a precursor to the cholesterol drug Lipitor. This was done using gene site saturation mutagenesis (GSSM), a form of saturation mutagenesis which involves replacing every amino acid residue in a protein with every other possible amino acid.

Screening of mutant libraries for hits is still a step which requires a lot of time and effort, slowing down the speed at which new biocatalysts can be produced (Otten *et al.*, 2010). In the case of enantioselectivity, methods of selecting the best hit are still difficult to design (Jaeger *et al.*, 2001), or are too slow compared to the high-throughput needs of industrial biocatalyst development (Sicard and Reymond, 2006). Many factors contribute to the increased speed and throughput of analysis methods used to identify 'hits' in an evolved biocatalyst library. As technology has improved, robots and instruments able to analyse multiple samples in parallel are being developed which both reduces overall sampling time and the manual labour involved (Wahler and Reymond, 2001). In many cases, screening tests have developed which are specific to the enzyme and reaction under study. A comprehensive account of some of the enzyme assays designed for industry, including enantioselective assays, can be found in Reymond (2006).

As well as improving the speed and sensitivity of assay methods, it is also possible to reduce the screening effort by producing smaller, more focused libraries (Lutz, 2010). This is accomplished by using in silico modelling techniques to identify possible hits instead of mutating all possible residues randomly (Kille *et al.*, 2011; Sandström *et al.*, 2012; Herwig *et al.*, 2017), as well as by using reducing amino acid alphabets (Reetz *et al.*, 2008; Kille *et al.*, 2013). These reduced alphabets replace a residue with only certain other residues using degenerate codons in the primer. In this way, residues that are representative of different classes of amino acids (e.g. basic, aliphatic etc.) can be introduced without having to test every amino acid in that position. This technique also the reduces the presence of redundant codons and stop codons (Tang *et al.*, 2012). NDT codon degeneracy is commonly used as it reduces the number of transformants which much be screened to ensure 95% library coverage, reducing the workload and time considerably. The basis of this technique is described in further detail by Reetz *et al.* (2008).

3.2 MATERIALS AND METHODS

3.2.1 Chemicals

Chemicals were obtained from Sigma-Aldrich (Now Merck; Modderfontein, South Africa) unless otherwise stated. Reagents for microbiology and protein biochemistry were obtained from BioRad (California, USA). Reagents from Thermo Fisher Scientific (Massachusetts, USA) were obtained from Lasec (Cape Town, South Africa), unless otherwise stated.

3.2.2 Expression systems

The NHase gene used in this study was recombinantly expressed due to the advantages of the recombinant systems over the native *R. rhodochrous* host. These include the presence of double antibiotic resistance markers in the plasmids containing the gene subunits, which ensures that contamination is minimized and also allows the presence of the plasmids to be determined based on colony growth. *E. coli* has a faster growth time compared to *R. rhodochrous,* with three hours to reach mid log phase compared to approximately eight hours for *R. rhodochrous* (Frederick, 2007) . Furthermore, plasmid extraction from *R. Rhodococcus* was slightly more time consuming as the cell walls are more difficult to lyse than those of *E. coli*. Both of the systems used were two-construct systems, separating the subunits into separate plasmid vectors, but expressing both vectors concurrently in a single host cell.

3.2.2.1 Recombinant expression of NHase using pRSFDuet-1 and pET21a(+)

Genscript plasmid design and gene optimisation

Based on the protocol developed by Frederick (2013), two plasmids were ordered from GenScript (USA). This was done to minimise initial cloning attempts, as well as to optimise the codon sequence for expression in *E. coli* BL21(DE3). The alpha subunit-chaperone gene sequence was inserted as a single construct into the second multiple cloning site (MCS) in the pRSFDuet-1 plasmid between the NdeI and PacI restriction sites. The beta subunit gene sequence was inserted into the MCS of pET-21a(+) between

the NdeI and XhoI restriction sites, and incorporating the C-terminal His-tag from the vector by removing the gene's native stop codon. The map of each plasmid is shown in Figures 3.4 and 3.5.

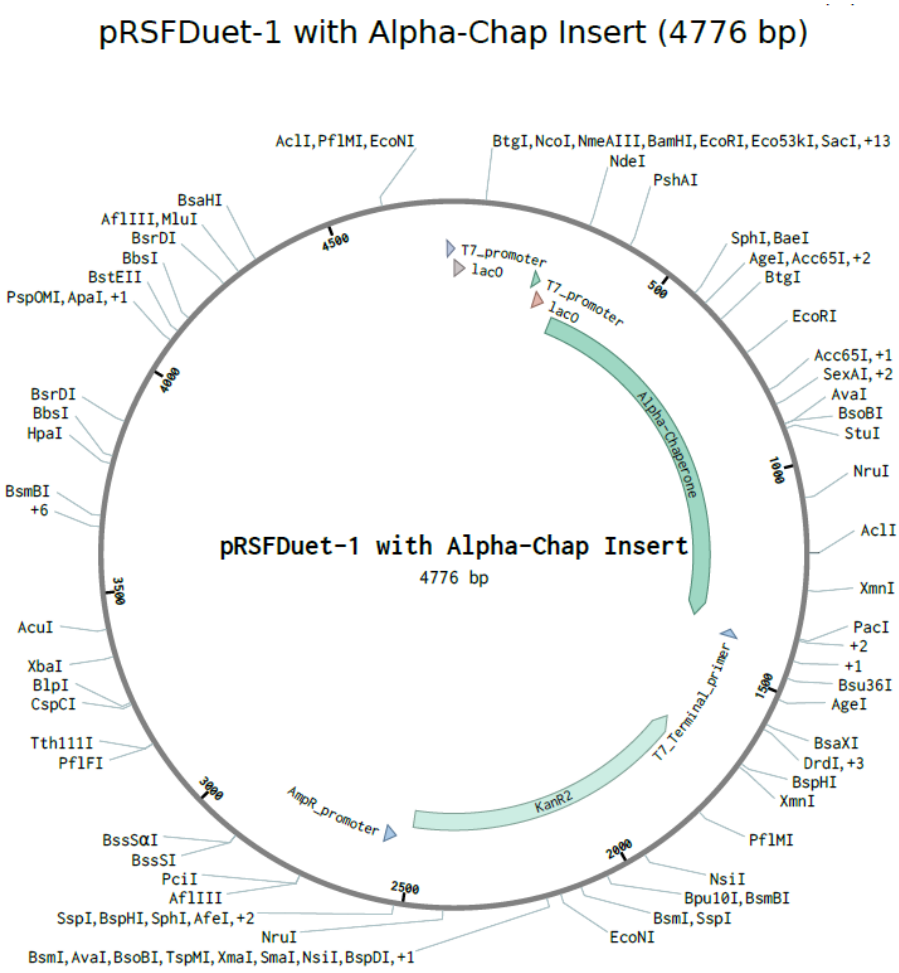


Figure 3.4: Plasmid map of pRSFDuet-1 with the alpha subunit and chaperone gene sequences inserted into MCS2 between NdeI and PacI. Image generated using the Benchling workspace at https://benchling.com (Benchling, 2017).

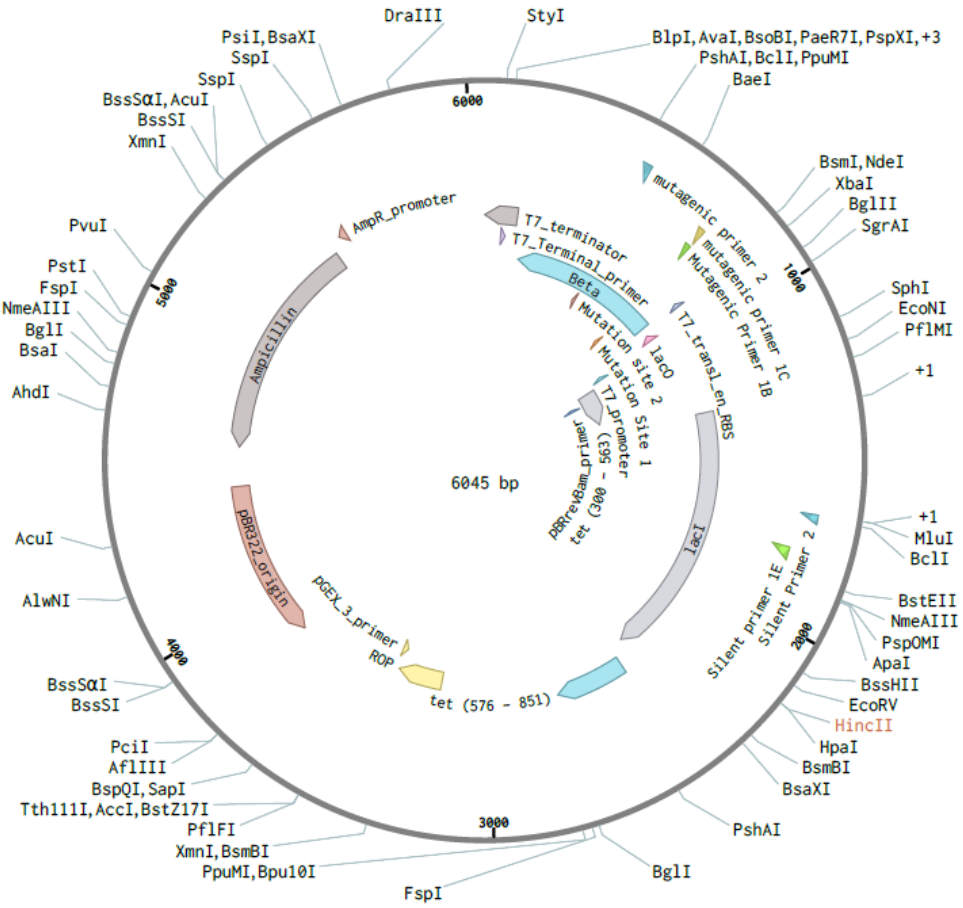


Figure 3.5: Plasmid map of pET-21a(+) with the beta subunit gene sequence inserted into the MCS between NdeI and XhoI. *Mutation site*: target sites for iterative saturation mutagenesis. Potential sites for primer pairs used during megaprimer PCR are indicated, numbered as set 1 and set 2. Mutagenic primers 1C and 1B are variations on the primer used to introduce mutations at mutation site 1. Image generated using the Benchling workspace at https://benchling.com (Benchling, 2017).

Frederick (2013) attempted to improve expression of the alpha subunit by mutating the RBS to one more easily recognised by the ribosome of *E. coli.* In this study, the native ribosome binding sites were removed from the alpha and beta subunit genes and the genes were incorporated into the plasmids from their start codons only. This was to allow the RBS of each plasmid to be the only site where ribosome binding could occur. However, as the chaperone sequence was within the alpha subunit, the chaperone RBS was not altered.

Transformation of vectors into *E. coli* BL21(DE3)

Supercompetent *E. coli* BL21(DE3) cells were produced using a protocol originally based on Hanahan *et al.* (1991). All glassware, plasticware and buffers were chilled before use and cells were kept on ice during the protocol, except where specified. A flask containing 250 ml SOB media (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% (w/v) NaCl, 2.5 mM KCl, 10 mM $MgCl_2$) was inoculated with a 1 ml stock of either *E. coli* JM109 or BL21(DE3) cells and grown for 16 hours at 20°C until an OD_{600} of ~0.3 was obtained. Readings were taken on a Boeco S-20 UV /Visible spectrophotometer. The cells were harvested by centrifugation at 3000 x g for 10 minutes at 4°C, resuspended in 40 ml of ice cold CCMB80 buffer, pH 6.4 (10 mM KOAc pH 7.0, 80 mM $CaCl_2.2H_2O$, 20 mM $MnCl_2.4H_2O$, 10 mM $MgCl_2.6H_2O$, 10% glycerol, pH adjusted with HCl) and then incubated on ice for 20 minutes. The cells were centrifuged again at 3000 x g for 10 minutes at 4°C and then resuspended in 10 ml CCMB80 buffer. To obtain an appropriate concentration of cells, 50 µL of the resuspended cells was added to 200 µL of SOC media (SOB media containing 10 mM glucose) and the OD_{600} measured. Additional CCMB80 buffer was added to the resuspended cells until the above test gave an OD_{600} of 1.5. After a final incubation of 20 minutes on ice, the cells were separated into 1 ml aliquots and stored at -80°C.

The transformation protocol was performed according to the protocol received with the plasmids (GenScript, USA). To transform the cells, 300 ng of plasmid DNA in DNase free water was added to 50 µL competent cells which had been thawed on ice for 20 minutes. The tube was placed on ice for 15 minutes and then submerged up to two-thirds of the volume of the tube in a water bath at 42°C for 3 minutes. The tube was then placed back on ice for 2 minutes. A volume of 800 µL of LB broth (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) was added to the tube and it was incubated at 37°C and 200 rpm for 40 minutes. The cells were then spread on LB agar plates (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 15 g/L agar) containing antibiotic in order to select for transformed cells, with plates containing 100 µg/ml ampicillin

to select for pET21a(+)-beta or 50 µg/ml kanamycin to select for pRSFDuet-1-alpha-chap, or a combination of both antibiotics to select for transformants containing both plasmids.

The vectors were transformed individually into JM109 to maintain the plasmids during storage. The vectors were also transformed into BL21(DE3) individually to yield BL21(DE3)-α-chap and BL21(DE3)-β recombinant cells, containing a single plasmid each. Each single-plasmid BL21(DE3) line was then made competent again and then transformed a second time with the remaining plasmid in order to produce BL21(DE3)-α-β-chap cells.

Plasmid extraction and restriction digest

Positive clones grown on antibiotic-containing LB agar plates were picked using a sterile pipette tip and placed in 20 ml of LB media at 37°C with shaking at 200 rpm and grown for approximately 16 hours. A plasmid extraction was then carried out using the Zyppy Plasmid MiniPrep kit (Zymo Research, USA). The concentration of the extracted DNA was measured using a Qubit (Thermo Fisher Scientific, USA) and the quality of the DNA was confirmed on a 1% agarose gel (0.8 g agarose in 50 ml 1X TAE buffer, 2 µL EtBr, 90V, 75 min. run time).

Plasmids introduced into BL21(DE3) and JM109 as single plasmids were extracted and the presence of an insert in each plasmid was confirmed by digesting pET-21a(+) with NdeI and XhoI and pRSFDuet-1 with NdeI and PacI. All restriction enzymes used were FastDigest enzymes (Thermo Fisher Scientific, USA). A 20 µL restriction digest was set up (1 µg of plasmid DNA, 2 µL 10X FastDigest buffer, 1 µL of each enzyme, 14 µL DNase free water). The reaction was incubated at 37°C for 5 minutes. The restriction enzymes were then inactivated by placing the reaction tubes in a heating block at 80°C for five minutes. The results were then viewed on a 1% agarose gel (0.8 g agarose in 50 ml 1X TAE buffer, 2 µL EtBr, 90V, 75 minutes run time).

To determine the presence of both plasmids in the recombinant BL21(DE3)-α-β-chap cells, the extracted plasmid DNA was digested using FastDigest HincII (Thermo Fisher Scientific, USA). The restriction digest (1 µg of plasmid DNA, 3 µL 10X FastDigest buffer, 1 µL enzyme, 13 µL DNase free water) was allowed to continue for 5 minutes at 37°C and then the enzyme was inactivated by placing the reaction tube in a

heating block at 65°C for five minutes. The results were then viewed on a 1% agarose gel (0.8 g agarose in 50 ml 1X TAE buffer, 2 µL EtBr, 90 V, 75 min. run time).

Protein expression

The method developed by Frederick (2013) was followed with slight modifications due to equipment limitations. A 20 ml overnight starter culture of BL21(DE3) containing both plasmids was grown in LB broth (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) with 50 µg/ml kanamycin and 100 µg/ml ampicillin at 37°C and shaking at 200 rpm for approximately 16 hours. This was used to inoculate a larger volume of LB broth for protein expression, by adding 400 µL starter culture per 200 ml LB broth containing 50 µg/ml kanamycin and 100 µg/ml ampicillin. This was grown to an OD_{600} of ~0.3 at 37°C with shaking at 200 rpm, following which 0.1 mM CoCl2 and 0.1 mM IPTG were added to induce protein expression. Cultures were then grown for 24 hours at 25°C, with shaking at 200 rpm.

The OD_{600} of the grown culture was measured. The cells were then harvested at 8000 x g for 15 minutes at 4°C in a Sorvall RC 6+ ultracentrifuge and resuspended in an appropriate volume of PBS (8 g/L NaCl, 0.2 g/L KCl, 1.44 /L Na_2HPO_4, 0.24 g/L KH_2PO_4, pH 7.4) to yield an OD of ten times the initially measured OD_{600}. Lysozyme (1 mg/ml) and PMSF (1 mM) were added and the cells were incubated at 4°C for 30 minutes with gentle rotation. The cells were then sonicated for five pulses of 30 seconds per pulse with 60 seconds on ice between pulses. DNaseI (3 units) and $MgCl_2$ (2 mM) were then added to the lysate and the tubes were again incubated at 4°C with gentle rotation for 15 minutes. The lysate was then centrifuged at 8000 x g for 15 minutes at 4°C and the pellet and supernatant were separated. The protein concentration was determined using a Qubit protein assay (Thermo Fisher Scientific, USA).

Determination of optimal protein storage conditions

Following protein extraction, the crude cell lysate was stored under varying conditions for five days prior to biocatalytic analysis in order to determine if storage conditions affected stability and catalytic activity. A fraction of the protein crude lysate was stored at 4°C; a second fraction was mixed 1:1 with 80% glycerol and frozen at -20°C; a third fraction was stored without added glycerol at -20°C. Samples were thawed on ice after ~24 hours. A small amount of substrate (~5 mg) was added to each sample and the reaction was rotated at 4°C. Samples were taken at time 0, 30 mins, and overnight to track the conversion of substrate

to product. This was done using thin layer chromatography and using a dropper to take samples of the mixture for TLC (using 70:30 ethyl acetate to hexane as the solvent).

3.2.2.2 Recombinant expression of NHase using pJOE6731.1 and pJOE6556.1

A second recombinant expression system for NHase was tested in this study and used as a positive control for NHase expression. The cells containing pJOE6731.1 and pJOE6556.1 (Figures 3.6 and 3.7) and the protocol for protein expression were provided by Prof Trevor Sewell and Dr Brandon Weber (Aaron Klug Centre for Imaging and Analysis, University of Cape Town). This protocol also used a two-construct system, with the two plasmids present in a single host cell for growth and protein expression.

Protein expression

An overnight starter culture of 20 ml LB 100 µg/ml ampicillin and 25 µ/ml chloramphenicol was inoculated with a single colony grown on an LB agar plate, or from a glycerol stock. This was grown for ~16 hours at 37°C with shaking at 200 rpm. A larger flask of LB was then inoculated from the starter culture by adding 400 µL starter culture per 200 ml LB broth containing 100 µg/ml ampicillin and 25 µ/ml chloramphenicol. This was grown at 37°C with shaking at 200 rpm until an OD_{600} of ~0.4 was reached. Protein expression was then induced by adding 0.2% (w/v) rhamnose and 1 mM $CoCl_2$ and the culture was then grown for ~24 hours at 25°C with shaking at 200 rpm. The cells were harvested and lysed using the same protocol as the pET21-a(+)/pRSFDuet-1 system (Section 3.2.1.1)

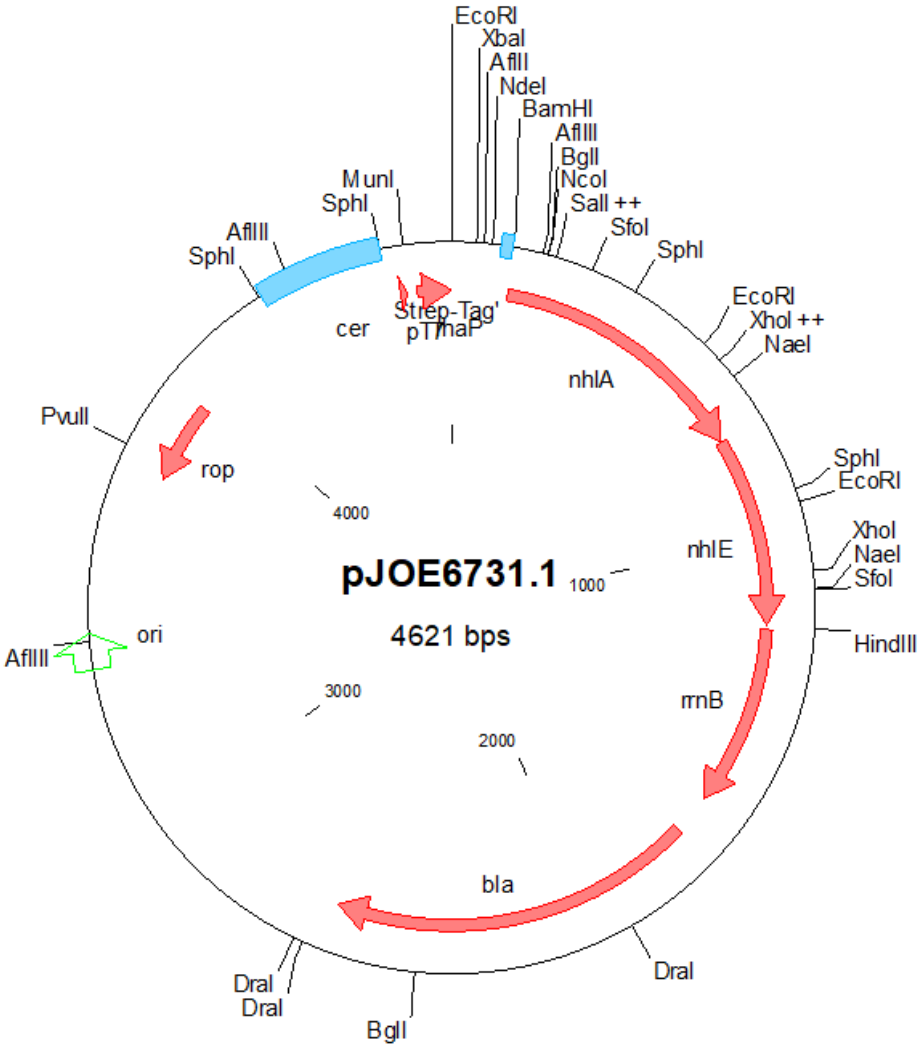


Figure 3.6: pJOE6731.1 vector for recombinant expression of the *R. rhodochrous* L-NHase alpha subunit and chaperone. nhlA: gene for alpha subunit. nhlE: gene for chaperone, bla: beta-lactamase gene conferring ampicillin resistance. Plasmid map provided by Prof Trever Sewell and Dr Brandon Weber (Aaron Klug Centre for Imaging and Analysis, University of Cape Town).

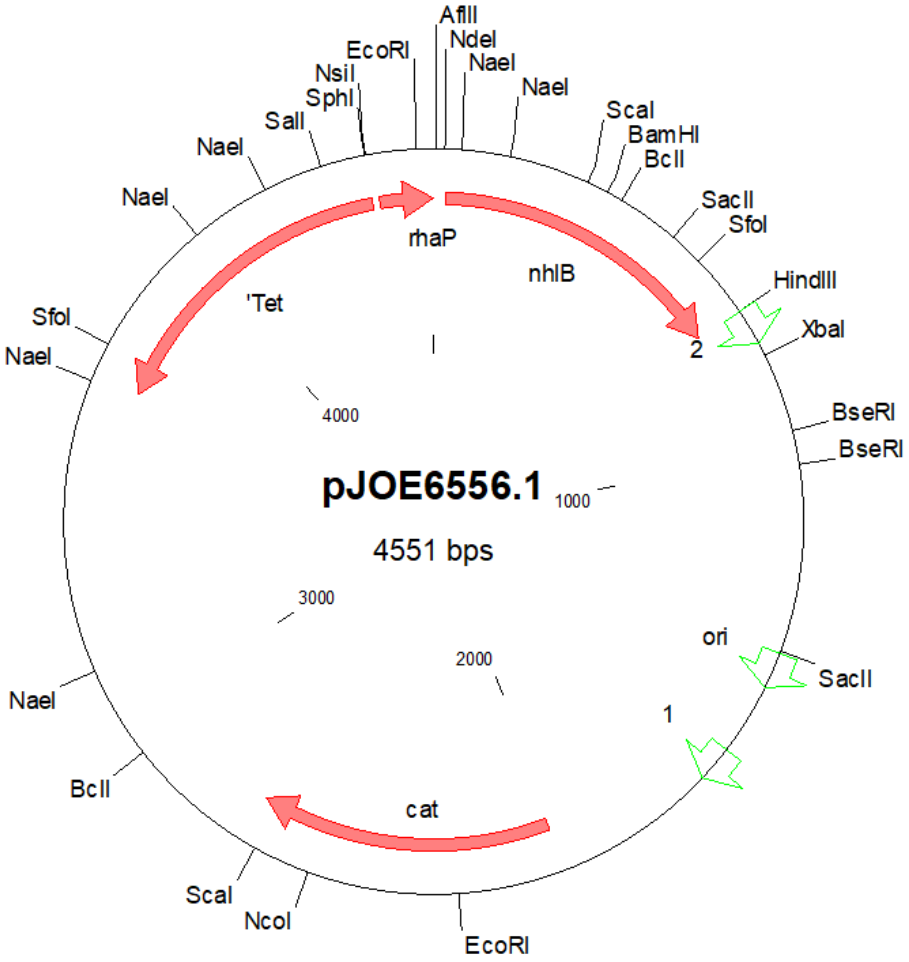


Figure 3.7: pJOE6556.1 vector for recombinant expression of the *R. rhodochrous* L-NHase beta subunit. nhlB: gene for beta subunit, cat: chloramphenicol acetyltransferase gene conferring chloramphenicol resistance. Plasmid map provided by Prof Trever Sewell and Dr Brandon Weber (Aaron Klug Centre for Imaging and Analysis, University of Cape Town).

3.2.3 Saturation mutagenesis by megaprimer PCR

The polymerase chain reaction protocol involved two aspects, designing a mutagenic primer and an antiprimer in order to generate a second megaprimer; and the PCR protocol for site saturation mutagenesis. Both aspects were based on the two-site mutagenesis method devised by Kirsch and Joly (1998) and later improved by Sanchis *et al.*, (2008). The suitability of a whole-plasmid PCR method for introducing mutations is further supported by the research of Cheng et al. (2018), who also introduced multiple mutations into the L-NHase from *R. rhodochrous* J1 in order to improve enantioselectivity using the whole plasmid method of Laible and Boonrod (2009). Whole-plasmid PCR is quicker and simpler as it removes the often tedious cloning step and the potential errors that can occur during cloning prior to protein expression.

3.2.3.1 Primer and megaprimer design

The primers were designed using Benchling (2017), which allowed multiple potential primers to be created and then analysed for the potential to form secondary structures. In this case, primer design was difficult due to the high GC content of the beta subunit gene. Figure 3.8 shows the multiple theoretical primers created and where they were located on the plasmid. The pair primers for the first PCR step consisted of a degenerate, mutagenic forward primer and a silent antiprimer, with a similar T_m to ensure binding to the target in the first step. The codons on the forward primer were randomised using NDT codon degeneracy, where N = Ade/Cyt/Gua/Thy; D = Ade/Gua/Thy; T = Thy. The primer pair for the first PCR step was as follows (underlined codons are the target codons for mutation):

Mutagenic primer (Reverse primer): 5'-CCATG<u>TTC</u>CCGGCGATG<u>GCG</u>CTGG-3'

Start	715
End	738
Length	24
GC	70.83%
Tm	68.2°C

Target codons: TTC (βPhe41) and GCG (βAla37)

Silent Primer (Forward primer): 5'-CCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCG-3'

Start	1779
End	1818
Length	40
GC	45.00%
Tm	65.5°C

The primers in the first PCR step were chosen to bind to the pET-21(a)-β plasmid at a sufficient distance apart to produce a large fragment of the plasmid template. This product is the megaprimer which was used in the second PCR step. The megaprimer was 4953bp, a large section of the 6045bp plasmid from position 703 to position 1796 and should in have theory have contained the added mutations. Using the method adapted by Sanchis *et al.* (2008) for a single two step reaction, the reaction mixture was then immediately subjected to a second round of PCR amplification but annealing occurs at the elongation temperature of the polymerase. Because the megaprimer is so large, it has a much higher T_m than the first set of primers and thus those primers can be left in the reaction mixture without interacting further with the template or megaprimer.

3.2.3.2 PCR conditions

The initial PCR master mix (Table 3.1) was designed based on the KOD Hot Start DNA Polymerase protocol (Novagen, Pretoria, South Africa), with the final DNA concentration based on the protocol by Sanchis *et al.* (2008). During the initial attempt, 50 ng of template DNA was used. After the first PCR step, a 5-10 µL sample was taken for agarose gel visualisation. The sample volume was then appropriately adjusted on the PCR cycler and the second reaction allowed to continue.

Table 3.1: Volumes and final concentrations of reagents in a single sample for megaprimer PCR

Reagent	Final Concentration	Volume in 1 X sample		Volume in 1 X sample	
10X buffer for KOD polymerase	1X	5.00	µL	5,00	µL
25 mM MgSO4	1.5 mM (up to 2.25 mM)	3.00	µL	3,00	µL
dNTPS (2mM each)	0.2 mM each	5.00	µL	5,00	µL
Forward primer (10 µM)	0.3 µM (15 pmol)	1.50	µL	1,50	µl
Reverse Primer (10 µM)	0.3 µM (15 pmol)	1.50	µL	1,50	µL
DNA	10-50 ng/µL (50-100 ng total DNA)	1,00 to 7,00	µL	1,00 to 2,00	µL
KOD polymerase	0.02 U/µL	1,00	µL	1,00	µL
PCR grade water (make up to volume)	N/A	1,00 to 7,00	µL	0,00 to 1,00	µL
TOTAL VOLUME		**25**	**µL**	**17**	**µL**

Initial cycling conditions, including the extension time per kb, were also mainly based on Sanchis *et al.* (2008), although an initial step of 95°C for 2 mins was added to activate the KOD polymerase, as required by the KOD protocol (Novagen, Pretoria, South Africa). The cycling conditions initially used are shown in Table 3.2. The first attempt at two-step PCR was not successful as no PCR product was present on a 1% agarose gel after either PCR step and therefore a number of alterations were made to the method and reagents. For the second attempt, the total concentration of template DNA was increased to 80ng and the volume of the other reagents reduced to compensate for the increased DNA volume. Due to the high cost and small volume of KOD polymerase available, the reaction volume was not increased. As well as this, the extension time was decreased to 15 sec/kb at 68°C.

Table 3.2: Cycling conditions for megaprimer PCR (first attempt)

Step	Temp	Time	Extension time per kb
Polymerase Activation	95 °C	2 min	
Initial Denaturing	95 °C	3 min	
5 cycles			
Denaturing	95 °C	30 sec	
Annealing	Tm-5°C	1 min	
Extension	72 °C	5 min	1 min/kb
20 cycles			
Denaturing	95 °C	30 sec	
Extension	68 °C	6 min 10 sec	1 min/kb
Final Extension	68 °C		2 min/kb
Hold	∞	12 °C	

3.2.3.3 Sequencing of PCR product

PCR results of both PCR steps were initially visualised on a 1% agarose gel, after which the results of the second PCR were sent to Inqaba Biotec (Pretoria, South Africa) for sequencing to determine the structure of the final PCR product. Although the full plasmid was subject to the PCR reaction, due to financial constraints, only the insert containing expected mutations was sequenced using standard T7 promoter and T7 terminator primers. It was assumed that all other regions of the plasmid should remain unmutated or if mutations were present, then the plasmid would be non-viable and not functional.

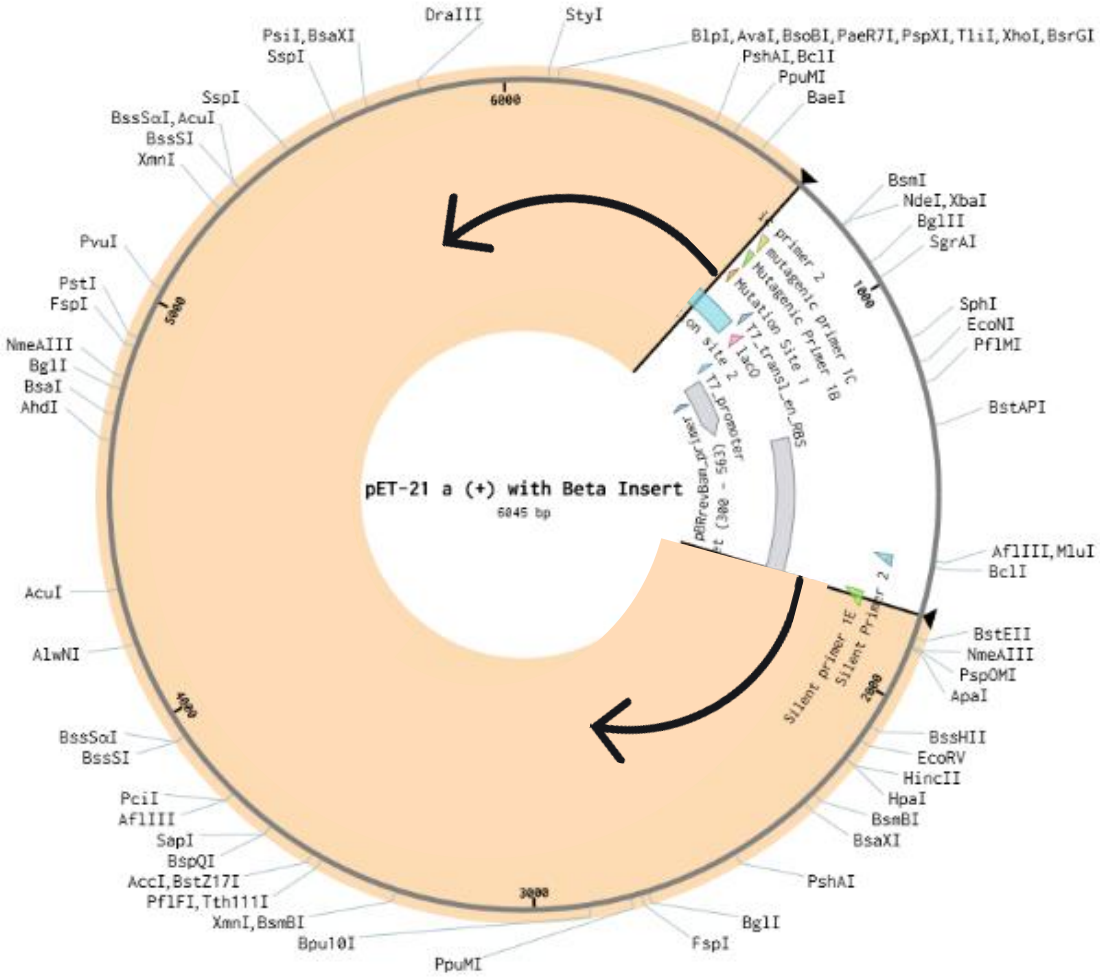


Figure 3.8: Plasmid map showing the position of mutagenic primers and the megaprimer formed in the first PCR step of iterative saturation mutagenesis. Mutagenic primer 1B and Silent primer 1E formed the pair of primers used to introduce mutations into Mutation Site 1. Black arrows show the direction of replication using these primers. Orange shading indicates the fragment of the plasmid replicated in the first PCR step which was used as the megaprimer in the second PCR step.

3.3 RESULTS AND DISCUSSION

3.3.1 Recombinant expression of NHase using pRSFDuet-1 and pET21a(+)

3.3.1.1 Restriction digests

To confirm the presence and structural integrity of the plasmids transformed into both JM109 and BL21(DE3), restriction digests were performed on extracted plasmid DNA. The individual digests of each plasmid clearly showed the expected digest results (Figure 3.9). In both cases, only two fragments were present, one short fragment indicating that the insert was correctly cut from the plasmid, and one longer fragment from the vector. Both NdeI/XhoI digests produced three bands. The insert was expected to produce a 1078 bp band and the linearised plasmid was expected to be 3698bp. As expected, a band of about 1100bp and one of about 3600bp was seen. The larger band at about 5000bp was potentially uncut plasmid in some form (supercoiled or relaxed). These results also indicated that no unwanted NdeI and XhoI cut sites had been introduced into the pRSF-Duet plasmid carrying the α-chaperone insert and that no unwanted NdeI and PacI cut sites were present in the pET21-a(+) plasmid carrying the β insert. The NdeI/PacI digests showed similar results, with expected bands of 682bp and 5363bp and observed bands at ~600bp and ~6000bp.

The HincII digest was done only on BL21(DE3) expression cells, as JM109 maintenance cells were only transformed with a single plasmid each. In this case, the two plasmids were digested simultaneously in the same reaction mixture using only HincII because it gave an easily recognisable digest pattern. The results showed all of the expected bands (Figure 3.10), although the two smallest bands (at ~355 bp and ~440 bp) were very faint, possibly due to a lower concentration and the small size of the DNA fragments resulting in very faint stain. However, they were assumed to be present based on a very light smear where the bands were expected, as well as on the presence of the other expected band sizes. Later protein expressed from the dual-construct cells showed good activity, further supporting that this double digest was successful.

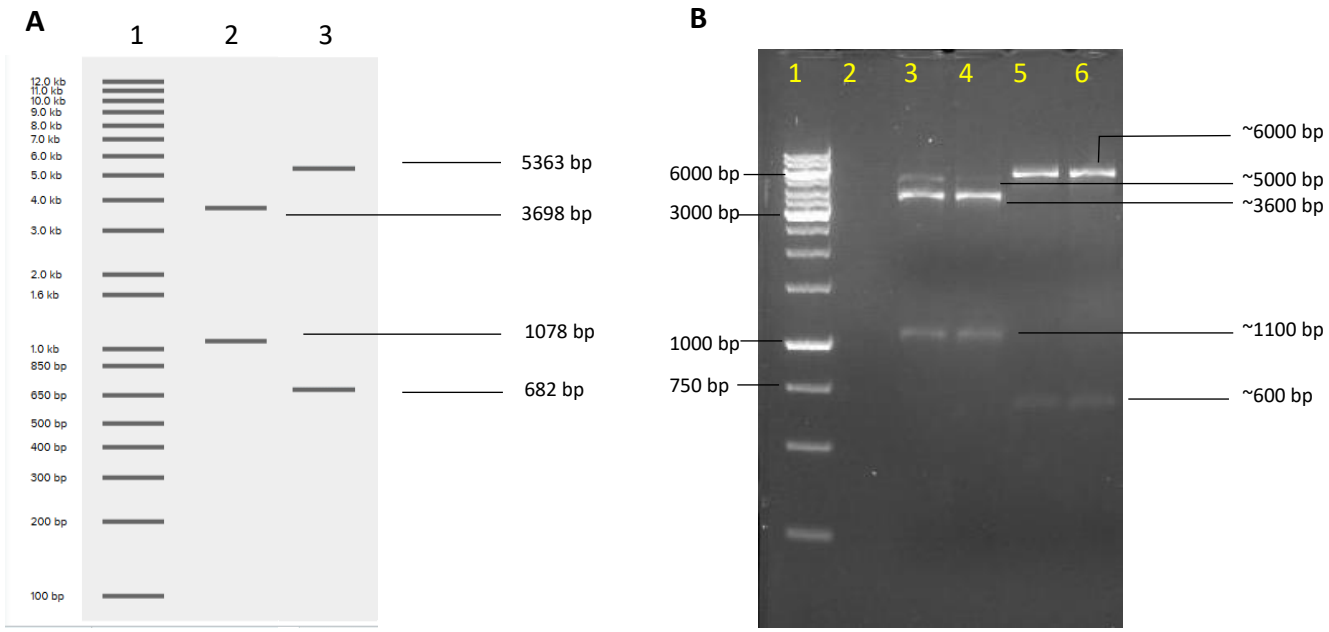


Figure 3.9: Restriction digests from single-plasmid transformations. A, **Expected results** using Benchling (2017**)**. Lane 1, MWM; Lane 2, pRSFDuet-1-αC digested with NdeI, XhoI; Lane 3, pET21a(+)-β digested with NdeI, PacI. **B, actual results** on 1 % agarose gel. Lane 1, GeneRuler 1kb; Lane 2, blank; Lane 3, pRSFDuet-1-αC from BL21(DE3); Lane 4, pRSFDuet-1-αC from JM109; Lane 5, pET21a(+)-β from BL21(DE3); Lane 6, pET21a(+)-β from JM109. Lanes 3 and 4 digested with NdeI, XhoI; Lanes 5 and 6 digested with NdeI, PacI.

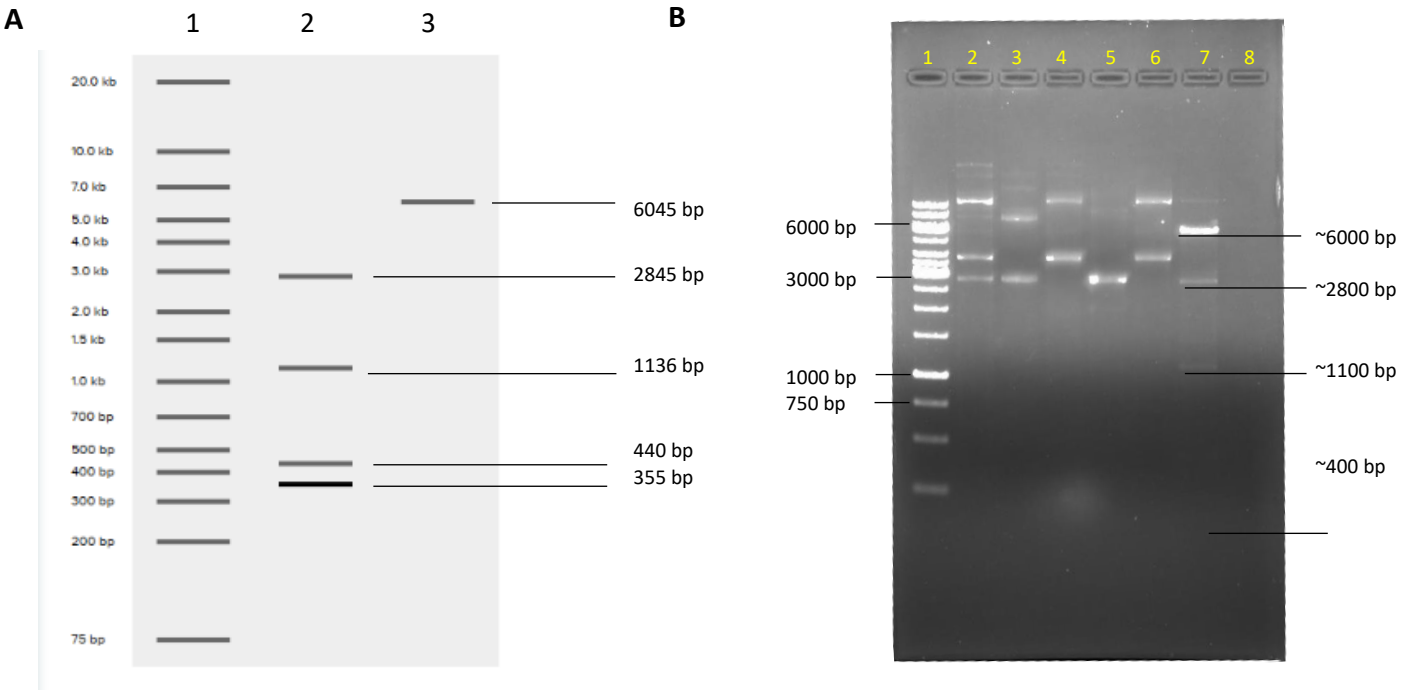


Figure 3.10: Combined restriction digest of two expression plasmids from *E. coli* BL21(DE3). A, **Expected results** using Benchling (2017**)**. Lane 1, MWM; Lane 2, pRSFDuet-1-αC digested with HincII; Lane 3, pET21a(+)-β digested with HincII. **B, actual results** on 1 % agarose gel. Lane 1, GeneRuler 1kb; Lane 2, undigested pRSFDuet-1-αC and pET21a(+)-β from BL21(DE3) dual construct; Lane 3, undigested pRSFDuet-1-αC from single-transformed BL21(DE3); Lane 4, undigested pET21a(+)-β from single-transformed BL21(DE3); Lane 5, undigested pRSFDuet-1-αC from single-transformed JM109; Lane 6, undigested pET21a(+)-β from single-transformed JM109; Lane 7, HincII digest of pRSFDuet-1-αC and pET21a(+)-β from BL21(DE3) dual construct.

3.3.1.2 Protein expression

Nitrile hydratase protein was overexpressed from both BL21(DE3) cells containing the two-construct system, as well as from pJOE expression cells provided by Prof Trevor Sewell and Dr Brandon Weber, containing the pJOE6731.1 and pJOE6556.1 plasmids. Expression cells were cultured to an OD_{600} of ~0.3, at which time protein expression was induced in the cultures, while one culture vessel received no inducer. The uninduced cultures were compared in order to determine if there was any leaky expression, as well as to confirm that nitrile hydratase was indeed overexpressed by the induced cultures. Initial rounds of expression included both pJOE and the BL21(DE3)-two construct cells. Protein was harvested following expression by cell lysis and then centrifuged to separate the pellet and supernatant. Samples of the pellet and supernatant from both induced and uninduced cells were taken and were separated using SDS-PAGE.

The SDS-PAGE results including nitrile hydratase expression samples from the pJOE cells were not successful due to experimental error, following which the samples became degraded and the protein was not re-expressed. pJOE expression is therefore not shown here. Following this, only BL21(DE3) cells containing the two-construct system were utilised and the following SDS-PAGE results are shown in Figure 3.11. The samples were left overnight in reducing sample buffer, resulting in degraded protein causing a smeared pattern. However, the supernatant and pellet from the induced two-construct cells displayed fairly clear wide bands in the expected regions for the NHase alpha and beta subunits, at approximately 23 kDA for the alpha subunit, 26 kDA for the beta subunit and 16 kDa for the chaperone, as reported by Frederick (2013). A thick band in the pellet was potentially caused by the unclear separation and high concentration of the alpha and beta subunits. Although the majority of expressed NHase appeared to be in the pellet, later biocatalytic tests on the protein supernatant produced activity with full substrate conversion.

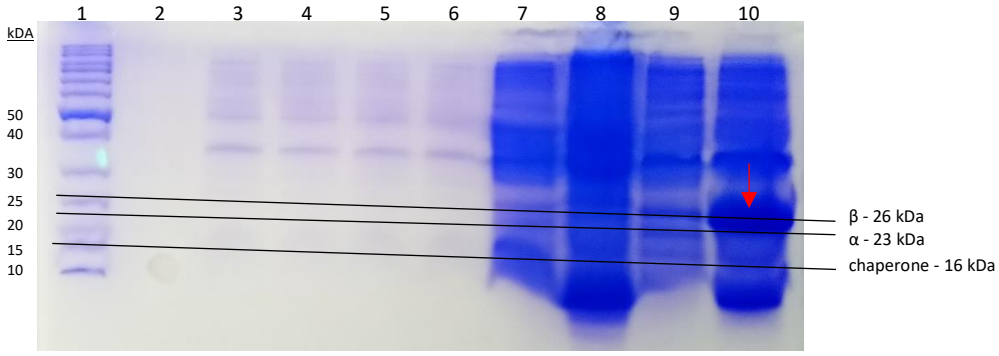


Figure 3.11: SDS-PAGE separation of induced and uninduced protein fractions from BL21(DE3) containing pRSFDuet-1-αC and pET21a(+)-β. S = supernatant; P = pellet. αβC = BL21(DE3) transformed with the two plasmid construct. Lane 1, PageRuler Unstained Protein Ladder (ThermoFisher, Massachusetts, USA); Lane 2, clear lane; Lane 3, BL21(DE3) control cells (S); Lane 4, BL21(DE3) control cells (P); Lane 5, αβC uninduced cells (S); Lane 6, αβC uninduced cells (P); Lane 7, BL21(DE3) control cells exposed to inducer (S); Lane 8, BL21(DE3) control cells exposed to inducer (P); Lane 9, αβC cells exposed to inducer (S); Lane 10, αβC cells exposed to inducer (P). The red arrow shows a large band possibly made by two separate bands, one for the 23 kDa α-subunit and one for the 26 kDa β-subunit.

3.3.1.3 Determination of optimal protein

3.3.1.4 storage conditions

Lysed protein samples stored under varying conditions were tested for activity to determine if storage conditions affected stability and activity. A fraction of the protein crude lysate was stored at 4°C; a second fraction was mixed 1:1 with 80% glycerol and frozen at -20°C; a third fraction was stored without added glycerol at -20°C. All samples were stored for 24 hours. A small amount of substrate (~5 mg) was added to each sample and all of the samples showed complete conversion of 4-methoxybenzonitrile (3a) and 3-Amino-3-p-tolylpropanenitrile (**2c**) within 30 minutes, while all the samples showed partial conversion of 3-Amino-3-(4-methoxyphenyl) propanenitrile (**3c**), even when left at 4°C for 24 hours with rotation. Therefore the storage conditions were deemed to be equally valid for short term protein storage (~24

hours), although this remains to be tested for longer periods. It was not possible to quantify the conversion of **3c** in each case, although visually it was not noticeably different as the diameter of the product spots on the TLC plate was similar for all sample reactions.

Stored glycerol stocks of the transformed cells have been grown and used to express active NHase for further bioconversion studies, as described by Soobben (2020). Therefore although the best storage conditions were not conclusively determined, this indicates that the cells were successfully transformed and appropriately stored.

3.3.2 Saturation mutagenesis by megaprimer PCR

It was not possible to visualise the results of either the first or second mutagenic PCR steps using 1% agarose gel electrophoresis. The gels showed no bands despite increasing the sample volume loaded on the gel as well as the concentration of ethidium bromide. Therefore to identify whether product was formed, the results of the PCR reaction were sent to Inqaba Biotec (Pretoria, South Africa) for sequencing.

As a PCR product was in fact present, there were many modifications that could have been made to the PCR protocol that may have resulted in the correct region being mutated. There were two main complications during this technique. First, the KOD polymerase is very expensive and is packaged in very small quantities, hence it was not possible to test many different samples, temperatures and other parameters repeatedly. This may have been resolved by developing both steps of the mutagenic PCR using a standard Taq polymerase, and then only using the KOD polymerase once the best method was established. The second problem in method development was that it was not clear whether the method of Sanchis *et al.* or the standard KOD polymerase method should be used, or whether a combination of steps from both methods should have been used. This would have been resolved by varying the method and testing multiple times for a PCR product. The KOD protocol recommends more cycles per PCR step (20 to 40 cycles) than the Sanchis *et al.* protocol (five cycles in step one and 20 cycles in step 2), thus an increased number of cycles may have provided a visible PCR product on an agarose gel and thus made method development easier.

Sequencing confirmed the presence of a PCR product, however an alignment of the original beta subunit gene with the PCR results (Figure 3.12) showed that mutations were present in the wrong part of the

target beta subunit gene while the target regions remained unchanged. The mutations occurred 90 nucleotides upstream of the target region, at codons βHis5 and βAsp6. The mutagenic primer had bound with a hairpin shape, introducing the mutations at two neighbouring residues, instead of three residues apart as intended. Sixteen sample reactions were tested and nine presented mutations at these codons, while one sample had both of these mutated codons, with an additional thymine nucleotide inserted between the codons. Such an insertion was not desirable and most likely would have resulted in an inactive protein. The intention with NDT codon degeneracy is to achieve an even distribution of 12 amino acids (Gly, Leu, Ile, Val, Tyr, His, Phe, Asn, Asp, Cys, Arg and Ser), as encoded by the reduced number of codons available using NDT degeneracy. From the 16 samples sequenced, it is possible to see that a balanced distribution of codons and therefore amino acids was not achieved. It is possible that the expected distribution may have been observed if a larger sample size was used, however it is more likely in this case that the distribution was not observed as the PCR conditions were not yet optimised. This is supported by Isalan (2006) who states that when a PCR technique is optimised, 80% to 90% of clones should contain mutations.

Looking at residue 5, the original codon was CAC. In all the ISM samples, the first and third nucleotides remained the same as the unmutated sequence (C in both positions). The second nucleotide was A in 8 out of 16 samples (50% occurrence) and M (A/C) in the other 50% of samples, whereas the second nucleotide was expected to be a D (A/G/T) in all of the samples. This shows that in codon 6, A occurred more frequently at the second position. Overall, this meant that the original codon was still abundant and had not been successfully mutated. Looking at residue 6, the original codon was GAT. In 4/16 (25%) of samples, the nucleotide was a G. In 10/16 (62.5%) of samples, there was a K (G/T) and in 1/16 (6.25%) there was an S (G/C). This shows that the most common nucletotide was the original G and not an equal distribution of all four nucleotides as was expected. In the second positon, the original A occurred 3/16 times (18.75%). In the second position, W (A/T) occurred 13/16 times (81.25%). Although this did indicate presence of A and T this was not in an even distribution as would be expected in NDT codon degeneracy. In the third nucleotide position, a T occurred 100% of the time, and although this is in line with the expected NDT codon degeneracy, it is also the same as the wild type codon. Bias towards the wild type codon is possible as reported by (Acevedo-Rocha *et al.*, 2015).

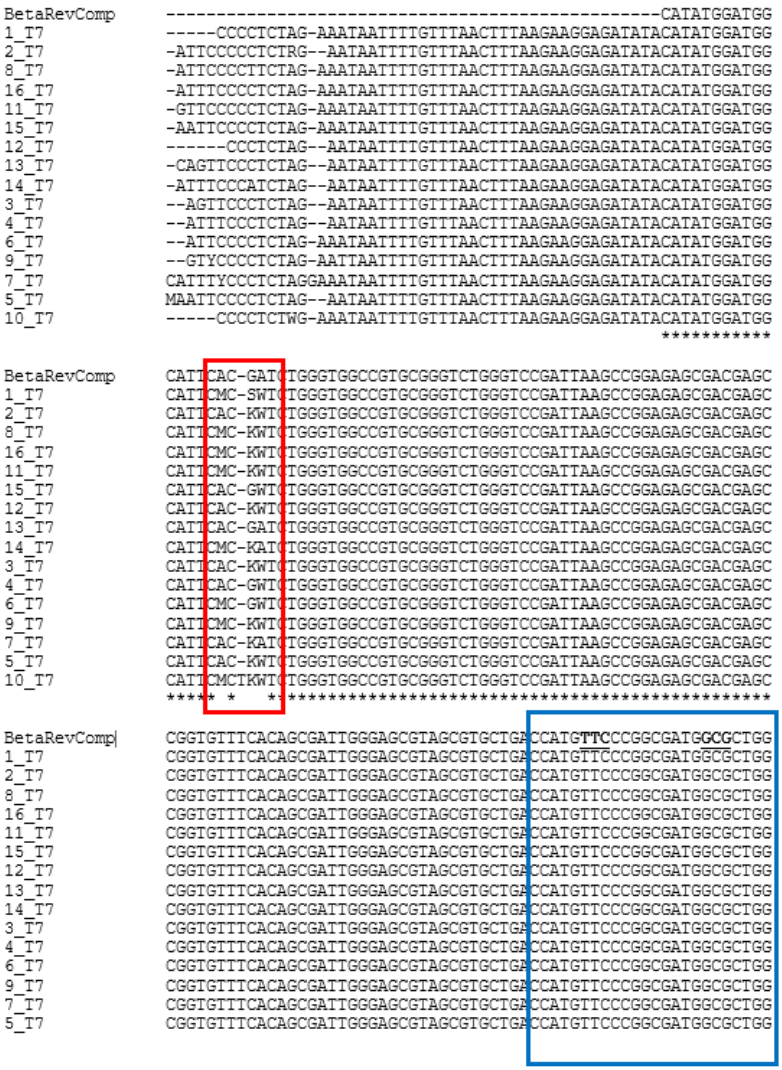


Figure 3.12: Excerpt of the MAFFT alignment of original BAA-870 L-NHase beta subunit gene with the 16 ISM PCR samples. The blue box shows the target codons for mutation (underlined and bold within the blue box). The red box shows the codons actually mutated during ISM PCR (underlined and bold within the red box). The target and actual mutated codons are also shown within the beta subunit protein sequence below the alignment. Blue: target codons. Red: actual mutated codons as identified by sequencing of the PCR product. Alignment done using MAFFT at https://www.ebi.ac.uk/Tools/msa/mafft/ (Madeira *et al.*, 2019).

The sequencing analysis was necessary to determine if a PCR product was present, however it is also a small-scale example of a technique dubbed Quick Quality Control (QQC) for determining codon diversity (Sanchis *et al.*, 2008; Kille *et al.*, 2013), where the entire plasmid library transformed with the mutated DNA is pooled and sequenced as a single reaction. As well as screening transformants for the desired protein characteristic, QQC is necessary to determine if the target codons were mutated in expected ratios. This informs whether or not the library covers a sufficient range of sequence space that it might reasonably contain a "hit". QQC should in fact be conducted before expending further time and money on a screening technique (Acevedo-Rocha *et al.*, 2015; Li *et al.*, 2018). For more detailed statistical analyses of ISM libraries, see Sullivan *et al.* (2013) and Acevedo-Rocha *et al.* (2015).

Codon alphabet selection is crucial to ensure that the reduced number of amino acids encoded by the library will be effective in improving the chosen enzyme characteristic. It has previously been shown that NDT codon degeneracy is effective when improving enzyme enantioselectivity (Reetz *et al.*, 2008; Reetz, 2011). The NDT method also greatly reduces the screening effort when screening transformants (Reetz and Carballeira, 2007), especially when focusing on enzyme enantioselectivity (Acevedo-Rocha *et al.*, 2014). The screening effort is the number of transformants needed to be tested in order to cover 95% of the possible mutants in a generated library (Reetz, 2016). When targeting a two-residue site using traditional NNK codon degeneracy (where N is any nucleotide and K is G or T), the number of transformants to be tested is 3066. For NDT codon degeneracy, it is reduced to 430 transformants (Reetz *et al.*, 2008), which is manageable when paired with a medium- or high throughput screening method. Unfortunately in this study, there was insufficient time to develop a screening assay, or indeed to develop the ISM PCR method fully. Assay development was also more complex as it would have been more challenging to create an assay for enantioselectivity rather than just basic enzyme activity.

Due to time constraints, a library of expressed NHase enzymes was not produced from the mutated NHase genes. However, the mutations introduced to βHis5 and βAsp6 were considered unlikely to be effective in altering enantioselectivity as they fall within a random coil region close to the surface of the enzyme and distant from the active site tunnel. It is possible that the position of the side chains might interact with nearby residues and thus distally cause subtle differences to the active site channel. However, determining exact side chain conformation requires more complex modelling and was not possible in this study. According to the study by Huang *et al.* (1997) on the Fe-type NHase, the N-terminal arm of the beta subunit wraps around the central part of the alpha subunit, so there may be a change in the binding

between the subunits with the mutations to βHis5 and βAsp6. This remains to be determined for the Co-type L-NHase in future work. In order to determine the effects of any mutations on the enzyme's enantioselectivity, it is necessary to isolate and purify the enzyme and to conduct kinetic studies (Reetz *et al.*, 2006). Further studies once an improved enzyme is produced involve systematically testing each residue to determine if it is actually contributing to the improved enantioselectivity (Reetz *et al.*, 2006).

3.4 CONCLUSIONS

This study was important in laying out the basic techniques to introduce mutations into the low molecular mass nitrile hydratase from R. rhodochrous ATCC BAA-870 to improve its enantioselectivity, however many aspects can be improved upon, while others remain to be developed.

Initial transformation and expression was successful. It was confirmed that the plasmids were genetically intact prior to expression. Problems related to sample storage and time constraints resulted in poor visualisation of protein expression, however biocatalytic tests proved the presence of active protein. The SDS-PAGE gels also showed that the biocatalytic activity was not due to a nitrile hydratase already present in uninduced cells as the uninduced cells did not have bands where NHase subunits would be expected. The NHase proved to be stable in multiple short term storage conditions including freezing slowly at -20°C, freezing slowly at -80°C and snap freezing with liquid nitrogen before being stored at -80°C, but longer storage under the same conditions needs to be tested. Further successful biocatalytic conversion has been achieved using the same parent stock of transformed BL21(DE3) cells made during the current study.

Although the work of Frederick (2013) concluded that a two-plasmid construct was preferable for the expression of the L-NHase from *R. rhodochrous* ATCC BAA-870, and the current study did produce sufficient amounts of soluble protein, the two-plasmid system complicates the experimental procedure considerably, in particular where plasmid transformation is needed. Therefore a single-plasmid system may be preferable in an industrial context where speed and efficiency are necessary. A single-plasmid expression system has been used to express large amounts of soluble recombinant Co-type NHase from *Bacillus pallidus* RAPc8 (Tsekoa *et al*., 2004; van Wyk, 2008; Kianja, 2016), and this system may be preferable for future directed evolution studies towards beta-substituted nitriles.

Saturation mutagenesis successfully introduced mutations into the protein, but not in the expected place (βPhe41 and βAla37). The mutations were on the N-terminal of the protein, at residues βHis5 and βAsp6. This is a less structured region more exposed to solvent. Based on knowledge about the less ordered regions of proteins, it was assumed that the mutations would not be functional and therefore the biocatalytic activity of the mutated protein was not tested. This may affect the enzyme via distant effects, or by altering the way the beta subunit wraps around the alpha subunit but remains to be determined in future work. Future studies should focus on testing various mutagenic primers, as well as determining the best size and position of megaprimer. The PCR protocol further needs to be optimised, including possibly increasing the yield. Iterative saturation mutagenesis needs to be performed with multiple cycles, testing the enantioselectivity after each round and choosing the best hit to subject to the next round of mutagenic PCR. If improved enantioselectivity is not achieved, it might be necessary to try a different amino acid alphabet. A screening assay needs to be developed for enantioselectivity, which would allow the required number of transformants to be tested after each round of mutation. Finally, each residue in the improved enzyme can be tested to determine its role in altering enzyme activity and whether the mutations are additive or cooperative in any way.

3.5 REFERENCES

Acevedo-Rocha, C.G., Hoebenreich, S. and Reetz, M.T., 2014. Iterative Saturation Mutagenesis: A Powerful Approach to Engineer Proteins by Systematically Simulating Darwinian Evolution. In: E.M.J. Gillam, J.N. Copp and D. Ackerley, eds. *Directed Evolution Library Creation*. [online] New York, NY: Springer New York.pp.103–128. Available at: <http://link.springer.com/10.1007/978-1-4939-1053-3_7> [Accessed 11 Apr. 2016].

Acevedo-Rocha, C.G., Reetz, M.T. and Nov, Y., 2015. Economical analysis of saturation mutagenesis experiments. *Scientific Reports*, 5(1), p.10654.

Arif, N.M., Ahmad, S.A., Syed, M.A. and Shukor, M.Y., 2012. Isolation and characterization of a phenol-degrading Rhodococcus sp. strain AQ5NOL 2 KCTC 11961BP. *Journal of Basic Microbiology*, 53(1), pp.9–19.

Arnold, F.H. and Volkov, A.A., 1999. Directed evolution of biocatalysts. *Current Opinion in Chemical Biology*, 3(1), pp.54–59.

Barettino, D., Feigenbutz, M., Valcarcel, R. and Stunnenberg, H.G., 1994. Improved method for PCR-mediated site-directed mutagenesis. *Nucleic acids research*, 22(3), pp.541–542.

Cheng, Z., Peplowski, L., Cui, W., Xia, Y., Liu, Z., Zhang, J., Kobayashi, M. and Zhou, Z., 2018. Identification of key residues modulating the stereoselectivity of nitrile hydratase toward rac-mandelonitrile by semi-rational engineering. *Biotechnology and Bioengineering*, 115(3), pp.524–535.

Dabbs, E.R., Gowan, B. and Andersen, S.J., 1990. Nocardioform arsenic resistance plasmids and construction of Rhodococcus cloning vectors. *Plasmid*, 23(3).

DeSantis, G., Wong, K., Farwell, B., Chatman, K., Zhu, Z., Tomlinson, G., Huang, H., Tan, X., Bibbs, L., Chen, P., Kretz, K. and Burk, M.J., 2003. Creation of a Productive, Highly Enantioselective Nitrilase through Gene Site Saturation Mutagenesis (GSSM). *Journal of the American Chemical Society*, 125(38), pp.11476–11477.

Fei, Q., Wewetzer, S.J., Kurosawa, K., Rha, C. and Sinskey, A.J., 2015. High-cell-density cultivation of an engineered Rhodococcus opacus strain for lipid production via co-fermentation of glucose and xylose. *Process Biochemistry*, 50(4), pp.500–506.

Frederick, J., 2007. *Characterisation of the nitrile biocatalytic activity of Rhodococcus rhodochrous ATCC BAA-870*. [online] Available at: <http://wiredspace.wits.ac.za/handle/10539/2016> [Accessed 12 Feb. 2016].

Frederick, J., 2013. *Genetic characterization of Rhodococcus rhodochrous ATCC BAA-870 with emphasis on nitrile hydrolysing enzymes*. [online] University of Cape Town. Available at: <http://open.uct.ac.za/handle/11427/4262> [Accessed 12 Feb. 2016].

Hammer, S.C., Knight, A.M. and Arnold, F.H., 2017. Design and evolution of enzymes for non-natural chemistry. *Current Opinion in Green and Sustainable Chemistry*, 7, pp.23–30.

Hanahan, D., Jessee, J. and Bloom, F.R., 1991. Plasmid transformation of Escherichia coli and other bacteria. *Methods in Enzymology*, 204, pp.63–113.

Hashimoto, Y., Nishiyama, M., Yu, F., Watanabe, I., Horinouchi, S. and Beppu, T., 1992. Development of a host-vector system in a Rhodococcus strain and its use for expression of the cloned nitrile hydratase gene cluster. *Microbiology*, 138(5), pp.1003–1010.

Herwig, L., Rice, A.J., Bedbrook, C.N., Zhang, R.K., Lignell, A., Cahn, J.K.B., Renata, H., Dodani, S.C., Cho, I., Cai, L., Gradinaru, V. and Arnold, F.H., 2017. Directed Evolution of a Bright Near-Infrared Fluorescent Rhodopsin Using a Synthetic Chromophore. *Cell Chemical Biology*, 24(3), pp.415–425.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R., 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, 77(1), pp.51–59.

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

Isalan, M., 2006. Construction of semi-randomized gene libraries with weighted oligonucleotide synthesis and PCR. *Nature Protocols*, 1(1), pp.468–475.

Jaeger, K.-E., Eggert, T., Eipper, A. and Reetz, M.T., 2001. Directed evolution and the creation of enantioselective biocatalysts. *Applied Microbiology and Biotechnology*, 55(5), pp.519–530.

Kesseler, M., Dabbs, E.R., Averhoff, B. and Gottschalk, G., 1996. Studies on the isopropylbenzene 2,3-dioxygenase and the 3-isopropylcatechol 2,3-dioxygenase genes encoded by the linear plasmid of Rhodococcus erythropolis BD2. *Microbiology*, 142(11), pp.3241–3251.

Kianja, J.M. 2016 MSc UCT *Investigation of the determinants of thermal stability of the nitrile hydratase from Geobacillus pallidus RAPc8*. [online] Available at: <http://hdl.handle.net/11427/24445> [Accessed 9 Aug. 2020].

Kille, S., Acevedo-Rocha, C.G., Parra, L.P., Zhang, Z.-G., Opperman, D.J., Reetz, M.T. and Acevedo, J.P., 2013. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synthetic Biology*, 2(2), pp.83–92.

Kille, S., Zilly, F.E., Acevedo, J.P. and Reetz, M.T., 2011. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nature Chemistry*, 3(9), pp.738–743.

Kirsch, R.D. and Joly, E., 1998. An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Research*, 26(7), pp.1848–1850.

Kurosawa, K., Radek, A., Plassmeier, J.K. and Sinskey, A.J., 2015. Improved glycerol utilization by a triacylglycerol-producing Rhodococcus opacus strain for renewable fuels. *Biotechnology for Biofuels*, 8, p.31.

Laible, M. and Boonrod, K., 2009. Homemade Site Directed Mutagenesis of Whole Plasmids. *Journal of Visualized Experiments*, [online] (27). Available at: <http://www.jove.com/index/Details.stp?ID=1135> [Accessed 23 May 2018].

Li, A., Acevedo-Rocha, C.G. and Reetz, M.T., 2018. Boosting the efficiency of site-saturation mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy. *Applied Microbiology and Biotechnology*. [online] Available at: <http://link.springer.com/10.1007/s00253-018-9041-2> [Accessed 29 May 2018].

Lin, Z.-J., Zheng, R.-C., Wang, Y.-J., Zheng, Y.-G. and Shen, Y.-C., 2012. Enzymatic production of 2-amino-2,3-dimethylbutyramide by cyanide-resistant nitrile hydratase. *Journal of Industrial Microbiology & Biotechnology*, 39(1), pp.133–141.

Lutz, S., 2010. Beyond directed evolution—semi-rational protein engineering and design. *Current Opinion in Biotechnology*, 21(6), pp.734–743.

Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D. and Lopez, R., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47(W1), pp.W636–W641.

Martinez, S., Wu, R., Sanishvili, R., Liu, D. and Holz, R., 2014. The Active Site Sulfenic Acid Ligand in Nitrile Hydratases Can Function as a Nucleophile. *Journal of the American Chemical Society*, 136(4), pp.1186–1189.

Miyazaki, K., 2011. Chapter seventeen - MEGAWHOP Cloning: A Method of Creating Random Mutagenesis Libraries via Megaprimer PCR of Whole Plasmids. In: C. Voigt, ed. *Methods in Enzymology*, Synthetic Biology, Part BComputer Aided Design and DNA Assembly. [online] Academic Press.pp.399–406. Available at: <http://www.sciencedirect.com/science/article/pii/B9780123851208000176> [Accessed 27 Sep. 2016].

Miyazaki, K. and Takenouchi, M., 2002. Creating random mutagenesis libraries using megaprimer PCR of whole plasmid. *BioTechniques*, 33(5), pp.1033–1034, 1036–1038.

Mot, R.D., Nagy, I., Schrijver, A.D., Pattanapipitpaisal, P., Schoofs, G. and Vanderleyden, J., 1997. Structural analysis of the 6 kb cryptic plasmid pFAJ2600 from Rhodococcus erythropolis NI86/21 and construction of Escherichia coli-Rhodococcus shuttle vectors. *Microbiology*, 143(10), pp.3137–3147.

Nguyen, Q.-T., de Gonzalo, G., Binda, C., Rioz-Martínez, A., Mattevi, A. and Fraaije, M.W., 2016. Biocatalytic Properties and Structural Analysis of Eugenol Oxidase from Rhodococcus jostii RHA1: A Versatile Oxidative Biocatalyst. *ChemBioChem*, 17(14), pp.1359–1366.

Otten, L.G., Hollmann, F. and Arends, I.W.C.E., 2010. Enzyme engineering for enantioselectivity: from trial-and-error to rational design? *Trends in Biotechnology*, 28(1), pp.46–54.

Pai, J.C., Entzminger, K.C. and Maynard, J.A., 2012. Restriction enzyme-free construction of random gene mutagenesis libraries in Escherichia coli. *Analytical Biochemistry*, 421(2), pp.640–648.

Pei, X., Wang, J., Guo, W., Miao, J. and Wang, A., 2017. Efficient biodegradation of dihalogenated benzonitrile herbicides by recombinant Escherichia coli harboring nitrile hydratase-amidase pathway. *Biochemical Engineering Journal*, 125, pp.88–96.

Quan, S. and Dabbs, E.R., 1993. Nocardioform Arsenic Resistance Plasmid Characterization and Improved Rhodococcus Cloning Vectors. *Plasmid*, 29(1), pp.74–79.

Reetz, M.T., 2011. Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions. *Angewandte Chemie International Edition*, 50(1), pp.138–174.

Reetz, M.T., 2016. *Directed Evolution of Selective Enzymes: Catalysts for Organic Chemistry and Biotechnology*. John Wiley & Sons.

Reetz, M.T. and Carballeira, J.D., 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature Protocols*, 2(4), pp.891–903.

Reetz, M.T., Kahakeaw, D. and Lohmer, R., 2008. Addressing the Numbers Problem in Directed Evolution. *ChemBioChem*, 9(11), pp.1797–1804.

Reetz, M.T., Prasad, S., Carballeira, J.D., Gumulya, Y. and Bocola, M., 2010. Iterative Saturation Mutagenesis Accelerates Laboratory Evolution of Enzyme Stereoselectivity: Rigorous Comparison with Traditional Methods. *Journal of the American Chemical Society*, 132(26), pp.9144–9152.

Reetz, M.T., Wang, L.-W. and Bocola, M., 2006. Directed Evolution of Enantioselective Enzymes: Iterative Cycles of CASTing for Probing Protein-Sequence Space. *Angewandte Chemie*, 118(8), pp.1258–1263.

Reetz, M.T. and Zheng, H., 2011. Manipulating the Expression Rate and Enantioselectivity of an Epoxide Hydrolase by Using Directed Evolution. *ChemBioChem*, 12(10), pp.1529–1535.

Reymond, J.-L., 2006. *Enzyme Assays: High-throughput Screening, Genetic Selection and Fingerprinting*. John Wiley & Sons.

Rosano, G.L. and Ceccarelli, E.A., 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*, [online] 5. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4029002/> [Accessed 23 Sep. 2017].

Rzeznicka, K., Schätzle, S., Böttcher, D., Klein, J. and Bornscheuer, U.T., 2010. Cloning and functional expression of a nitrile hydratase (NHase) from Rhodococcus equi TG328-2 in Escherichia coli, its purification and biochemical characterisation. *Applied Microbiology and Biotechnology*, 85(5), pp.1417–1425.

Sanchis, J., Fernández, L., Carballeira, J.D., Drone, J., Gumulya, Y., Höbenreich, H., Kahakeaw, D., Kille, S., Lohmer, R., Peyralans, J.J.-P., Podtetenieff, J., Prasad, S., Soni, P., Taglieber, A., Wu, S., Zilly, F.E. and Reetz, M.T., 2008. Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates. *Applied Microbiology and Biotechnology*, 81(2), pp.387–397.

Sandström, A.G., Wikmark, Y., Engström, K., Nyhlén, J. and Bäckvall, J.-E., 2012. Combinatorial reshaping of the Candida antarctica lipase A substrate pocket for enantioselectivity using an extremely condensed library. *Proceedings of the National Academy of Sciences*, 109(1), pp.78–83.

Sankararaman, S., Kolaczkowski, B. and Sjölander, K., 2009. INTREPID: a web server for prediction of functionally important residues by evolutionary analysis. *Nucleic Acids Research*, 37(suppl_2), pp.W390–W395.

Sarkar, G. and Sommer, S.S., 1990. The 'megaprimer' method of site-directed mutagenesis. *BioTechniques*, 8(4), pp.404–407.

Soobben, M., 2020. *Production of Nitrile Hydrolysing Biocatalysts from Rhodococcus rhodochrous*. University of the Witwatersrand.

Stanislauskiene, R., Gasparaviciute, R., Vaitekunas, J., Meskiene, R., Rutkiene, R., Casaite, V. and Meskys, R., 2012. Construction of Escherichia coli- Arthrobacter- Rhodococcus shuttle vectors based on a cryptic plasmid from Arthrobacter rhombi and investigation of their application for functional screening. *FEMS Microbiology Letters*, 327(1), pp.78–86.

Sullivan, B., Walton, A.Z. and Stewart, J.D., 2013. Library construction and evaluation for site saturation mutagenesis. *Enzyme and Microbial Technology*, 53(1), pp.70–77.

Tang, L., Gao, H., Zhu, X., Wang, X., Zhou, M. and Jiang, R., 2012. Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *Biotechniques*, 52(3), pp.149–158.

Tsekoa, T.L., Sayed, M.F., Cameron, R.A., Sewell, B.T. and Cowan, D.A., 2004. Purification, crystallization and preliminary X-ray diffraction analysis of thermostable nitrile hydratase: research letter. *South African journal of science*, 100(9–10), pp.488–490.

Van Wyk, J.C. 2008. *The relationship between structure and thermostability of a nitrile hydratase from Geobacillus pallidus RAPc8*. Available at: <http://etd.uwc.ac.za/xmlui/handle/11394/2442> [Accessed 9 Aug. 2020].

Vandergeize, R. and Dijkhuizen, L., 2004. Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications. *Current Opinion in Microbiology*, 7(3), pp.255–261.

Wahler, D. and Reymond, J.-L., 2001. High-throughput screening for biocatalysts. *Current opinion in biotechnology*, 12(6), pp.535–544.

Yam, K.C., Okamoto, S., Roberts, J.N. and Eltis, L.D., 2010. Adventures in Rhodococcus — from steroids to explosives. *Canadian Journal of Microbiology*, 57(3), pp.155–168.

CHAPTER 4: ORGANIC SYNTHESIS AND BIOCONVERSION OF BETA-AMINO NITRILES

4.1 INTRODUCTION

4.1.1 Amides and their derivatives in medicine and industry

Amides are one of the most common molecules in nature and amide bond formation is one of the most important reactions, as it is the bond which links amino acids in proteins and peptides (Pitzer and Steiner, 2016). The general amide functional group consists of a carbon with three attached electron domains: a double-bonded oxygen, an amine or ammonia group and an R group (where R refers to a hydrogen or another hydrocarbon moiety). Amides may be primary, secondary or tertiary, where a primary amide has the nitrogen bound to two hydrogens and the acyl group of the amide; in a secondary amide, the nitrogen has one hydrogen and one R group in addition to the acyl group; and in a tertiary amide, the nitrogen has two R groups aside from the acyl moiety of the amide (Figure 4.1. Resonance over the N-CO bond contributes to a slightly heightened energetic barrier to the rotation of this bond (Greenberg, 2002).

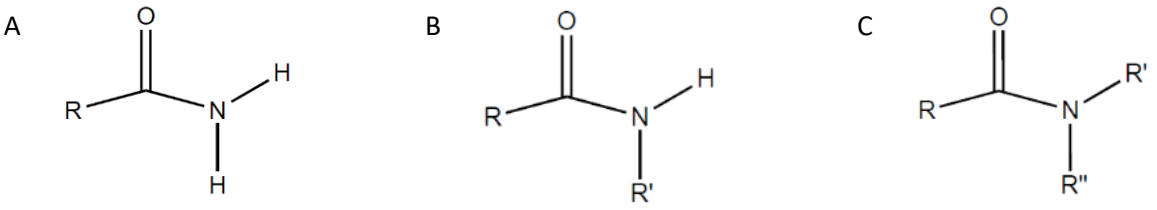


Figure 4.1: General amide structure. A: primary amide; B: secondary amide; C: tertiary amide. Image created using ChemDraw v.12.0.

The amide bond within proteins and peptides occurs on the ribosome. Transporter RNA (tRNA) molecules attach amino acids to the growing peptide chain, using a messenger RNA (mRNA) template. The amide bond (Figure 4.2) is a covalent bond formed via a condensation reaction between the carboxylic acid group of one amino acid and the amino group of another amino acid, with the loss of a water molecule. In this way, a continuous chain of amino acids held together by amide bonds is formed (Petsko and Ringe, 2004).

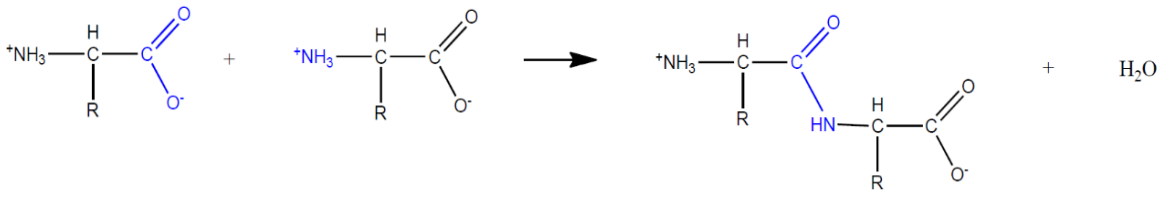


Figure 4.2: The formation of the peptide bond between amino acids is a condensation reaction between the carboxylic acid group (-COO^-) of one amino acid and the amine group (-NH_3^+) of a second amino acid, forming an amide or peptide bond (-CONH-) and releasing a water molecule. Image created using ChemDraw v.12.0.

Amides have uses as potential pharmaceuticals and industrial products, or as precursors of other products. Due to this varied potential, they have been of interest to researchers for at least two decades. Thiollet and Musikas (1989) suggested their use as extractants of actinide ions in the nuclear industry. These findings were later supported by Ansari *et al.* (2005) who identified N,N,N',N'-Tetraoctyl diglycolamide (TODGA) as a potential nuclear waste extractant. By 2000, Troeberg *et al.* had identified certain disubstituted aliphatic amides to have activity against *Trypanosoma* species, which cause human sleeping sickness and other diseases in cattle. Development of antiparisitic drugs targeting *Trypanosoma* are still of interest, and research continues to focus on amide derivatives. In 2009, Konkle *et al.* found that indomethacin-amides were active against *Trypanosoma cruzi* sterol 14α-demethylase and in 2016, Pirttimaa *et al.* identified diterpenoid amides with activity against both *Trypanosoma cruzi* and *Leishmania donovani*. Amides are also used as flavouring agents in the food industry (Tachdjian *et al.*, 2006, 2010).

4.1.2 Current synthesis techniques for production of amides

Amides may be produced using organic or inorganic methods and catalysts, as side products of other processes or as products extracted from plants. In 2005, 65% of 128 potential drug candidates surveyed included an amide bond formation step (Constable *et al.*, 2007). Common coupling reagents for this bond formation included *N*-ethyl-*N'*-(3-dimethylaminopropyl) carbodiimide hydrochloride catalysed by hydroxybenzotriazole; 1-propylphosphonic acid cyclic anhydride; and *N,N'*-carbonyldiimidazole.

Jain *et al.* (2014) produced amides by a reaction of an aldehyde with a nitro-containing compound in acetic acid over a manganese dioxide (MnO_2) catalyst. This was successful with aromatic (aryl) nitro compounds,

but not aliphatic nitro compounds, and poor yields were obtained with iodo- and bromo- substitutions on the benzene ring (Jain *et al.*, 2014). Primary amides can be produced using the method of Oishi, *et al.* (2009), whereby nitriles are synthesised from primary alcohols and ammonia and the nitriles are then hydrated under argon. Up to 70% yields were obtained by this method. A method of producing primary aromatic amides by reacting benzylic alcohols and aqueous ammonia utilising I_2/TBHP catalyst system in acetonitrile was developed by Xie *et al.* (2017).

4.1.3 Greener amide synthesis techniques

As environmentalism has become more mainstream, the chemical industry has begun to look for new means to develop chemicals, pharmaceuticals and other chemical processes which uphold the values of green chemistry (Poliakoff *et al.*, 2002; Heinrichs and Thum, 2005; Anastas and Eghbali, 2009; Choi *et al.*, 2015). In 2005, the ACS GCI Pharmaceutical Roundtable (ACS-GCIPR) was developed to encourage the funding and development of greener pharmaceutical production. One of the key areas identified was the production of amides, with the goal of replacing or decreasing the use of reagents with poor atom economy, a contributor to waste output. Decreasing waste in this area is desirable, as 9.1% of pharmaceutical production identified in the 2005 Roundtable involved amides (Constable *et al.*, 2007). The coupling reactions used to produce amide bonds produce more waste, while the hydroxybenzotriazole catalyst is a potential explosive. It was predicted as a result of the focus on green chemistry that biocatalysts would begin to replace such processes by 2010 (Constable *et al.*, 2007).

Although greener organometallic catalysts have been developed, a disadvantage of reactions involving amides with these catalysts is that the amides still need to be Boc-protected (Andrews *et al.*, 2010). Therefore it is still important to develop reactions involving either milder conditions, or using more specific processes that do not cause the unwanted reaction or alteration of the amide group.

The 2005 Roundtable on green chemistry identified poor substrate scope as a disadvantage of biocatalysts (Constable *et al.*, 2007). Nitrile hydratases however have much potential in this case, as they have a wide substrate scope. Many NHases have been identified in highly varied environments and their activity ranges from aliphatic to aromatic nitrile substrates (Brandao *et al.*, 2003). Therefore, it should be possible to identify or even engineer an NHase biocatalyst with a desired substrate specificity (Yu *et al.*, 2006; Tao and Xu, 2009).

4.1.4 The potential of nitrile hydratase as a biocatalyst to produce enantiopure beta substituted nitriles

β-substituted nitriles are those with substituent groups at the beta carbon, as determined by the position of the nitrile group. The production of both β-substituted hydroxynitriles and aminonitriles has been the focus of recent studies as they form the precursors to pharmaceuticals (Preiml *et al.*, 2003; Wu and Li, 2003) such as β-blockers, 1,3-amino alcohols (Wu and Li, 2003) and the anti-tumour drug paclitaxel while also possessing potential antibiotic and antifungal effects alone (Winkler *et al.*, 2005). In certain bacteria and fungi, there is an alternate pathway besides the ribosomal pathway which is used to produce peptides known as nonribosomal peptides (NRP) on large enzyme complexes called nonribosomal peptide synthetases (NRPSs) (Caboche *et al.*, 2008). These peptides act as antiviral, antibiotic and anti-tumour agents, as well as immunosuppressants and surfactants. They have also been found to act as siderophores (Pitzer and Steiner, 2016), which chelate metal ions and transport them into cells under conditions of low ion availability (Ahmed and Holmström, 2014). NRPs may contain moieties not found in ribosomal proteins, such as glycosylated, methylated or phosphorylated modifications, heterocycles, D-amino acids and, notably for this study, β-amino acids (Aguilar *et al.*, 2007; Pitzer and Steiner, 2016). It has previously been theorised that artificial NRPs can be synthesised as novel pharmaceuticals and therefore it would be useful to synthesise β-substituted amino acids and their precursor amides (Caboche *et al.*, 2008; Chhiba *et al.*, 2012; Sawant *et al.*, 2014.). Nitrile hydratases have been identified as a potential biocatalysts which can be used to produce these compounds (Winkler *et al.*, 2005; Chhiba *et al.*, 2012).

In order to reduce waste and further synthesis steps, it is preferable to be able to produce these amides as pure enantiomers. When developing new pharmaceuticals, often only one enantiomer is active, while in some cases one enantiomer can be toxic (Simon *et al.*, 2013). When conducting in vitro tests and clinical trials, testing a single enantiomer at a time will provide clearer results as to whether a drug is safe and effective. Nitrile hydratases have already been used in partnership with other enzymes to produce enantiopure chemicals. It has previously been found that amidases show greater enantioselectivity than their associated NHases, usually towards the (*S*)-amide (Faber, 1997; Brady *et al.*, 2004), while highly enantioselective NHases are less common. Gilligan *et al.* (1993) were able to produce a pure solution of R-(-)-2-phenylpropionamide by using an enantioselective amidase on a racemic solution of (R,S)-2-phenylpropionamide, following the production of the amide by a nitrile hydratase from *Rhodococcus equi* TG328. The nitrile hydratase from *R. rhodochrous* been used to produce amides which are then

racemically purified by an enantioselective amidase, as in the production of (S)-Pipecolic acid and (R)- and (S)- piperazinic acid from 2-Pyridinamide and pyrazinamide (Schmid *et al.*, 2002). When 3-hydroxy-4-phenoxybutanenitrile was hydrolysed by *Rhodococcus sp*. CGMCC 0497 at 20°C, it was found to be preferentially hydrolysed by a nitrile hydratase/amidase system at this temperature with (R)-amides and (S)-acids produced with a maximum enantiomeric excess of 87% (Wu and Li, 2003). In all of these cases, the NHase itself was not enantioselective and was paired with an amidase that produced racemically pure carboxylic acids. However, this requires both extra time and cost and therefore a process using only the NHase and not the amidase as well is more desirable. Therefore a goal of current research is to develop an NHase that could produce enantiopure amides in a single step.

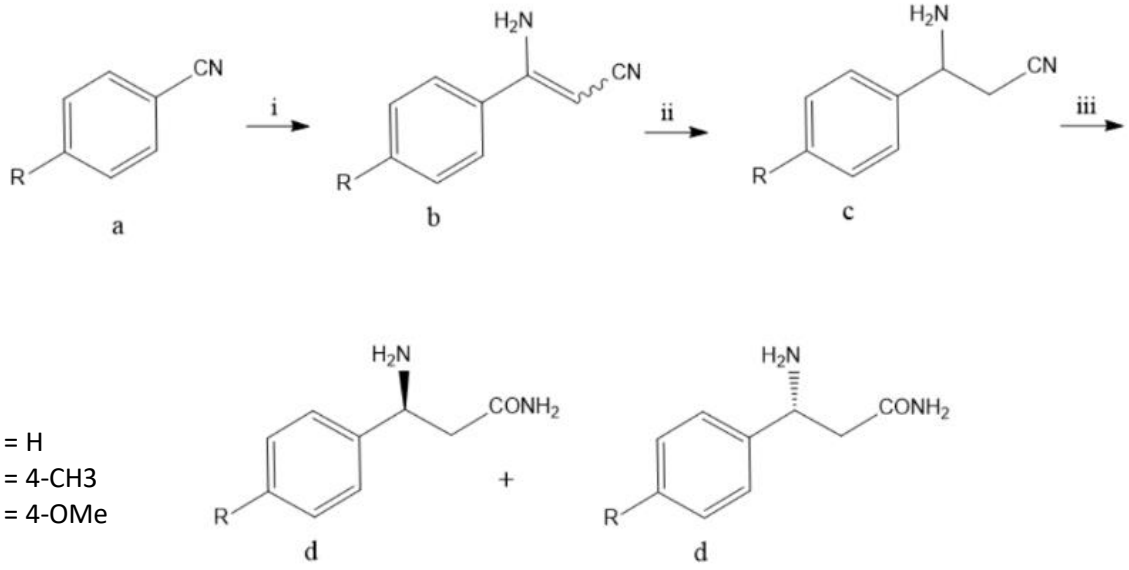
The first Co-type NHase showing stereoselective activity was characterised from *Pseudomonas putida* NRRL-18668 by Payne *et al.* (1997). In a study of various *Rhodococci* in whole cell reactions, Winkler *et al.* (2005) found that *trans*-five-membered aminonitriles were hydrolysed with a greater enantiomeric excess than *trans*-six-membered amino nitrile substrates. The *cis*-five- and six-membered amino nitriles were transformed with a lower enantiomeric excess in all cases, possibly due to steric hindrance caused by the configuration. Similarly, the NHase from *R. equi* TG328-2 showed some enantioselectivity towards aromatic substrates but not aliphatic ones, suggesting that the way in which bulkier substrates fit into the substrate channel plays a role in forcing enantioselective hydrolysis (Rzeznicka *et al.*, 2010). Chhiba *et al.* (2012) studied the enantioselective hydrolysis of β-substituted nitriles by whole cells of *R. rhodochrous* ATCC BAA-870 and found that β-amino nitriles with substituted aromatic rings as well as *para*-substituted rings led to an increased enantiomeric excess of the (*S*)-enantiomer of the amide. A very high enantiomeric excess is necessary for pharmaceutical syntheses, and a highly enantioselective nitrile hydratase is not naturally available. Furthermore, most studies reporting enantioselectivity achieved this using whole cells, not an isolated NHase. Therefore developing an enantioselective NHase is of interest. The Fe-type NHases do not usually show enantioselectivity (van Pelt *et al.*, 2011), thus the Co-type NHases are of more interest to the current study.

4.2 MATERIALS AND METHODS

4.2.1 General

Unless otherwise specified, all chemicals and reagents were obtained from Sigma-Aldrich (now Merck; Modderfontein, South Africa).

4.2.2 Organic synthesis of beta-substituted nitriles



1 R = H
2 R = 4-CH3
3 R = 4-OMe

Reagents (i) CH_3CN, KOtBu, 24 h. (ii) $NaBH_3CN$, HCl, EtOH, 2 h. iii. *R. rhodochrous sp.*, MeOH in 100mM Tris buffer, pH 9.0 at 4°C.

Figure 4.3: Reaction scheme showing the synthesis of 3-amino-3-phenyl acrylonitriles (**b**) from substituted benzonitriles (**a**) followed by the synthesis of racemic 3-amino-3-phenyl propane nitriles (**c**) and finally the enzymatic transformation of compounds **c** to their β-amino amide products (**d**). Image created using ChemDraw v.12.0.

The method of Chhiba *et al.*, (2012) was used to synthesis beta-substituted amino nitriles as substrates for enzyme catalysis. The first step involved the synthesis of 3-amino-3-acrylonitriles from substituted benzonitriles. The substituted benzonitrile starting material (minimum 4g) was placed in a round bottom flask and dissolved in 150 mL of dry toluene. Acetonitrile was added in the ratio of 2.01 mmol acetonitrile to 1.00 mmol starting material. Potassium tert-butoxide was added in small amounts (2.30 mmol to 1.00 mmol starting material) with continuous stirring and the reaction was left to stir overnight at room temperature (approximately 20°C). Equal volumes of water and diethyl ether (150 mL each) were added to the reaction mixture in a separating funnel and the aqueous layer was extracted and the organic layer isolated. A further 150 mL of diethyl ether was added to the aqueous layer and a second extraction was carried out. The organic layers were combined, washed with brine and dried over $MgSO_4$. The organic layer was filtered via gravity flow through a sintered funnel and collected in a new round bottom flask. The solvent was removed by rotary evaporation on a Büchi RII apparatus (Büchi Labortechnik AG, Switzerland) and then fully dried *in vacuo* to yield a brown oil. Approximately 20 mL of diethyl ether were added and hexane was added in portions until the solution became permanently cloudy. This mixture was left at 4°C overnight to allow crystal formation. White crystals of acrylonitrile product were washed with ice cold hexane and collected via suction filtration.

To synthesise the 3-amino-3-phenylpropanenitrile derivatives (beta-substituted amino amides), the acrylonitrile starting material (4 g) was dissolved in 50 mL absolute ethanol in a round bottom flask and sodium cyanoborohydride was added in portions with stirring (1.10 mmol to 1.00 mmol starting material). Bromocrescol green (3 drops of a 0.5 % solution in ethanol) was added as a pH indicator. Concentrated HCl (32%) was then added dropwise until the solution turned an opaque lemon yellow. The reaction was allowed to continue with stirring at room temperature overnight (~16 hrs). After overnight stirring, the reaction was stopped by adding water (150 mL), followed by aqueous ammonia (25%) in small amounts until pH 11 was reached, as determined using universal pH paper. The aqueous layer was extracted with three 100 mL volumes of diethyl ether, washed with brine and dried with $MgSO_4$. The organic phase was gravity filtered through a sintered funnel to remove the $MgSO_4$ and the solvent removed in *vacuo*, yielding a brown oil. Uncorrected melting points were determined using a Stuart SMP10 melting point apparatus.

4.2.3 Column chromatography

Macherey-Nagel silica gel 60 (particle size 0.063 mm to 0.200 mm) was used to purify the amino amide and gravity flow column chromatography was performed on the brown oil. The column was packed with a slurry of silica gel and 50:50 hexane to ethyl acetate. The sample was loaded on the column after dissolving in a minimal amount of 80:20 hexane to ethyl acetate. Elution was carried out using a gradient mobile phase, initially using 80:20 hexane to ethyl acetate, then adjusted to 50:50 hexane to ethyl acetate and finally adjusted to 80:20 hexane to ethyl acetate to elute the amino amide. All solvents were distilled prior to use to remove impurities.

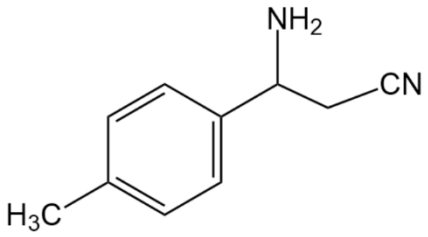
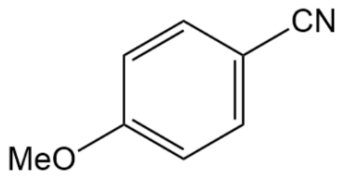
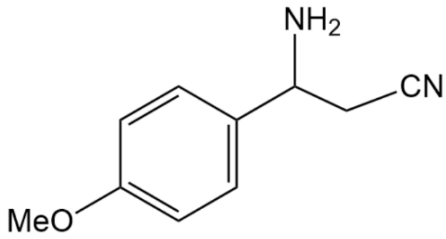
4.2.4 Biocatalytic conversion of beta-substituted nitriles using crude protein extract

Following recombinant NHase expression in both the BL21(DE3) two-construct system and the pJOE system (provided by Prof Trevor Sewell and Dr Brandon Weber of the Aaron Klug Centre for Imaging and Analysis, University of Cape Town), crude cell extract containing protein was used for bioconversion of nitrile substrates. Based on the protein concentration of the cell-free extract, 1 to 5 mg of protein was added to a 1.5 mL Eppendorf tube with 5 mg of substrate dissolved in 10 to 100 mL of methanol. The reaction volume was made up to between 1 mL and 1.5 mL with 100 mM Tris-HCl buffer pH 9 (12.114 g Tris-HCl in 1 L dH_2O, pH adjusted with 1 M HCl, autoclaved at 121°C for 20 min). The reaction was then allowed to progress at 4°C with shaking. Thin layer chromatography was done at 30 min to determine the reaction progress, with all reactions reaching completion after a maximum of 30 min. The reaction was then centrifuged at 10 000 rpm for 5 min.

4.2.4.1 Substrates

Multiple substrates were used to identify and quantify both enzyme activity and enantioselectivity (Table 4.1). The substituted benzonitrile **2a** was used to determine if activity was present. Following this, racemic β-aminonitriles were then used to test the enantioselectivity of the recombinant NHases.

Table 4.1: Nitrile substrates used in this study

Code	Substrate Name	Structure
2c	3-Amino-3-*p*-tolylpropanenitrile	
3a	4-Methoxybenzonitrile	
3c	3-Amino-3-(4-methoxyphenyl) propanenitrile	

4.2.4.2 Thin-layer chromatography

Reaction progress was monitored by thin layer chromatography (TLC) on aluminium-backed silica gel 60 F_{254} plates (Merck, Modderfontein, South Africa). The developing solvent used was a mixture of hexane and ethyl acetate, ranging from 70:30 hexane to ethyl acetate to 100% ethyl acetate based on the polarity of the compounds. Plates were visualised under UV light at λ = 254 nm.

4.2.4.3 Extraction

Once the enzyme reaction was complete, as determined by thin layer chromatography (between 30 min and 24 hrs after initiation of the reaction), the reactions were centrifuged at 13000 rpm for five min to remove protein and other debris. The supernatant was transferred to a 50 mL separating funnel. The aqueous layer was extracted using 2 mL of ethyl acetate, the organic layer was washed with brine and the solvent was removed *in vacuo*. The dried product was then analysed using NMR to confirm its structure and purity (data unavailable due to national lockdown restrictions).

A modified extraction procedure was used prior to chiral HPLC analysis. The enzyme reaction mixture was transferred to a 15 mL Falcon tube. Sodium hydroxide solution (pH 10) was added to the reaction mixture to approximately double the volume. The increased pH caused the amino ($-NH_2$) group of the amino amide product to become unprotonated, from $-NH_3^+$ to $-NH_2$, to make the product less water soluble during separation. A 2 mL – 4 mL volume of ethyl acetate was added and the tube shaken vigorously for 10 sec. The organic and aqueous layers were allowed to separate and the organic layer was removed via pipetting into a fresh Eppendorf tube. A second volume of ethyl acetate was added to the aqueous layer in the Falcon tube, and the tube was shaken and the organic layer removed. Samples of both the organic and aqueous phases were separated using TLC to confirm the presence of the amino amide in the organic layer. A small spatula tip of $MgSO_4$ was added to the organic layer and the tube was shaken and allowed to settle. The organic phase was then filtered through cotton wool into a fresh Eppendorf tube and dried in an Eppendorf Concentrator Plus vacuum concentrator for 4 x 15 min at 60°C.

4.2.4.4 ^{1}H and ^{13}C nuclear magnetic resonance spectroscopy

A minimum of 5 mg of sample was dissolved in deuterated chloroform ($CDCl_3$), deuterated methanol (CD_3OD) or dimethyl sulfoxide (DMSO) and ^{1}H and ^{13}C nuclear magnetic resonance was recorded at room temperature on a Bruker 300 MHz NMR spectrometer using tetramethylsilane as an internal standard. Data processing was done using MestReNova v.9.0.1 (Mestrelab Research, S.L., Santiago de Compostela, Spain). The number of scans was set to 96 for ^{1}H NMR and 1024 or 2048 for ^{13}C NMR.

4.2.4.5 Chiral high-performance liquid chromatography of amides

The method of (Chhiba *et al.*, 2012) for separating the chiral amide products was used to determine the relative amounts of enantiomers and to potentially determine enantiomeric excess (ee) and enantiomeric ratio (E). Amino amides were extracted according to the procedure in Section 4.2.4.3 and the dry product was dissolved in aqueous perchloric acid ($HClO_4$, 16.3 g/L, pH 1.98).

The liquid chromatography system comprised an S200 iso-pump, an S200 UV/Vis detector and an S225 autosampler from Perkin Elmer (Massachusetts, USA). The software used for data acquisition was TotalChrom version 6.3 (Perkin Elmer, Massachusetts, USA) and data analysis was done with OpenChrom Analytics Edition 1.4.0 (Wenig and Odermatt, 2010). A Crownpak CR (+) 150 mm × 4 mm column was used, provided by Dr Varsha Chhiba-Govindjee of the Council for Scientific and Industrial Research (CSIR, Pretoria, South Africa) and was stored below 10°C when not in use. The eluent used was aqueous perchloric acid (16.3 g/L, pH 1.98) and the column was flushed for a minimum of one hour with a flow rate of 0.2 mL/min with solvent before loading any samples. Samples were diluted with an appropriate volume of perchloric acid to ensure that the amount of sample was not more than 10^{-6} mol per sample. Initial method development was done using samples of **2d** and therefore the absorbance was set to 210 nm (Chhiba *et al.*, 2012). Initially a 10 µL injection volume was used at a flow rate of 0.2 mL/min with an experimental run time of 80 min. The injection volume was then increased to 25 µL and then 50 µL with a maximum of 120 min run time. The flow rate was tested by decreasing the rate to 0.15 mL/min and increasing it to 0.25 mL/min.

4.3 RESULTS AND DISCUSSION

Two chiral substrates, 3-amino-3-*p*-tolylpropanenitrile and 3-Amino-3-(4-methoxyphenyl)propanenitrile were synthesised and used to test the activity of the recombinantly expressed nitrile hydratases from both the two-construct expression system as well as the cells provided by Prof Trevor Sewell and Dr Brandon Weber. A simpler non-chiral substrate, 4-methoxybenzonitrile (**3a**), used as the precursor to 3-amino-3-(4-methoxyphenyl) propanenitrile (**3c**) synthesis, was used to confirm enzyme activity occurred. The two-construct system originally developed by Frederick (2013) was used in an effort to optimise a chiral HPLC method to determine the enantiomeric excess and enantioselectivity of rationally designed NHase variants, however technical problems prevented the optimisation of this method.

4.3.1 Organic synthesis of beta-substituted nitriles

The acrylonitriles **2b** and **3b** were successfully synthesised. As the acrylonitrile compounds were intermediates in the synthesis and the synthesis method has already been optimised by Chhiba *et al.* (2012), only proton NMR was done to check the structures of the acrylonitriles. The enzyme substrates **2c** and **3c** were then successfully synthesised from their precursors, **2b** and **3b** and both proton and carbon-13 NMR was done to confirm the structure of each final amino nitrile. The yields at each step were moderate, between 30% and 60%, despite varying reaction and purification conditions multiple times. The variations attempted during reactions included changing the stirring time, increasing the stirring temperature to 30°C and using varying amounts of reagents and solvents. However, no significant improvement was observed and thus these are not reported in detail.

Proton and carbon NMR indicated that the substrates were successfully synthesised, although with some impurities in the case of **2c** and **3c**. Representative NMR spectra are shown in the Appendix, Figures A1-A4. Figure 4.4 is a selection of the fractions taken during gravity-flow column chromatography purification of the substrate **3c**, showing the co-elution of the substrate with an impurity. TLC also showed the presence of this impurity as a spot (R_f = 0.43, 100% ethyl acetate) above the substrate spot (R_f = 0.22, 100% ethyl acetate) and it was not possible to remove using any combination of solvents, column length and diameter and silica type during the chromatography step. As **3c** is in the form of an oil, no melting point analysis was done in this study, as well as in previous work (Chhiba *et al.*, 2012).

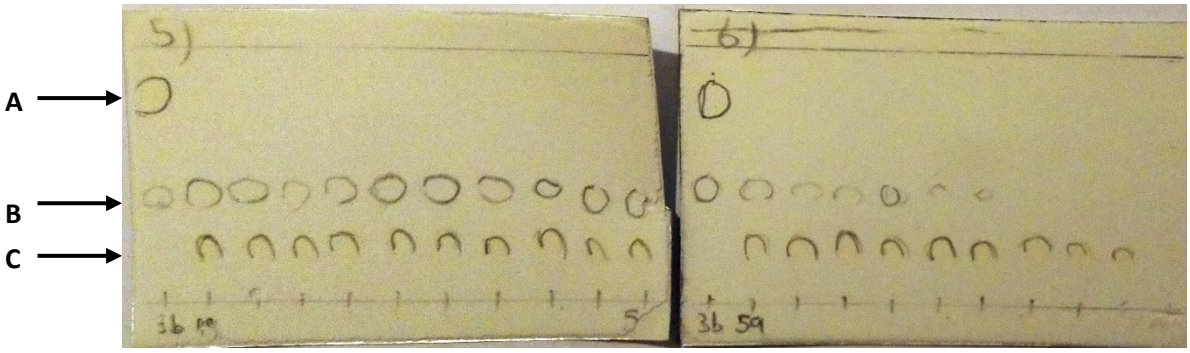
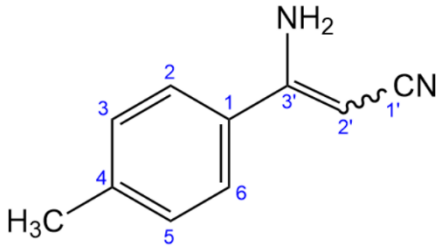


Figure 4.4: Thin-layer chromatography plates showing samples taken from fractions during workup using gravity-flow column chromatography purification of substrate 3c, showing elution of an impurity with the product. Spot A: **3b** 3-Amino-3-(4-methoxyphenyl) acrylonitrile, precursor to **3c** (R_f = 0.81); spot B: unknown impurity (R_f = 0.43); spot C: substrate **3c** (3-Amino-3-(4-methoxyphenyl) propanenitrile) (R_f = 0.21). The polarity of the solvent was increased during elution from 50:50 ethyl acetate to hexane to 100% ethyl acetate. TLC plates were developed using 100% ethyl acetate. Yellowing is due to the presence of the amine groups in **3b** and **3c**.

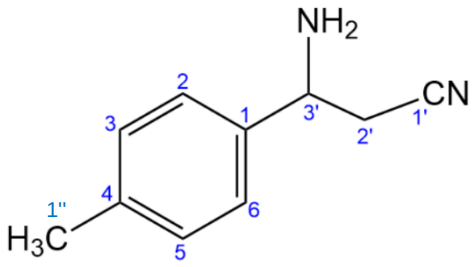
4.3.1.1 Amino-3-*p*-tolyl acrylonitrile (2b)



3-Amino-3-*p*-tolyl benzonitrile (5g). Stirring time = 16 hrs. Purification by washing with ice cold hexane (3 x 20 mL) to yield a pale yellow powder (2.8634 g; 37 % yield). **R_f** (70:30 hexane to EtOAc) = 0.43. **^{1}H NMR (300 MHz, chloroform-*d*)**: δ 7.39 (d, *J* = 8.2 Hz, 2H), 7.23 (m, 2H), 4.89 (br s, 2H), 4.24 (s, 1H), 1.56 (s, 4H). **^{13}C NMR as predicted by ChemDraw v.12.0 (300 MHz, chloroform-*d*):** δ 156.3 (C3'), 137.6 (C4), 131.2 (C1), 129.0 (C2, C6), 128.9 (C3, C5), 118.6 (C1'), 58.3 (C2'), 21.3 (C1'').

Melting point (MP) analysis was not carried out, however literature reports it to be 117-110°C (Coward and Bruice, 1969) and 103.5°C (Chhiba *et al.*, 2012). As this is an intermediate product, only ^{1}H NMR was used to confirm the structure before undertaking the next synthesis step.

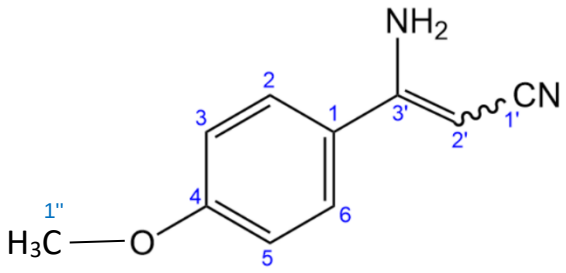
4.3.1.2 3-Amino-3-*p*-tolyl propanenitrile (2c)



Stirring time = 16 hrs. Purification by column chromatography (elution 70:30 EtOAc to hexane) to yield a yellow-brown oil which then solidified to a yellow solid (0.33 g; 15 %). **R_f** (60:40 EtOAc to hexane) = 0.39. **MP** = 98 – 100°C. **^{1}H NMR (300 MHz, chloroform-*d*)**: δ 7.28-7.26 (m, 2H), 7.25-7.17 (m, 2H), 4.1 (dd, J = 7.2, 5.7 Hz, 1H), 2.67-2.64 (m, 1H), 2.35 (s, 3H), 1.69 (br s, 2H). **^{13}C NMR (300 MHz, chloroform-*d*):** δ 129.58 (C3, C5), 125.80 (C2, C6), 52.47 (C3'), 28.58 (C2'), 21.12 (C1'').

Due to the low yield, insufficient product was available for ^{13}C NMR and therefore the peaks were difficult to distinguish from the baseline. The peaks that did not show clearly on the spectrum were δ 139.58 (C1) and 137.90 (C4) as previously determined by Chhiba *et al.* (2012). Literature melting point values have not been reported, as the product is an oil. It is possible that the solidification observed in this study was due to impurities in the product.

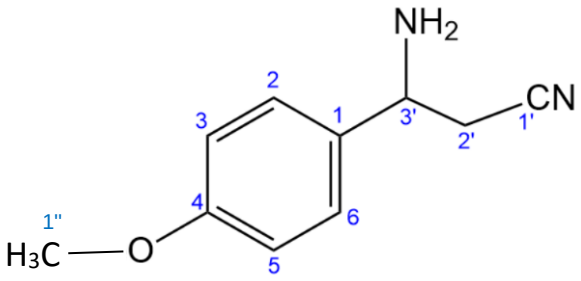
4.3.1.3 3-Amino-3-(4-methoxyphenyl) acrylonitrile (3b)



Stirring time = 16 hrs. Purification by washing with ice cold hexane to yield a pale yellow powder (2.11 g; 32 %). **R_f** (70:30 hexane to EtOAc) = 0.51. **MP** = 115-119°C. **^{1}H NMR (300 MHz, chloroform-*d*)**: δ 7.32-7.29 (m, 2H), 6.91-6.88 (m, 2H), 4.67 (s, 1H), 3.83 (s, 3H), 2.02 (s, 2H) **^{13}C NMR (300 MHz, chloroform-*d*):** δ 161.87 (C4), 161.24 (C3'), 134.07 (C2, C6), 127.54 (C1), 119.31 (C3, C5), 114.85 (C1'), 62.66 (C2'), 55.64 (C1'').

The MP was within the range previously reported by Coward and Bruice (1969) and Chhiba *et al.* (2012), who reported the MP as 117-118°C and 119.4°C respectively.

4.3.1.4 3-Amino-3-(4-methoxyphenyl) propanenitrile (3c)



Stirring time = 16 hrs. Purification by column chromatography (elution 70:30 EtOAc to hexane) to yield a brown oil (0.4289 g; 29 %). **R_f** (70 ETOAc: 30 hexane) = 0.29. **^{1}H NMR (300 MHz, chloroform-*d*)**: δ 7.30-7.25 (m, 2H), 6.91-6.85 (m, 2H), 4.28-4.24 (m, 1H), 3.77 (s, 3H), 2.61 (dd, *J* = 6.4, 3.4 Hz, 1H), 1.69 (s, 2H). **^{13}C NMR (300 MHz, chloroform-*d*):** δ 159.39 (C4), 134.57 (C1), 127.08 (C2, C6), 118.07 (C1'), 114.19 (C3, C5), 55.30 (C1''), 52.09 (C3'), 28.56 (C2').

No melting point analysis was carried out as the product was an oil (Chhiba *et al.*, 2012).

4.3.2 Biocatalytic conversion of beta-substituted nitriles using crude protein extract

Enzyme reactions using both NHase expression systems were tested. Both BL21(DE3) cells containing the two construct system and the pJOE cells were cultured and harvested, following which the cells were suspended in Tris buffer (pH 9). A volume of crude protein extract containing 5 mg of protein (approximately 300 µL of extract) was combined with 5mg of substrate dissolved in 10 to 100 µL of methanol. Further Tris buffer was added to make an approximately 1 to 1.5 mL reaction volume. The reaction products were monitored using TLC (Figure 4.5). Two samples of substrate **3c** were used, one with an impurity that it was not possible to remove during column purification of the substrate. Despite the presence of this impurity, all biocatalytic enzyme reactions using the synthesised substrate **3c** were successful, with all reactions going to completion within 30 min. There was full conversion of **3c** to its product amide **3d** while the unidentified impurity showed no change, suggesting that it did not interact with the enzyme. The structures of the enzyme reaction products from the substrates used in this study have previously been characterised using NMR (Chhiba *et al.*, 2012).

The two-construct system in BL21(DE3) was then used in an attempt to develop a method to quantify enantioselectivity. The NHase expressed by the two-construct system was subject to reactions with three substrates, one non-chiral (4-methoxybenzonitrile, **3a**) and two chiral substrates (**2c** and **3c**). In order to test for the presence and potential activity of native *E. coli* NHases, BL21(DE3) cells not containing the two-construct system were also cultured and provided with the same inducing conditions and then used in identical enzyme reactions. All three substrates were converted to amide products within 30 min by the NHase from the two-construct system, while the induced untransformed BL21(DE3) cells showed no product formation, indicating that there was no other NHase expressed by these cells.

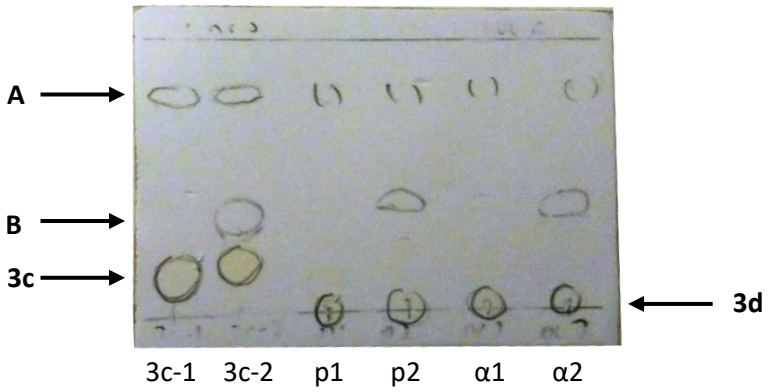
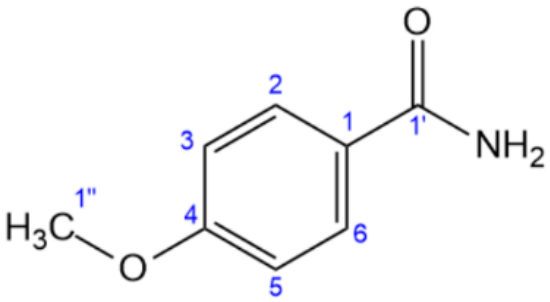


Figure 4.5: Enzyme conversion reactions after 30 min using cells containing either the BL21(DE3) two-construct system or the pJOE system. 3c-1: cleaner sample of substrate **3c**; 3c-2: impure sample of **3c**; p1 and p2: enzyme reaction supernatant from cells containing pJOE expression system; α1 and α2: enzyme reaction supernatant from cells containing two-construct system in BL21(DE3). A: unreacted **3b** remaining from the synthesis of **3c**; B: additional impurity found in substrate 3c-2 only; **3c**: 3-Amino-3-(4-methoxyphenyl) propanenitrile; **3d**: 3-Amino-3-(4-methoxyphenyl) propanamide. TLC plates were developed using 100% ethyl acetate. Yellowing is due to amine groups in **3b** and **3c**. A (R_f = 0.81); B (R_f = 0.0.35); **3c** (R_f = 0.14); **3d** (R_f = 0.00).

4.3.2.1 4-methoxybenzamide

Full conversion of **3a** to 4-methoxybenzamide occurred within 30 min at 4°C using the NHase expressed by the two-construct system while BL21(DE3) control cells showed no conversion of **3a** (Figure 4.6).



Reaction time: 30 min. Vacuum concentration for 4 x 15 min at 60°C yielded a white crystalline layer. $\mathbf{R_f}$ (50:50 ethyl acetate to hexane) = 0.14

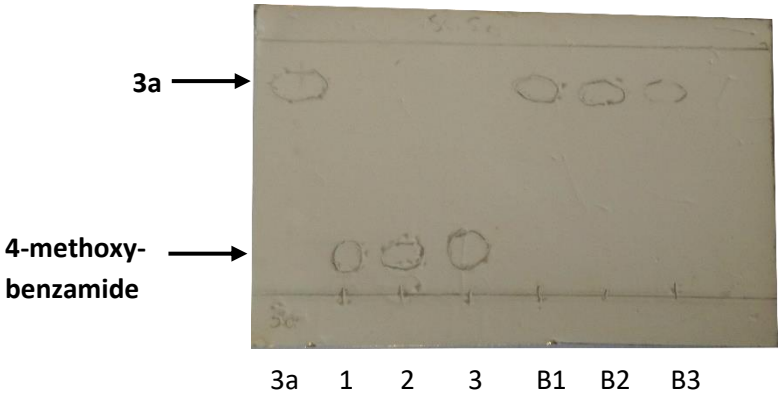
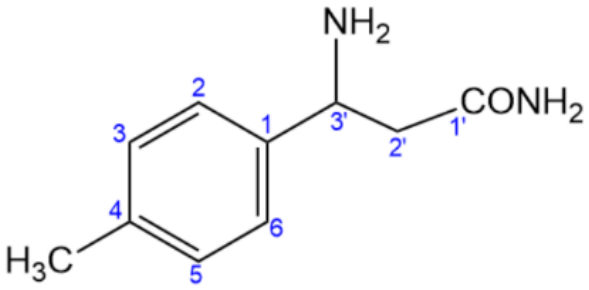


Figure 4.6: Conversion of substrate 3a to 4-methoxybenzamide using the two-construct expression system for NHase expression. No yellowing of spots occurred as this substrate and product do not contain amine groups. 1, 2, 3: supernatant containing expressed NHase from three separate samples of the two construct system. B1, B2, B3: supernatant from three separate samples of BL21(DE3) control cells induced but not containing the two-construct system. The control cells showed no conversion of product indicating there were no nitrile-hydrolysing enzymes present. TLC plate developed using 50:50 ethyl acetate to hexane. **3a** (R_f = 0.81); 4-methoxy benzamide (R_f = 0.14).

4.3.2.2 3-Amino-3-*p*-tolyl propanamide (2d)

Full conversion of **2c** to **2d** occurred within 30 min at 4°C using the NHase expressed by the two-construct system while BL21(DE3) control cells showed no conversion of **2c** (Figure 4.7)



Reaction time: 30 min. Vacuum concentration for 4 x 15 min at 60°C yielded a white crystalline layer. R_f (50:50 ethyl acetate to hexane) = 0.026

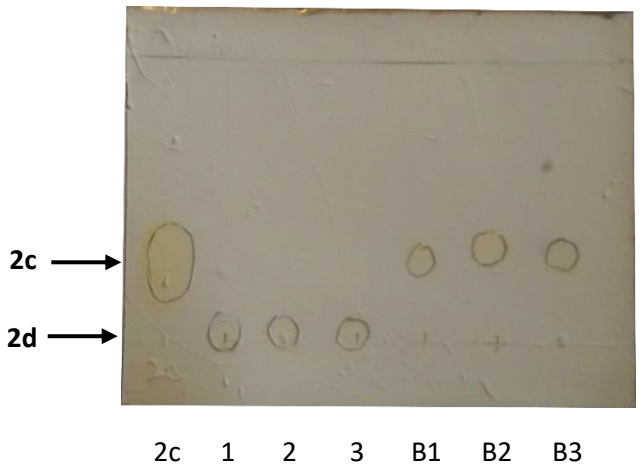
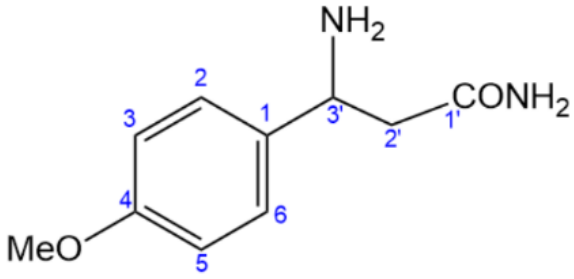


Figure 4.7: Conversion of substrate 2c to 2d using the two-construct expression system for NHase expression. Yellowing of spots occurred due to the presence of amine groups. 1, 2, 3: supernatant containing expressed NHase from three separate samples of the two construct system. B1, B2, B3: supernatant from three separate samples of BL21(DE3) control cells induced but not containing the two-construct system. The control cells showed no conversion of product indicating there were no nitrile-hydrolysing enzymes present. TLC plate developed using 50:50 ethyl acetate to hexane. **2c** (R_f = 0.29); **2d** (R_f = 0.029).

4.3.2.3 3-Amino-3-(4-methoxyphenyl) propanamide (3d)

Full conversion of **3c** to **3d** occurred within 30 min at 4°C using the NHase expressed by the two-construct system while BL21(DE3) control cells showed no conversion of **3c** (Figure 4.8)



Reaction time: 30 min. Vacuum concentration for 4 x 15 min at 60°C yielded a white crystalline layer. $\mathbf{R_f}$ (50:50 ethyl acetate to hexane) = 0.00

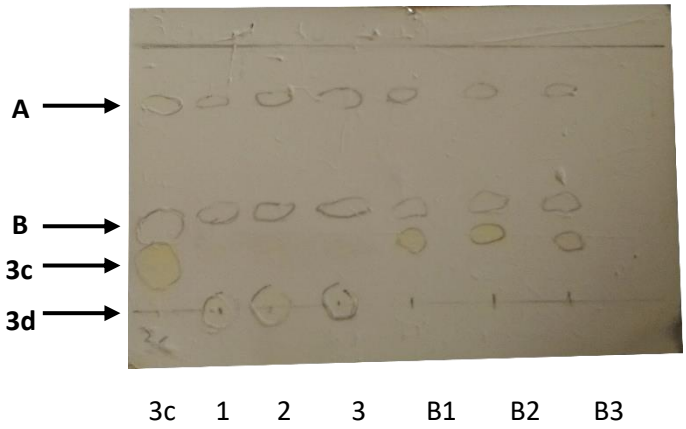


Figure 4.8: Conversion of substrate 3c to 3d using the two-construct expression system for NHase expression. Yellowing of spots occurred due to the presence of amine groups. A: unreacted **3b** from the synthesis of **3c**; B: unidentified impurity from the synthesis of **3c**; 1, 2, 3: supernatant containing expressed NHase from three separate samples of the two construct system; B1, B2, B3: supernatant from three separate samples of BL21(DE3) control cells induced but not containing the two-construct system. The control cells showed no conversion of product indicating there were no nitrile-hydrolysing enzymes present. TLC plate developed using 50:50 ethyl acetate to hexane. A (R_f = 0.77); B (R_f = 0.31); **3c** (R_f = 0.19); **3d** (R_f = 0.00).

4.3.3 Chiral high performance liquid chromatographic analysis of enzyme reaction products

The enzyme product **2d** was expected to elute at 24.5 min and 30.6 min with an absorption maximum at 210 nm and **3d** was expected to elute at 20.0 min and 23.4 min with a max absorption peak at 225 nm. Initial chiral high performance liquid chromatography (HPLC) tests were done using samples of **2d**, thus the detector was set to 210 nm. However, all results were inconclusive despite alterations made to the method, including increasing the sample injection volume, increasing the experiment time and increasing sample concentration. All blanks showed three large combined peaks eluting within the first ten min (Figure 4.9).

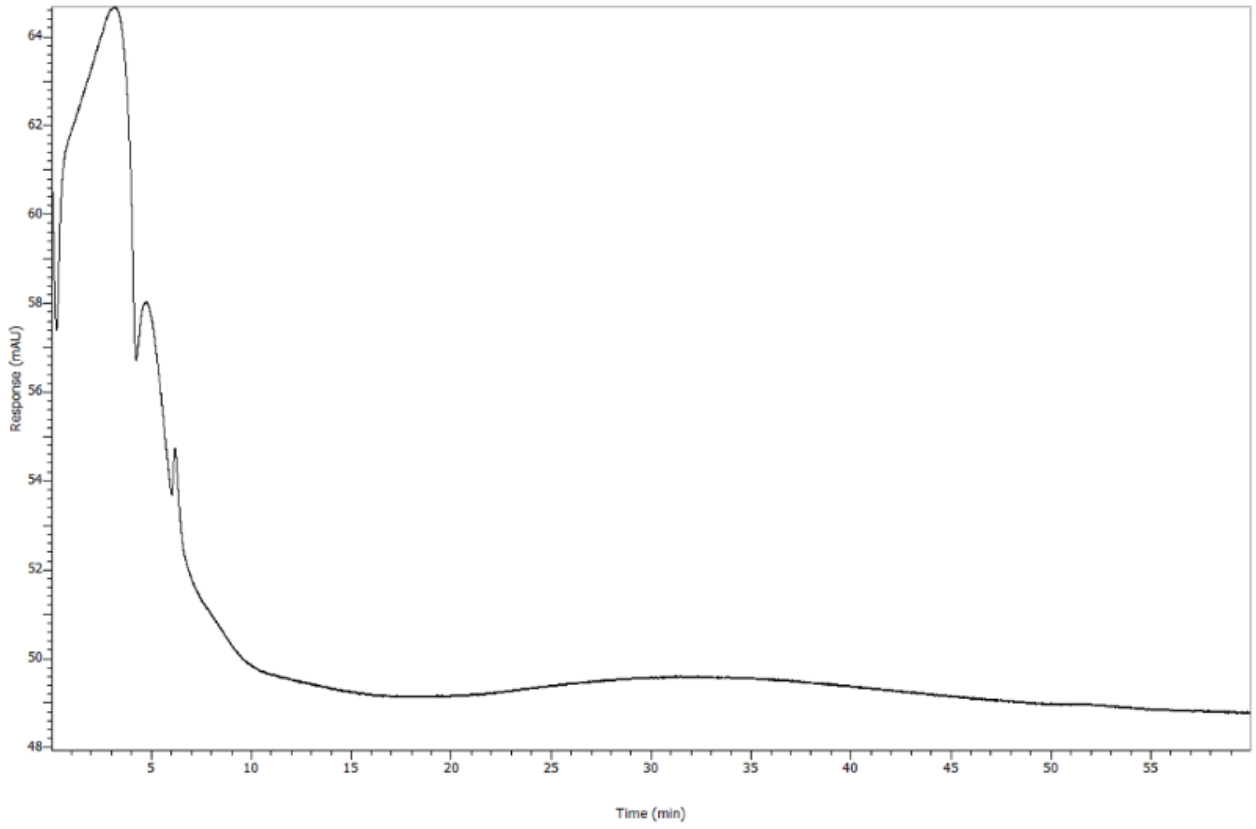


Figure 4.9: Raw HPLC Chromatogram of the blank sample used to calibrate the HPLC system before reading samples of chiral NHase enzyme products. Three large unresolved peaks are visible eluting within 10 min. A 25 µL injection volume was used with a 0.200 mL/min flow rate. Eluent: aqueous perchloric acid (16.3 g/L, pH 1.98). Results visualised using TotalChrom v. 6.3.0 (Perkin Elmer, Massachusetts, USA).

All **2d** samples showed a single large peak also eluting within 10 min, approximately at 6.5 min. Some samples had a second, smaller peak which eluted before the large peak, at approximately 4 to 5 min (Figure 4.10). It is possible that these were the enantiomers of the product, but it was not possible to confirm this as the peaks did not elute at the expected times. As previously shown in Figures 4.6 and 4.7, TLC results seemed to show complete conversion of chiral nitrile to amide, with no remaining nitrile evident on the TLC plate. If this was the case, then the HPLC peaks should have been of equal size, as there would have been equal amounts of each amide enantiomer. The unequal sized HPLC peaks, however, would suggest that the enzyme had preferentially hydrolysed one nitrile enantiomer, and thus there should have been some remaining nitrile in the reaction. Unfortunately none of this could be confirmed due to the unreliable HPLC results.

In the case of all blanks and samples, there were no further eluted peaks after 15 min total run time. This indicated that all substances were eluting from the column very quickly, and therefore not interacting with the column as they should. This suggested that the column's stationary phase had possibly degraded. To test this, a standard solution of **2d** was prepared using lyophilised NHase from cells of *R. rhodochrous* ATCC BAA-870 (prepared by Dr Varsha Chhiba-Govindjee, CSIR). The reaction was allowed to proceed overnight and then the product was prepared as described in Section 4.2.4.3. A TLC separation of this product as well as some of the HPLC samples showed a visible amide spot on the baseline in both the standard and sample 17. Sample 16 showed no spot, although this may have been due to a low concentration of product (Figure 4.11). When the standard was run through the same HPLC column using the same method, it also eluted within 10 min and not at the expected time. This confirmed that the column was most likely degraded. Degradation could have occurred due to improper storage temperature prior to receiving the column. It is recommended to flush the column with distilled water after use as the acidic mobile phase can damage the column, which may not have been done previously (Daicel Corporation, Japan).

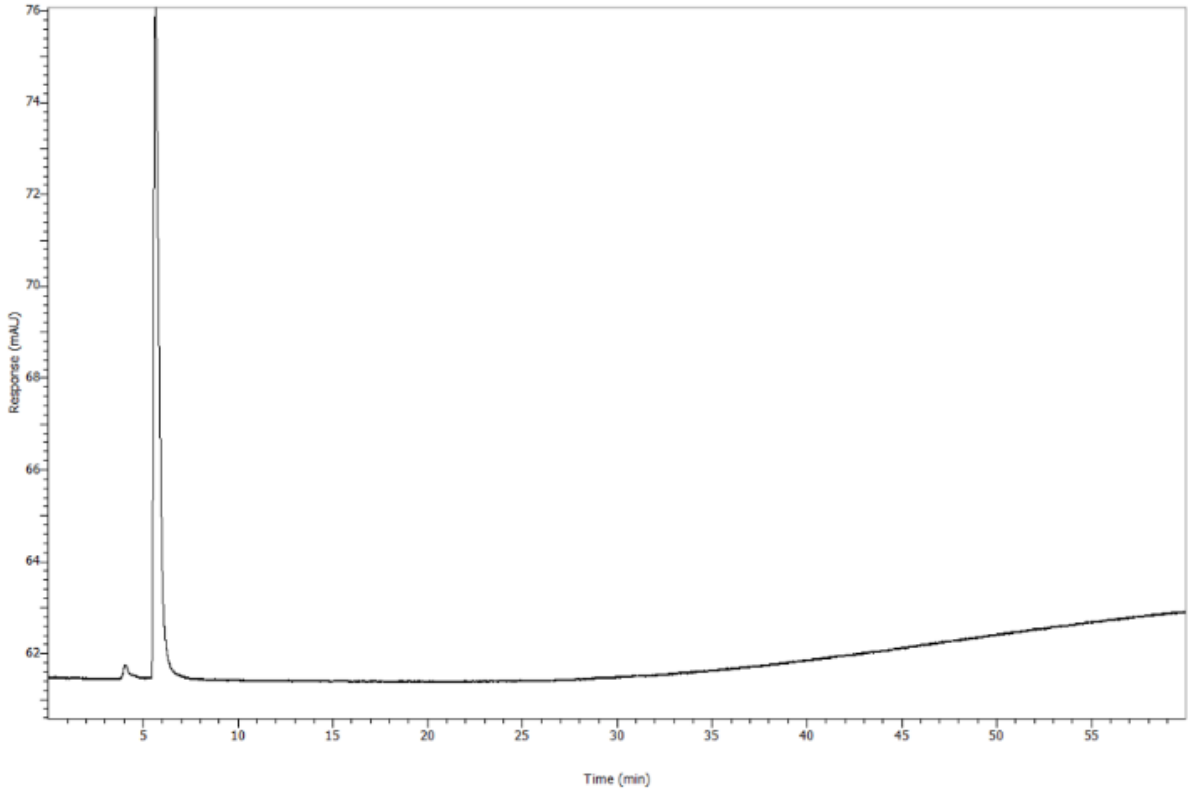


Figure 4.10: Raw HPLC Chromatogram of NHase product 2d (3-Amino-3-*p*-tolyl propanamide). All substances eluted within 10 min, with a large peak at approximately 6 min and a small peak at approximately 4 min, both of which were not the expected times for the enantiomers of **2d**. A 25 µL injection volume was used with a 0.200 mL/min flow rate. Eluent: aqueous perchloric acid (16.3 g/L, pH 1.98). Results visualised using TotalChrom v. 6.3.0 (Perkin Elmer, Massachusetts, USA).

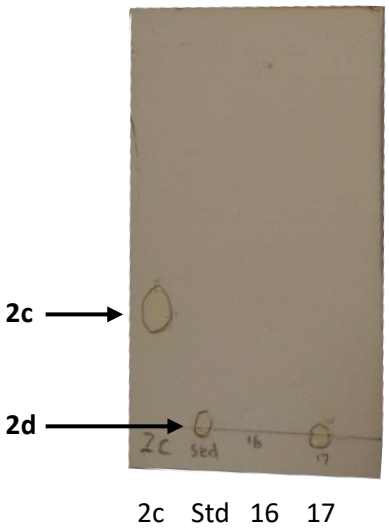


Figure 4.11: Standard amide 2d was produced using lyophilised NHase from cells of *R. rhodochrous* ATCC BAA-870 and compared to the amide produced by the two-construct expression system in BL21(DE3). Both reactions appear to have gone to completion, with all nitrile converted to amide. This was used to confirm the presence of the amide in the HPLC samples. Std: standard reaction; 16, 17: HPLC samples. TLC plate developed using 80:20 ethyl acetate to hexane.

4.3.4 Development of a high-throughput assay to monitor enantioselectivity

One of the goals of this study was to develop a high-throughput assay to monitor not only activity, but enantioselectivity of the engineered enzymes in order to quickly and easily identify 'hits' or enzymes with improved enantioselectivity. This proved to be beyond the scope of this project and therefore chiral HPLC was used to test enantioselectivity in the absence of a high-throughput assay. The research done by Cheng *et al.* (2018) which also focused on improving the enantioselectivity of the L-NHase from *R. rhodochrous* J1 also used only chiral HPLC to measure enantioselectivity. This supports the assumption that development of a high-throughput assay requires a separate research effort and this goal remains to be achieved in future using the two-construct expression system in this study.

Microtitre plates, most commonly the 96-well format, are commonly used in screening assays and enzyme assays have been miniaturised to be performed in such plates (Xiao *et al.*, 2015). These assays are

accessible and can be done even without robots and although this increases the labour intensity of the task, it is still higher-throughput than traditional HPLC (Xiao *et al.*, 2015). Such assays have already been applied to identify nitrilases or NHase/amidase systems expressed by *R. erythropolis* CGMCC 1.2362 and *Alcaligenes sp.* ECU040 using ferric hydroxamate spectrophotometry to identify carboxylic acids (He *et al.*, 2011).

There are significant challenges to the development of an assay for enzyme activity, and therefore developing one for enantioselectivity encompasses the challenges of both a basic assay and the identification of activity towards separate enantiomers. High-throughput NMR and mass spectrometry analysis of *pseudo-meso*-compounds in microtitre plates has been done but requires specialised equipment (Reetz, 2003b.). Gas chromatography has also been employed by using one sampler and two instruments for simultaneous analysis of the kinetic resolution of (*R*)-and (*S*)-2-phenyl-1-propanol by the lipase from *Pseudomonas aeruginosa* but this also requires specialised equipment (Reetz, 2003a.).

Due to the cost and equipment requirements, a simpler colourimetric assay was selected which could be visualised using UV or fluorescence spectroscopy. The assay measures the amount of ammonia released during an enzyme reaction. Glutamate dehydrogenase, α-ketoglutarate and NADPH are added to the reaction and the glutamate dehydrogenase then uses up the ammonia to convert α-ketoglutarate to L-glutamine (Figure 4.12) (Sonke *et al.*, 2006). During this reaction, NADPH is oxidised to $NADP^+$. NADPH absorbs at 340 nm and therefore the more NADPH is used up, the more the absorbance decreases. This is directly related to the amount of ammonia originally present, as more ammonia present will allow more NADPH to be converted and thus the absorbance will decrease more.

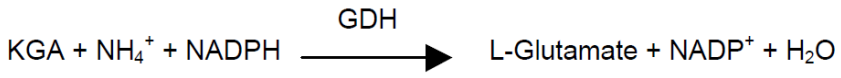


Figure 4.12: Schematic of the reaction catalysed by glutamate dehydrogenase. KGA: α-ketoglutarate. The absorbance of NADPH at 340 nm decreases as it is used up by the reaction and is proportional to the concentration of ammonia originally present. Reproduced from the Ammonia Assay Kit technical bulletin (Sigma-Aldrich, St Louis, MO, USA).

As the reaction being measured is the production of an amide by nitrile hydratase, it would also be necessary to couple an amidase to the assay. According to Sonke *et al.* (2006), the initial enzyme reaction (in this case, the conversion of nitrile to amide by NHase) is carried out in the well of a microtitre plate. The reaction is stopped using acid inactivation. The plate is centrifuged to remove protein or cell debris. An amidase is then added to convert the amide to a carboxylic acid and ammonia. Finally the assay components glutamate dehydrogenase, α-ketoglutarate and NADPH are added and the absorbance at 340 nm is measured.

There were certain disadvantages to this method that would need to be addressed. The NHase reaction occurs in Tris buffer pH 9, therefore it would be necessary to check how this would affect the pH when attempting acid inactivation of the reaction. An ammonia assay kit based on this method and available from Sigma-Aldrich was selected to be used in this method. The kit requires a pH of 7 to 8 to operate. Additionally, the kit was designed for reactions of 1 ml volume and would need to be downscaled to be done in microtitre plate wells. This results in some steps requiring buffer exchange, which requires a vacuum centrifuge with a plate holder, or requiring excess volumes from each well to be removed manually to reduce the volume per well, which increases the labour intensity of the test. This assay has been adapted to measurements of enantioselectivity by simply coupling the assay with an enantioselective amidase, although it may be difficult to acquire a commercially available enantioselective amidase.

4.4 CONCLUSIONS

The chiral substrates 3-Amino-3-*p*-tolyl propanenitrile (**2c**) and 3-Amino-3-(4-methoxyphenyl) propanenitrile (**3c**) were both successfully synthesised, although there were some with impurities present in **3c** which could not be removed during purification with gravity flow column chromatography. These chiral substrates and a non-chiral substrate, 4-methoxybenzonitrile, were successfully used in reactions with recombinantly expressed L-NHase enzyme to produce the product amides 4-methoxybenzamide, 3-Amino-3-*p*-tolyl propanamide (**2d**) and 3-Amino-3-(4-methoxyphenyl) propanamide (**3d**).

All enzyme reactions went to completion very quickly as observed during TLC monitoring of the reactions, with storage conditions of the enzyme showing no effects on enzyme activity. The TLC results showed complete conversion of nitrile to amide for both **2c** and **2d**, suggesting that the enzyme had little to no

enantioselectivity for the selected nitriles. This was in opposition to the results found by Chhiba *et al.* (2012), but it was not possible to confirm the extent of enantioselectivity as chiral amide analysis using HPLC was not conclusive most likely due to column degradation. It is possible that reactions could be monitored with shorter sampling intervals to determine if enantioselectivity occurs before 30 min. It is also possible that enantioselectivity is more evident in larger scale reactions but this remains to be determined. Due to the inconclusive HPLC results it was not possible to quantify conversion of nitriles to amides or to calculate ee_p, ee_s or enantiomeric ratio (E). Furthermore, ^{1}H and ^{13}C NMR need to be done on enzyme reaction products to confirm their structure, although a simple TLC comparison of standard **2d** and an HPLC sample of **2d** did give spots with the same R_f and yellowing indicating presence of an amine group.

Further work would require the chiral amide HPLC procedure to be corrected, and the results can be reinforced using additional chiral nitrile analysis on reaction components to identify the extent of nitrile conversion. It would be ideal to acquire separate standards of both the chiral substrates and products for easier characterisation in future as it was not possible to easily identify which enantiomer was the (*R*)-enantiomer and which was the (*S*)-enantiomer.

In order to meet the demands required by industrial research processes, a high-throughput screening assay for enantioselectivity needs to be developed to screen the recombinantly expressed enzymes for improved enantioselectivity towards these substrates. In the case that an improved enzyme is identified, the enzyme needs to be characterised kinetically, the protein sequence determined and a crystal structure produced.

4.5 REFERENCES

Aguilar, M.-I., Purcell, A.W., Devi, R., Lew, R., Rossjohn, J., Smith, A.I. and Perlmutter, P., 2007. β-Amino acid-containing hybrid peptides—new opportunities in peptidomimetics. *Organic & Biomolecular Chemistry*, 5(18), p.2884.

Ahmed, E. and Holmström, S.J.M., 2014. Siderophores in environmental research: roles and applications. *Microbial Biotechnology*, 7(3), pp.196–208.

Ammonia Assay Kit technical bulletin, 09/15-1. Sigma Aldrich, St Louis, MO, USA, 2015.

Anastas, P. and Eghbali, N., 2009. Green Chemistry: Principles and Practice. *Chemical Society Reviews*, 39(1), pp.301–312.

Andrews, I., Cui, J., Dudin, L., Dunn, P., Hayler, J., Hinkley, B., Hughes, D., Kaptein, B., Lorenz, K., Mathew, S., Rammeloo, T., Wang, L., Wells, A. and White, T.D., 2010. Green Chemistry Articles of Interest to the Pharmaceutical Industry. *Organic Process Research & Development*, 14(4), pp.770–780.

Ansari, S.A., Pathak, P.N., Manchanda, V.K., Husain, M., Prasad, A.K. and Parmar, V.S., 2005. N,N,N′,N′-Tetraoctyl Diglycolamide (TODGA): A Promising Extractant for Actinide-Partitioning from High-Level Waste (HLW). *Solvent Extraction and Ion Exchange*, 23(4), pp.463–479.

Brady, D., Beeton, A., Zeevaart, J., Kgaje, C., van Rantwijk, F. and Sheldon, R.A., 2004. Characterisation of nitrilase and nitrile hydratase biocatalytic systems. *Applied Microbiology and Biotechnology*, 64(1), pp.76–85.

Brandao, P.F.B., Clapp, J.P. and Bull, A.T., 2003. Diversity of Nitrile Hydratase and Amidase Enzyme Genes in Rhodococcus erythropolis Recovered from Geographically Distinct Habitats. *Applied and Environmental Microbiology*, 69(10), pp.5754–5766.

Caboche, S., Pupin, M., Leclère, V., Fontaine, A., Jacques, P. and Kucherov, G., 2008. NORINE: a database of nonribosomal peptides. *Nucleic Acids Research*, 36(suppl_1), pp.D326–D331.

Cheng, Z., Peplowski, L., Cui, W., Xia, Y., Liu, Z., Zhang, J., Kobayashi, M. and Zhou, Z., 2018. Identification of key residues modulating the stereoselectivity of nitrile hydratase toward rac-mandelonitrile by semi-rational engineering. *Biotechnology and Bioengineering*, 115(3), pp.524–535.

Chhiba, V., Bode, M.L., Mathiba, K., Kwezi, W. and Brady, D., 2012. Enantioselective biocatalytic hydrolysis of β-aminonitriles to β-amino-amides using Rhodococcus rhodochrous ATCC BAA-870. *Journal of Molecular Catalysis B: Enzymatic*, 76, pp.68–74.

Choi, J.-M., Han, S.-S. and Kim, H.-S., 2015. Industrial applications of enzyme biocatalysis: Current status and future aspects. *Biotechnology Advances*, 33(7), pp.1443–1454.

Constable, D.J.C., Dunn, P.J., Hayler, J.D., Humphrey, G.R., Leazer, Jr., J.L., Linderman, R.J., Lorenz, K., Manley, J., Pearlman, B.A., Wells, A., Zaks, A. and Zhang, T.Y., 2007. Key green chemistry research areas—a perspective from pharmaceutical manufacturers. *Green Chem.*, 9(5), pp.411–420.

Coward, J.K. and Bruice, T.C., 1969. Catalytic reactions involving azomethines. XII. Mechanism of hydrolysis of primary and secondary enamines. *Journal of the American Chemical Society*, 91(19), pp.5329–5339.

Faber, K., 1997. *Biotransformations in organic chemistry: a textbook*. 3rd, completely rev. ed ed. Berlin ; New York: Springer Verlag.

Frederick, J., 2013. *Genetic characterization of Rhodococcus rhodochrous ATCC BAA-870 with emphasis on nitrile hydrolysing enzymes*. [online] University of Cape Town. Available at: <http://open.uct.ac.za/handle/11427/4262> [Accessed 11 Feb. 2016].

Gilligan, T., Yamada, H. and Nagasawa, T., 1993. Production of S-(+)-2-phenylpropionic acid from (R, S)-2-phenylpropionitrile by the combination of nitrile hydratase and stereoselective amidase in Rhodococcus equi TG328. *Applied microbiology and biotechnology*, 39(6), pp.720–725.

Greenberg, A., 2002. The amide linkage as a ligand: its properties and the role of distortion. In: *The Amide Linkage: Structural Significance in Chemistry, Biochemistry, and Materials Science*. John Wiley & Sons.

He, Y.-C., Ma, C.-L., Xu, J.-H. and Zhou, L., 2011. A high-throughput screening strategy for nitrile-hydrolyzing enzymes based on ferric hydroxamate spectrophotometry. *Applied Microbiology & Biotechnology*, 89(3), pp.817–823.

Heinrichs, V. and Thum, O., 2005. Biocatalysis for the Production of Care Specialties. *Lipid Technology*, 17(4).

Instruction manual for Crownpak ® Cr(+) / Cr(-), 07/2013LE. Daicel Corporation, Tokyo, Japan, 2013.

Jain, S.K., Kumar, K.A.A., Bharate, S.B. and Vishwakarma, R.A., 2014. Facile access to amides and hydroxamic acids directly from nitroarenes. *Organic & Biomolecular Chemistry*, 12(33), pp.6465–6469.

Konkle, M.E., Hargrove, T.Y., Kleshchenko, Y.Y., Kries, J.P. von, Ridenour, W., Uddin, M.J., Caprioli, R.M., Marnett, L.J., Nes, W.D., Villalta, F., Waterman, M.R. and Lepesheva, G.I., 2009. Indomethacin Amides as a Novel Molecular Scaffold for Targeting Trypanosoma cruzi Sterol 14α-Demethylase. *Journal of Medicinal Chemistry*, 52, pp.2846–2853.

Oishi, T., Yamaguchi, K. and Mizuno, N., 2009. Catalytic Oxidative Synthesis of Nitriles Directly from Primary Alcohols and Ammonia. *Angewandte Chemie International Edition*, 48(34), pp.6286–6288.

Payne, M.S., Wu, S., Fallon, R.D., Tudor, G., Stieglitz, B., Turner, I.M. and Nelson, M.J., 1997. A stereoselective cobalt-containing nitrile hydratase. *Biochemistry*, 36(18), pp.5447–5454.

Petsko, G.A. and Ringe, D., 2004. *Protein Structure and Function*. New Science Press.

Pirttimaa, M., Nasereddin, A., Kopelyanskiy, D., Kaiser, M., Yli-Kauhaluoma, J., Oksman-Caldentey, K.-M., Brun, R., Jaffe, C.L., Moreira, V.M. and Alakurtti, S., 2016. Abietane-Type Diterpenoid Amides with Highly Potent and Selective Activity against Leishmania donovani and Trypanosoma cruzi. *Journal of Natural Products*, 79(2), pp.362–368.

Pitzer, J. and Steiner, K., 2016. Amides in Nature and Biocatalysis. *Journal of Biotechnology*, 235, pp.32–46.

Poliakoff, M., Fitzpatrick, J.M., Farren, T.R. and Anastas, P.T., 2002. Green Chemistry: Science and Politics of Change. *Science*, 297, pp.807–810.

Preiml, M., Hillmayer, K. and Klempier, N., 2003. A new approach to β-amino acids: biotransformation of N-protected β-amino nitriles. *Tetrahedron Letters*, 44(27), pp.5057–5059.

Reetz, M. T., 2003a. An overview of high-throughput screening systems for enantioselective enzymatic transformations. In: F. H. Arnold and G. Geogiou, eds. 2003. *Directed Enzyme Evolution*. New York: Springer. Ch.26.

Reetz, M. T., 2003b. Select protocols of high-throughput ee-screening systems for assaying enantioselective enzymes. In: F. H. Arnold and G. Geogiou, eds. 2003. *Directed Enzyme Evolution*. New York: Springer. Ch.27.

Rzeznicka, K., Schätzle, S., Böttcher, D., Klein, J. and Bornscheuer, U.T., 2010. Cloning and functional expression of a nitrile hydratase (NHase) from Rhodococcus equi TG328-2 in Escherichia coli, its purification and biochemical characterisation. *Applied Microbiology and Biotechnology*, 85(5), pp.1417–1425.

Sawant, S.D., Nerkar, A.G., Pawar, N.D. and Velapure, A.V., 2014. DESIGN, SYNTHESIS, QSAR STUDIES AND BIOLOGICAL EVALUATION OF NOVEL TRIAZOLOPIPERAZINE BASED B-AMINO AMIDES AS DIPEPTIDYL PEPTIDASE-IV (DPP-IV) INHIBITORS: PART-II. *International Journal of Pharmacy and Pharmaceutical Sciences*, 6(Suppl. 2), pp.812–817.

Schmid, A., Hollmann, F., Park, J.B. and Bühler, B., 2002. The use of enzymes in the chemical industry in Europe. *Current Opinion in Biotechnology*, 13(4), pp.359–366.

Simon, R.C., Mutti, F.G. and Kroutil, W., 2013. Biocatalytic synthesis of enantiopure building blocks for pharmaceuticals. *Drug Discovery Today: Technologies*, 10(1), pp.e37–e44.

Sonke, T., Duchateau, L., Schipper, D., Euverink, G.-J., van der Wal, S., Henderickx, H., Bezemer, R., Vollebregt, A. 2006. Industrial Perspectives on Assays. In: J.-L. Reymond, ed. 2006. *Enzyme Assays*. Weinheim: Wiley-VCH. Ch.4.

Tachdjian, C., Patron, A., Qi, M., Adamski-Werner, S., Tang, X.-Q., Chen, Q., Darmohusodo, V., Lebl-Rinnova, M. and Priest, C., 2006. *Aromatic amides and ureas and their uses as sweet and/or umami flavor modifiers, tastants and taste enhancers*. [online] US20060045953A1. Available at: <https://patents.google.com/patent/US20060045953A1/en> [Accessed 11 Oct. 2018].

Tachdjian, C., Patron, A.P., Bakir, F., Averbuj, C., Priest, C., Adamski-Werner, S.L., Chen, Q., Darmohusodo, V., Lebl-Rinnova, M., Kimmich, R.D.A., Tang, X.-Q. and Shigemura, R., 2010. *Bis-aromatic amides and their uses as sweet flavor modifiers, tastants, and taste enhancers*. [online] US7842324B2. Available at: <https://patents.google.com/patent/US7842324B2/en> [Accessed 11 Oct. 2018].

Tao, J. and Xu, J.-H., 2009. Biocatalysis in development of green pharmaceutical processes. *Current Opinion in Chemical Biology*, 13(1), pp.43–50.

Thiollet, G. and Musikas, C., 1989. SYNTHESIS AND USES OF THE AMIDES EXTRACTANTS. *Solvent Extraction and Ion Exchange*, 7, pp.813–827.

Troeberg, L., Chen, X., Flaherty, T.M., Morty, R.E., Cheng, M., Hua, H., Springer, C., McKerrow, J.H., Kenyon, G.L., Lonsdale-Eccles, J.D., Coetzer, T.H.T. and Cohen, F.E., 2000. Chalcone, Acyl Hydrazide, and Related Amides Kill Cultured Trypanosoma brucei brucei. *Molecular Medicine*, 6(8), pp.660–669.

van Pelt, S., Zhang, M., Otten, L.G., Holt, J., Sorokin, D.Y., van Rantwijk, F., Black, G.W., Perry, J.J. and Sheldon, R.A., 2011. Probing the enantioselectivity of a diverse group of purified cobalt-centred nitrile hydratases. *Organic & Biomolecular Chemistry*, 9(8), p.3011.

Wenig, P. and Odermatt, J., 2010. OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data. *BMC Bioinformatics*, 11, p.405.

Winkler, M., Martínková, L., Knall, A.C., Krahulec, S. and Klempier, N., 2005. Synthesis and microbial transformation of β-amino nitriles. *Tetrahedron*, 61(17), pp.4249–4260.

Wu, Z.-L. and Li, Z.-Y., 2003. Enantioselective biotransformation of α,α-disubstituted dinitriles to the corresponding 2-cyanoacetamides using Rhodococcus sp. CGMCC 0497. *Tetrahedron: Asymmetry*, 14(15), pp.2133–2142.

Xiao, H., Bao, Z. and Zhao, H., 2015. High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Industrial & Engineering Chemistry Research*, 54(16), pp.4011–4020.

Xie, Z., Chen, R., Du, Z., Kong, L., Li, Z., Li, Z., Wang, N. and Liu, J., 2017. Iodine-Catalyzed Formation of Amide Bond: Efficient Strategy for the Synthesis of Aromatic Primary Amides. *Asian Journal of Organic Chemistry*, 6(2), pp.157–160.

Yu, H.-M., Shi, Y., Luo, H., Tian, Z.-L., Zhu, Y.-Q. and Shen, Z.-Y., 2006. An over expression and high efficient mutation system of a cobalt-containing nitrile hydratase. *Journal of Molecular Catalysis B: Enzymatic*, 43(1–4), pp.80–85.

4.6 APPENDIX

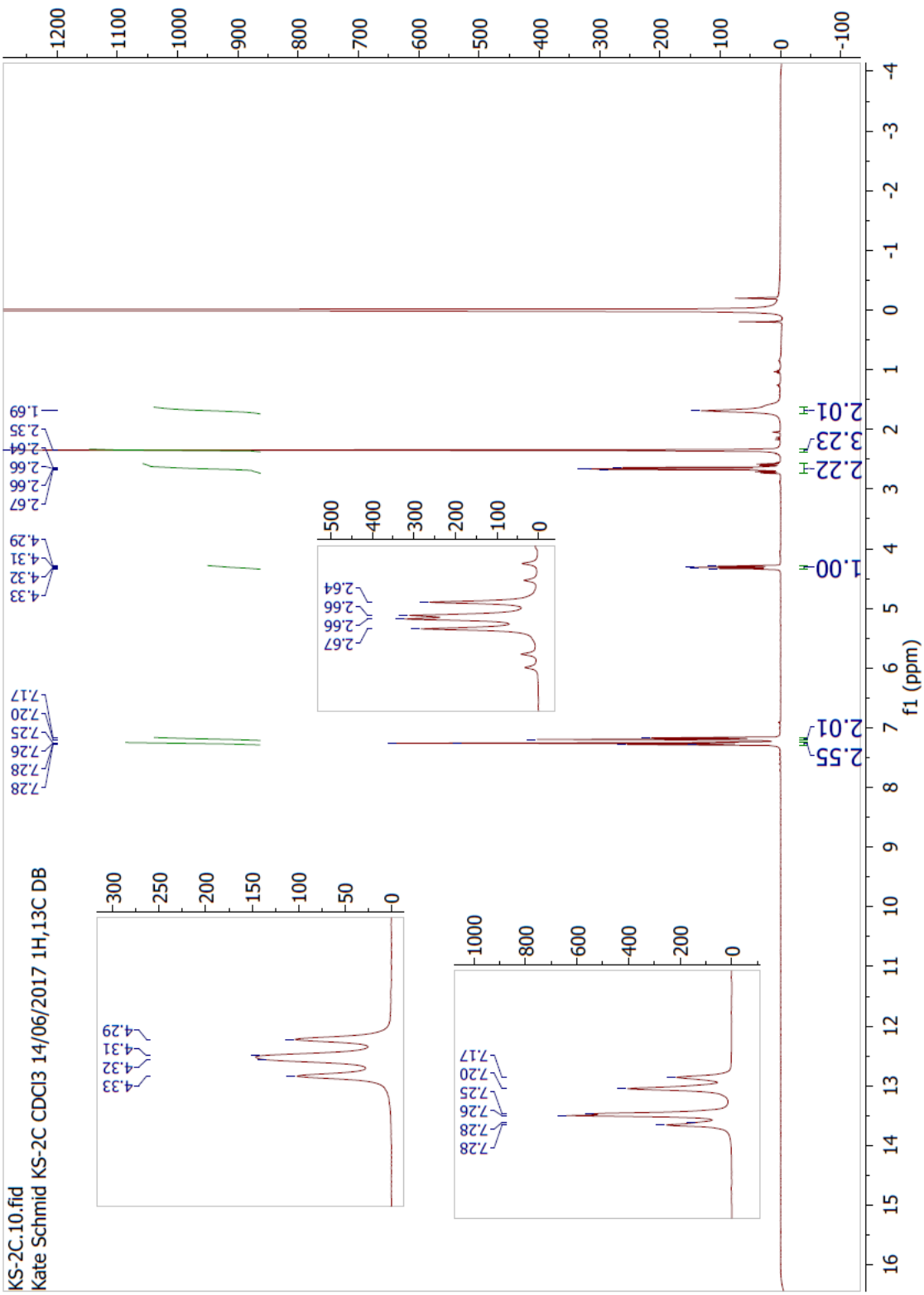


Figure A1.1 ^{1}H NMR spectrum of substrate 2c

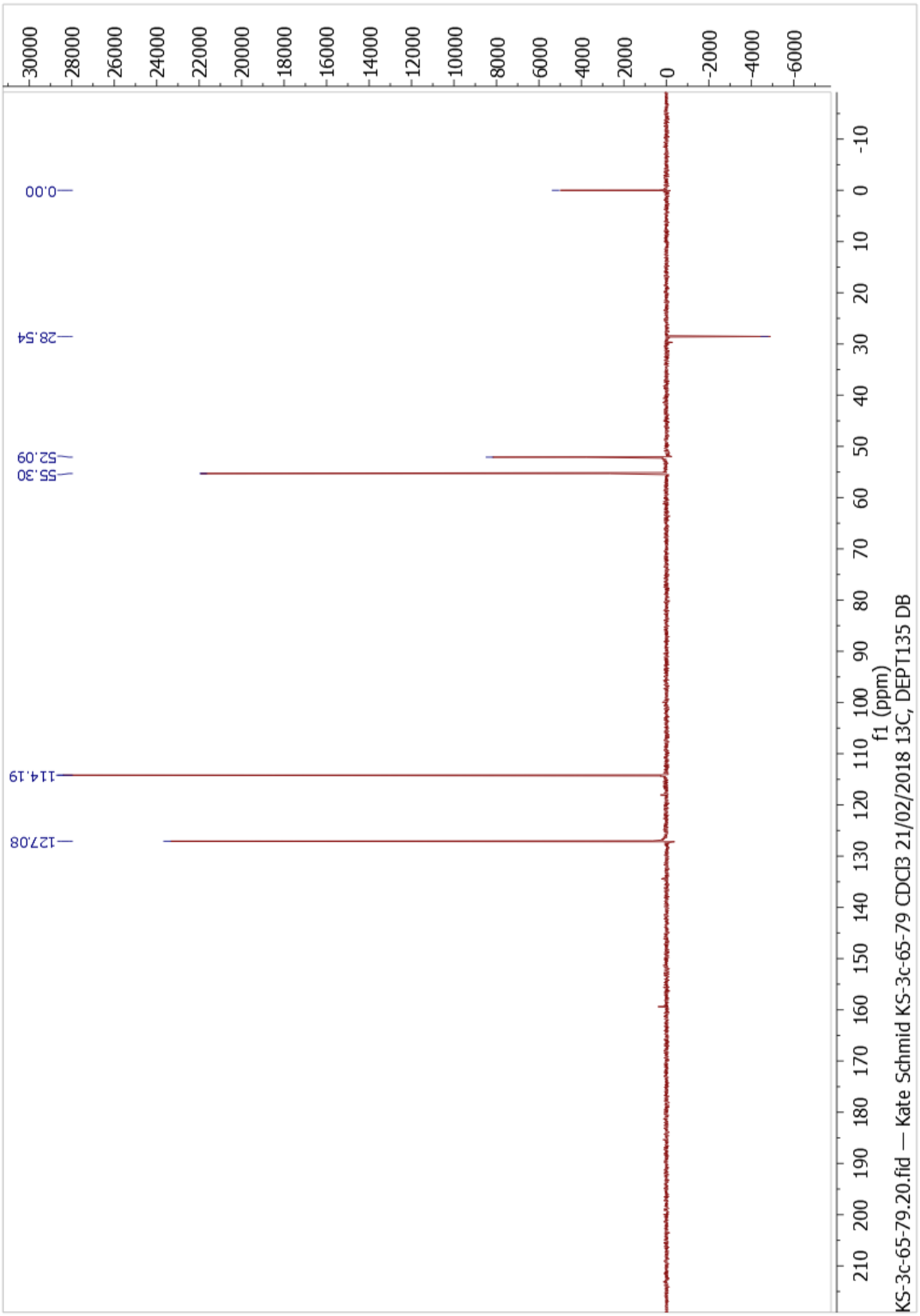


Figure A1.2 DEPT135 NMR spectrum of substrate 2c

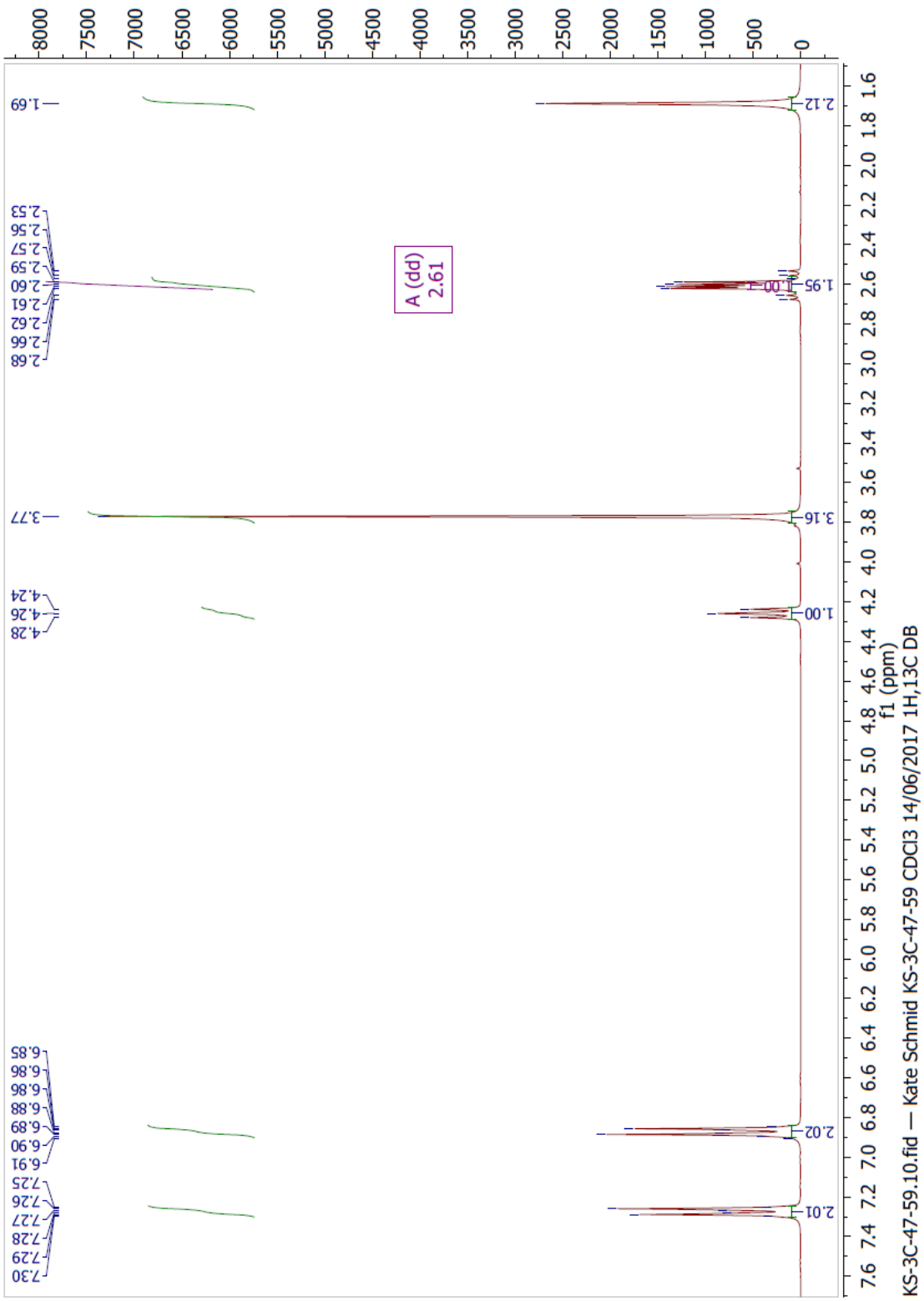


Figure A1.3 ^{1}H NMR spectrum of substrate 3c

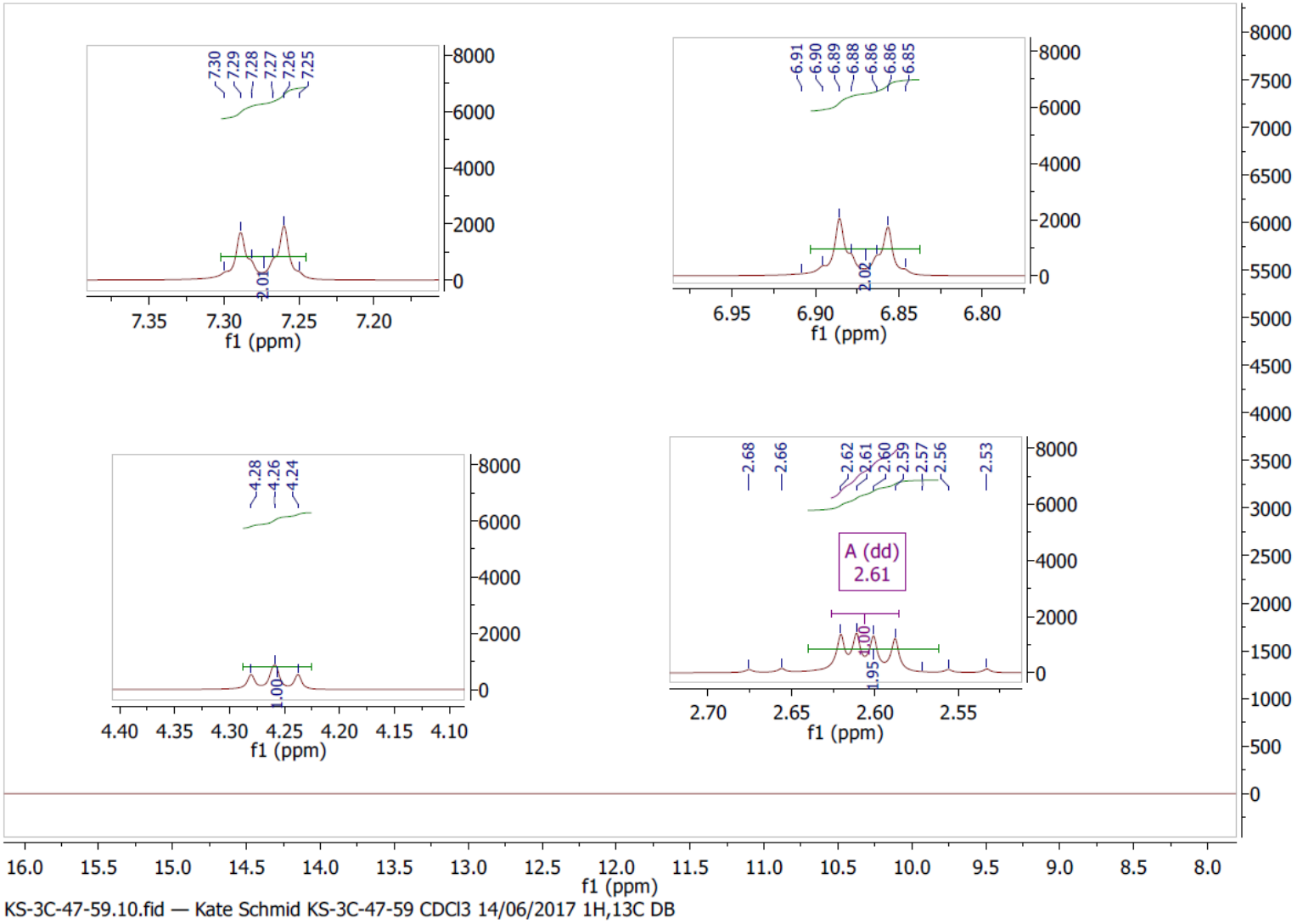


Figure A1.4 Magnified regions of ^{1}H NMR spectrum of substrate 3c showing multiplets

CONCLUSION AND FUTURE WORK

Members of the genus *Rhodococcus* are a source of highly useful enzymes for industrial biocatalysis, in particular the nitrile hydrolysing enzymes. The high-molecular mass NHase (H-NHase) from *R. rhodochrous* J1 has already been developed as a viable industrial biocatalyst, but the low-molecular mass NHase (L-NHase) which is more commonly found *in R. rhodochrous* strains, also has potential use in the production of chemicals, including enantiopure beta-substituted amino-nitriles, which formed the basis of the current study. However, this enzyme is not naturally enantioselective for most substrates and therefore rational design must be applied to produce an enantioselective NHase. Rational design of enzymes is an important research subject in the current movement towards greener industrial processes. Enzymes can be designed to be highly specific and to produce large yields with decreased costs and energy consumption, while producing less waste.

This study aimed to recombinantly express the NHase from *R. rhodochrous* ATCC BAA-870 (a close relative of the L-NHase from *R. rhodochrous* J1) and determine which amino acid residues may be involved in substrate specificity by comparison to related NHase sequences that were determined during this study. The second major goal was to be able to modify the substrate profile of the enzyme by introducing mutations, in this case to improve its enantioselectivity towards beta-substituted nitriles for future use in pharmaceutical production.

A two-construct method for recombinant expression of the L-NHase from *R. rhodochrous* ATCC BAA-870 in BL21(DE3) involved expressing the alpha subunit and chaperone in the pRSFDuet-1 plasmid and the beta subunit in the pET-21-a(+) plasmid and transforming one culture with both plasmids. The current study improved an earlier version of this expression system using codon optimisation to improve expression of the NHase. This system also allowed for independent mutation of the beta subunit without possibly affecting the plasmid containing the alpha subunit and chaperone sequences, which were necessary for enzyme stability. The activity of the recombinant enzyme was confirmed and compared to an alternative single construct expression system developed by a collaborating research group.

Data-mining analysis was done on the NHase structure and substrate profile of multiple recorded *Rhodococci* including five new *Rhodococci* from South African soil samples closely related to *R. rhodochrous* ATCC BAA-870 and *R. rhodochrous* J1. This was done to identify patterns associated with

substrate catalysis that could be linked to genetic characteristics in order to identify target residues for directed evolution. Although this analysis was inconclusive, it did indicate a possible origin for the H-NHase found in *R. rhodochrous* J1 as being acquired via horizontal gene transfer from another *Rhodococcus* such as *R. pyridivorans*. A homology model of the L-NHase from *R. rhodochrous* ATCC BAA-870 was produced and the substrate channel residues were identified as reported in previous structural studies. As well as this, the residues within 10 Å of the active site were also located as residues in this region have previously been reported to play a role in enzyme enantioselectivity and to be viable targets for directed evolution. Based on the homology model and previously reported substrate-channel information, the residues chosen for mutation were βPhe41 and βAla37.

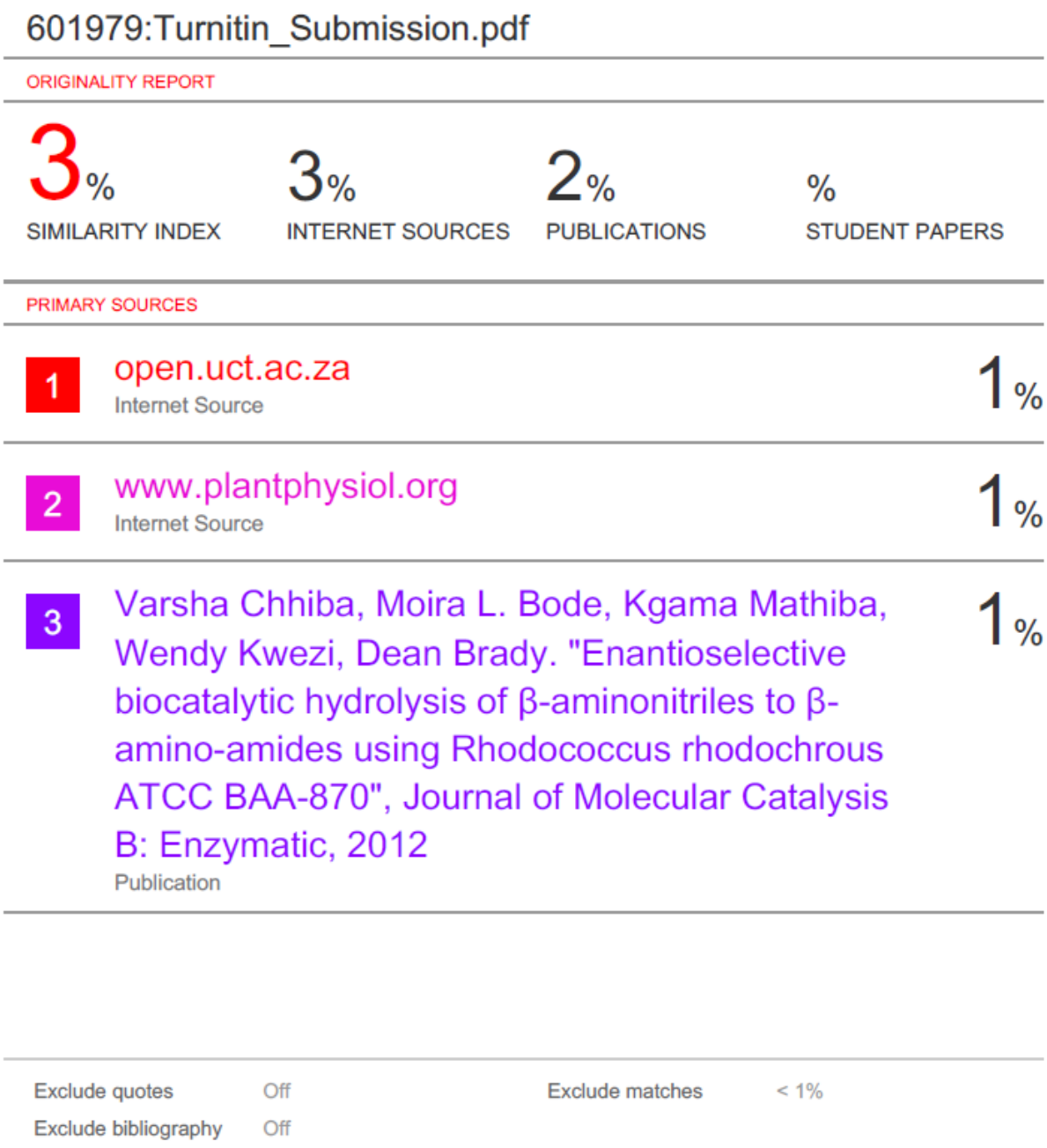
Directed evolution of enzymes has been applied to alter many enzyme characteristics, including thermostability, pH tolerance, substrate selectivity and enantioselectivity, including prior successful attempts to alter the stereoselectivity of other NHases. Semi-rational design has been found to be successful and involves using prior knowledge of enzyme-substrate activity, visualising interactions using molecular modelling and rationally selecting residues or sites containing multiple residues for mutation. These sites are then randomly mutated to produce a library of potential improved enzymes. The strategy selected for this study involved successive rounds of iterative saturation mutagenesis via polymerase chain reaction (ISM-PCR) using the Combinatorial Active Site Saturation Test (CASTing) which focuses on mutating residues within the binding pocket. ISM-PCR was done using NDT-degenerate primers and using the megaprimer method for whole-plasmid PCR which removed laborious cloning steps following PCR. It was found that mutations were introduced, but not at the expected site. The mutations were introduced at two adjacent residues at the N-terminus, βHis5 and βAsp6, indicating that the mutational primer had bound in the shape of a hairpin. The ratios of introduced nucleotides did not follow the expected pattern. The PCR technique was therefore partially successful and the protocol and primer design must be optimised to produce preferred results.

Two chiral beta-amino nitrile substrates, 3-amino-3-*p*-tolyl propanenitrile and 3-amino-3-(4-methoxyphenyl) propanenitrile were successfully synthesised and their structure confirmed using NMR. A non-chiral substrate, 4-methoxybenzonitrile, was used to confirm the activity of the recombinant L-NHase. Biocatalytic activity was successfully demonstrated on all three substrates, Chiral HPLC was attempted to quantify enantioselectivity of the enzyme, but the results were inconclusive due to column performance degradation. Monitoring of enzyme reactions using thin-layer chromatography

demonstrated complete conversion of nitriles to amides, indicating a lack of absolute enantioselectivity, but we were unable to confirm this with chiral HPLC.

This study has assisted in laying a foundation for future rational engineering of the nitrile hydratase from *R. rhodochrous* ATCC BAA-870 by providing key data and evaluating the relevant methods. Future work should focus on optimising the PCR method to introduce targeted mutations into the beta subunit and on adjusting and successfully using chiral HPLC to monitor enantioselectivity of the recombinant enzymes It is necessary to develop a high-throughput screening assay for faster, automated screening of enantioselectivity. The enantioselectivity of the NHase for other substrates can be explored in combination with molecular modelling and molecular dynamics simulations to identify the residues involved in interactions with specific substrates. Therefore future projects may focus on adjusting the enantioselectivity of the L-NHase for other substrates. An improved NHase could then undergo kinetic studies, sequencing and crystallisation to provide further information about how the enzyme acquired improved enantioselectivity.

Sources of <1% filtered out:



601979:Turnitin_Submission.pdf

ORIGINALITY REPORT

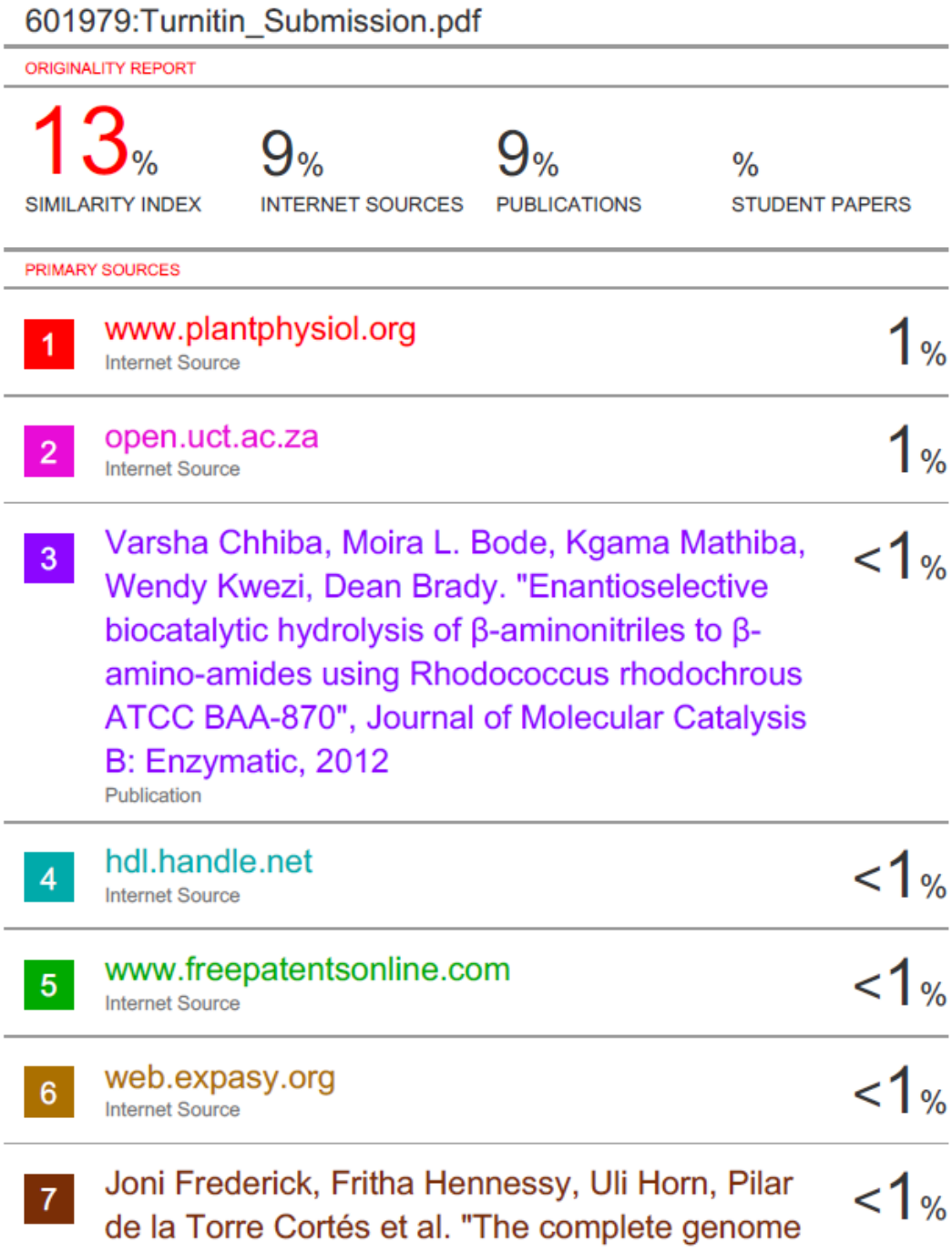
3%	3%	2%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	open.uct.ac.za Internet Source	1%
2	www.plantphysiol.org Internet Source	1%
3	Varsha Chhiba, Moira L. Bode, Kgama Mathiba, Wendy Kwezi, Dean Brady. "Enantioselective biocatalytic hydrolysis of β-aminonitriles to β-amino-amides using Rhodococcus rhodochrous ATCC BAA-870", Journal of Molecular Catalysis B: Enzymatic, 2012 Publication	1%

Exclude quotes	Off	Exclude matches	< 1%
Exclude bibliography	Off		

When including sources of <1%:



601979:Turnitin_Submission.pdf

ORIGINALITY REPORT

13%	9%	9%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.plantphysiol.org Internet Source	1%
2	open.uct.ac.za Internet Source	1%
3	Varsha Chhiba, Moira L. Bode, Kgama Mathiba, Wendy Kwezi, Dean Brady. "Enantioselective biocatalytic hydrolysis of β-aminonitriles to β-amino-amides using Rhodococcus rhodochrous ATCC BAA-870", Journal of Molecular Catalysis B: Enzymatic, 2012 Publication	<1%
4	hdl.handle.net Internet Source	<1%
5	www.freepatentsonline.com Internet Source	<1%
6	web.expasy.org Internet Source	<1%
7	Joni Frederick, Fritha Hennessy, Uli Horn, Pilar de la Torre Cortés et al. "The complete genome	<1%