Genetic characterization of *Rhodococcus rhodochrous* ATCC BAA-870 with emphasis on nitrile hydrolysing enzymes

Joni Frederick

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Molecular and Cell Biology,
University of Cape Town

Supervisor: Professor B. T. Sewell
Co-supervisor: Professor D. Brady

February 2013

Keywords

Nitrile hydrolysis

Biocatalysis

Rhodococcus rhodochrous ATCC BAA-870

Genome sequencing

Nitrilase

Nitrile hydratase

Abstract

Rhodococcus rhodochrous ATCC BAA-870 (BAA-870) had previously been isolated on selective media for enrichment of nitrile hydrolysing bacteria. The organism was found to have a wide substrate range, with activity against aliphatics, aromatics, and aryl aliphatics, and enantioselectivity towards beta substituted nitriles and beta amino nitriles, compounds that have potential applications in the pharmaceutical industry. This makes *R. rhodochrous* ATCC BAA-870 potentially a versatile biocatalyst for the synthesis of a broad range of compounds with amide and carboxylic acid groups that can be derived from structurally related nitrile precursors. The selectivity of biocatalysts allows for high product yields and better atom economy than non-selective chemical methods of performing this reaction, such as acid or base hydrolysis.

In order to apply BAA-870 as a nitrile biocatalyst and to mine the organism for biotechnological uses, the genome was sequenced using Solexa technology and an Illumina Genome Analyzer. The Solexa sequencing output data was analysed using the Solexa Data Analysis Pipeline and a total of 5,643,967 reads, 36-bp in length, were obtained providing 4,273,289 unique sequences. The genome sequence data was assembled using the software Edena, Velvet, and Staden. The best assembly data set was then annotated automatically using dCAS and BASys. Further mate-paired sequencing, contracted to the company BaseClear® BV in Leiden, the Netherlands, was performed in order to improve the completeness of the data. The scaffolded Illumina and mate-paired sequences were further assembled and annotated using BASys. BAA-870 has a GC content of 65% and contains 6997 predicted protein-coding sequences (CDS). Of this, 54% encodes previously identified proteins of unknown function. The completed 5.83 Mb genome (with a sequencing coverage of 135 X) was submitted to the NCBI Genome data bank with accession number PRJNA78009. The genome sequence of *R. rhodochrous* ATCC BAA-870 is the seventh rhodococcal genome to be submitted to the NCBI and the first *R. rhodochrous* subtype to be sequenced. An analysis of the genome for nitrile-

metabolising enzymes revealed that BAA-870 contains both a single nitrilase and a low molecular weight nitrile hydratase. An amidase is associated with the nitrile hydratase, although many other amidases are also present on the genome. Additionally, the genes for numerous other enzymes from commercially used enzyme classes are identified, including proteases, reductases, transaminase, epoxide hydrolase, monooxygenase, cytochrome P450 and haloperoxidase.

The nitrilase gene was isolated by PCR and directionally ligated into the expression vector pET28a(+). The nitrilase sequence was checked against a nucleotide database using BLAST, and a 98% and 97% identity at the nucleotide level to the nitrilase gene from *R. rhodochrous* tg1-A6 and *R. rhodochrous* J1 (J1) respectively was found. This was confirmed independently by purification of the enzyme from the wild type culture and sequence determination using mass spectrometry. Observing the purified wild-type nitrilase by electron microscopy showed that, like the nitrilase from *R. rhodochrous* J1, it formed fibre assemblies.

The recombinantly expressed nitrilase in *E. coli* was extracted by cell disruption, and purified by Protino-Ni resin and Sephacryl S200 size exclusion chromatography. Activity was followed using benzonitrile as a substrate, by both colorimetric ammonia assay and by HPLC analysis. BAA-870 nitrilase forms helices with molecular masses of over 600 kDa. The diameter of the BAA-870 fibre was measured from scaled electron micrographs, and found to be 120 Å. Before fibre-formation, the 'c'-shaped nitrilase structures were observed using electron microscopy. Single particles were selected using sxboxer in the SPARX suite of programs, and the set of images manually filtered, and translationally and rotationally aligned. Alignment was done iteratively over many cycles and classes classified into k means. The reconstructed 'c' was compared to the structure of the J1 nitrilase. The reconstructed J1 fibre's 3D information was used to generate models of J1 nitrilase oligomers containing only 8 and 10-monomers. The BAA-870 nitrilase 'c' structure was structurally aligned with the 8-monomer volume of J1, suggesting that the BAA-870 nitrilase pre-fibre form likely consists of 8 monomers. The BAA-870 nitrilase pre-fibre form was found to be a left-handed, c-shaped particle. Final 3D reconstruction showed that there are 9-10

subunits (~5 dimers) forming the 1-start, left-handed, c-shaped oligomer, suggesting that the c-shaped nitrilase could form either an octamer or a decamer.

The low-molecular mass nitrile hydratase (L-NHase) gene from *R. rhodochrous* ATCC BAA-870 was amplified from the genome of the organism using PCR. The amplified L-NHase gene cassette contained the beta subunit, alpha subunit and chaperone as confirmed by DNA sequencing. The PCR product was directionally ligated into the expression vector pET28a(+) and confirmed by sequencing. Expression studies of the L-NHase showed that the beta subunit and the L-NHase chaperone were overexpressed, but not the alpha subunit. The L-NHase-containing clone was additionally mutated at the ribosome binding site (RBS) for the alpha subunit, as it was found to be a particularly rare site. Further expression of L-NHase in *E. coli* could be achieved, but at low protein yields. The L-NHase was then successfully cloned using a two construct vector system, for coexpression in *E. coli*. The alpha subunit and chaperone, and beta subunit with a C-terminus His-tag, were expressed independently in the same culture medium using IPTG and cobalt chloride. The recombinantly expressed nitrile hydratase was extracted by cell disruption and purified by Protino-Ni affinity and ion exchange chromatography. Activity of NHase was followed by continuous spectrophotometric assay, and by HPLC, using benzonitrile as substrate. The nucleotide sequence of L-NHase from BAA-870 was 97% identical to that from *R. rhodochrous* J1, the differences occurring only in the beta subunit. NHase from BAA-870 was modelled by secondary structural prediction and a structural alignment performed using NHase from *B. pallidus* Rapc8.

The enzymes responsible for nitrile degradation in *R. rhodochrous* ATCC BAA-870 are now independently accessible in order to provide complete characterization without interference from potentially competing enzymes. Structural insights have been obtained using modelling and EM reconstruction (for nitrilase). These enzymes may now be more easily genetically modified to provide improved stability or change substrate preference. The knowledge of the overall genome may also provide insights into how to modify the cell metabolism to prevent further catabolism of biocatalytic products, and also to mine for new biocatalytically useful enzymes.

Declaration

I declare that the thesis entitled

"Genetic characterization of *Rhodococcus rhodochrous* ATCC BAA-870 with emphasis on nitrile hydrolysing enzymes"

is my own work, that it has not been submitted before for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged as complete references.

Joni Frederick February 2013

Signed: Signed by candidate

Signature Removed

Acknowledgements

My heartfelt thanks to the supervisors and mentors of this project: Dr Dean Brady of the Enzyme Technologies Group at CSIR Biosciences in Modderfontein, Johannesburg, and Prof Trevor Sewell, of the Electron Microscope Unit at the University of Cape Town. The support and patience of Prof Brady and Prof Sewell are duly appreciated. CSIR Biosciences is gratefully acknowledged for the support and funding during the duration of the project. The National Research Foundation is thanked for the funding of a visit to Synchrotron SOLEIL. Proxima I Beamline at Synchrotron SOLEIL and Delft Technical University (especially Dr Linda Otten) are thanked for hosting me during parts of the project.

A special thanks to those who helped technically during this project. Your input, training, advice and/or discussion are greatly appreciated. Specifically I thank Dr Fritha Hennessy (for knowledgeable bioinformatics, sequencing and genomics assistance), Mr Uli Horn (for technical help in genome sequence assembly pipelines and computing), Dr Daniel Visser (for assistance in batch fermentation of wildtype nitrile hydratase), Dr Brandon Weber (for extensive column chromatography and protein preparation discussion), Mr Mohamed Jaffer (for making fancy microscopes seem less scary), Ndoriah Thuku (for nitrilase insights and huge encouragement), Jeremy Woodward, Jason van Rooyen and Jean Watermeyer (for EM reconstruction), Dr Koni Rashamuse (for many cloning discussions, advice and input), Dr Arvind Varsani (for much insight into cloning, sequencing and molecular work), and Joy Blain (for countless paperwork and administration).

Dad, Mom, Bonnie – thanks for always showing such interest and admiration of my work. Your pride, love and encouragement have helped me tremendously, made me feel like a hero, and kept me going. Dad, thank you for always helping me financially so I could focus on becoming an independent woman…

I thank also my very good, understanding, immensely patient and beautiful friends, for helping me to find puzzle pieces, for the tissues and shoulders, the social occasions that kept things normal, sharing wine and meals, and the lending of ears and sometimes couches. Fritha (owed a lifetime of chocolate for many reasons), Varsha (an inspirational force), Amanda and James and Sausage and Weasel (formidable, unflailing friends and life savers), Anu, Asha, Tina, Morgwyn, Kelly, Nichole, Justin, Ashleigh: you all have enriched my life and helped in more ways than you may ever know.

And to the only one who ever really truly understood, listened, loved unconditionally and stood by me through everything without any hesitation, I thank my dog, Bob the Bunny.

This thesis is dedicated to the memory of a fascinating and talented man, David Llewellyn Frederick. I always promised you I would do this, and I wish you could have stayed around longer to see it.

Forward

Cobalt

n.

Symbol Co.

Hard ferromagnetic silver-white bivalent or trivalent metallic element. Trace element in plant and animal nutrition.[1]

German: Kobalt, Kobolt; a goblin, evil spirit, malicious sprite, malevolent earth spirit.[2]

[1] American Heritage® Science Dictionary

[2] Germanenglishwords.com; wordinfo.info

Table of Contents

List of Tables

List of Figures

List of Equations

Abbreviations List

A	absorbance
Å	Ångstrom
ABC	ATP-Binding Cassette superfamily
ACN	acetonitrile
ACV	alpha-aminoadipyl-L-cysteinyl-D-valine
Amd	amidase
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
BAA-870	*Rhodococcus rhodochrous* ATCC BAA-870
BASys	Bacterial Annotation System web server
BLAST	Basic local alignment search tool
bp	base pairs
BSA	bovine serum albumin
BTB	bromothymol blue
BVMO	Baeyer-Villiger monooxygenase
CDMO	cyclododecanone monooxygenase
CDS	protein-coding sequences
CHEF	clamped homogeneous electric fields
CHMO	cyclohexanone monooxygenase
CMN	Corynebacterium, Mycobacterium and Nocardia
COG	Clusters of Orthologous Groups
Contig	contiguous sequence
CSIR	Council for Scientific and Industrial Research
Da	Dalton
DCase	*N*-Carbamyl-D-amino acid amido-hydrolase
DHB	dihydroxy-benzoate
DHFR	Dihydrofolate reductase
DHQase	dehydroquinate dehydratase
DTT	dithiothreitol

E.C.	enzyme commission
Edena	Exact DE Novo Assembler programme
EDTA	ethylene diamine tetraacetic acid
ee	enantiomeric excess
EM	electron microscopy
ENDOR	electron nuclear double resonance
EPR	electron paramagnetic resonance
EXAFS	extended x-ray absorption fine structure
FABI	Forestry and Agricultural Biotechnology Institute
FAD	flavin adenine dinucleotide
Fhit	fragile histidine triad protein
FSC	Fourier shell correlation
GO	Gene Ontology
3HB-co3HV	co-polyester of 3-hydroxybutyrate and 3-hydroxyvalerate
HGT	horizontal gene transfer
H-NHase	high molecular mass nitrile hydratase
HPLC	high performance liquid chromatography
IPTG	isopropyl-β-D-1-thiogalactopyranoside
Ismall	*Rhodococcus rhodochrous* ATCC BAA-870
Kav	partition coefficient
kb	Kilo base
k_{cat}	enzyme catalytic constant
k_{cat}/K_m	enzyme efficiency/ turnover
kDa	kilodaltons
K_m	Michaelis constant
LB	Luria Bertani medium
LB-ampicillin	LB medium containing 100 µg/ml ampicillin
LB-kanamycin	LB medium containing 50 µg/ml kanamycin
L-NHase	low molecular mass nitrile hydratase
m	minimum contig size
M (assembly)	overlap length

MALDI-TOF	Matrix-assisted laser desorption/ionization, time-of-flight mass spectroscopy
Mbp	mega base pairs
mce	mammalian cell entry protein/ gene
MDa	mega dalton
MFS	Major Facilitator Superfamily
MRA	Multi reference alignment
Mr	relative molecular mass
N50	contig size of value X such that at least half of the genome is contained in contigs of size X or larger
NaCl	sodium chloride
NAD	nicotinamide adenine dinucleotide
NADP(H)	nicotinamide adenine dinucleotide phosphate (reduced)
NCBI	National Center for Biotechnology Information
NHase	nitrile hydratase
NHLS	National health Laboratory Service
Nit	nitrilase
NitFhit	nitrilase-fragile histidine triad fusion protein
NO	nitric oxide
NSAID	non-steroidal anti-inflammatory drug
nt	nucleotide
OD	optical density
OPA	*o*-phthaldialdehyde
ORF	open reading frame
PCB's	polychlorinated biphenyls
PCP	pentachlorophenol
PCR	polymerase chain reaction
PDA	photo diode array
PDB	protein data bank
PFG	pulsed-field gel
PFGE	pulsed-field gel electrophoresis
PSIPRED	protein secondary structure prediction

RP-HPLC	reversed-phase high performance liquid chromatography
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SRS	short read sequence
TBE	Tris-borate-EDTA
TCA	tricarboxylic acid cycle
TE	Tris-EDTA
TEM	transmission electron microscopy
TEMED	N,N,N'N'-tetramethylethylenediamine
Tm	melting temperature
TFA	trifluoroacetic acid
Tris	Tris(hydroxymethyl)-aminomethane
TSA	tryptone soya agar
TSB	tryptone soya broth
TU Delft	Delft Technical University
UV	ultraviolet
V_e	elution volume
V_o	void volume
V_t	total column volume
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
Δϕ	rotation per dimeric subunit
Δz	axial rise per dimer

The IUPAC-IUBMB one and three letter codes are used for amino acids.

Chapter One

1. Identifying nitrile degrading *Rhodococci*, and outlining toolboxes for nitrile biocatalysis

1.1. Introduction

1.1.1. The rationale for nitrile conversion (amides and acids, and their uses)

Nitrile compounds (R-C≡N) contain the cyanide group (-CN), and therefore used to be referred to generally as organic cyanides. They are widespread in nature where they are mainly present as cyanoglycosides, cyanolipids, ricinine and phenylacetonitrile produced by plants (Conn 1981). Synthetically, nitriles can be made by one of several routes, including the dehydration of amides, or addition of hydrogen cyanide to carbonyl groups (aldehydes and ketones). Nitriles are important and versatile intermediates in organic synthesis, and are used extensively in industry for the production of polymers and chemicals. For example, nylon-6,6 polymers and polyacrylonitrile (a plastics and acrylic fibre precursor) are produced from adiponitrile and acrylonitrile, respectively. Nitriles are also used as solvents, extractants, pharmaceuticals, feedstock, drug intermediates (chiral synthons), and pesticides, as well as in the organic synthesis of amines, amides, amidines, esters, carboxylic acids, aldehydes, ketones and heterocyclic compounds.

In the current study of particular interest is the fact that nitriles are important sources (as precursors) of carboxyamides and carboxylic acids via hydrolysis (Figure 1.1).

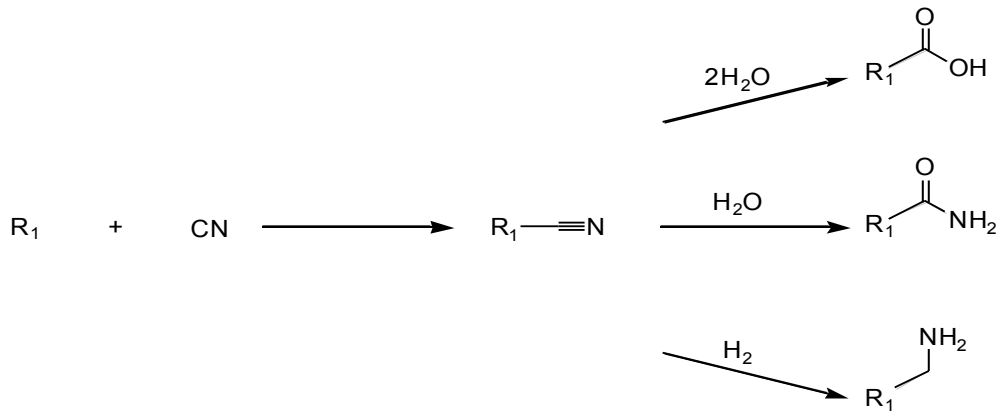


Figure 1.1: Schematic addition of cyanide to a molecule providing an additional carbon and nitrogen for subsequent conversion to other functional groups.

Nitriles can be hydrolysed to carboxylic acids, hydrolysed to amides, reduced to primary amines or selectively converted by regio-specific catalysts.

Traditionally, acid or alkaline hydrolysis of nitriles is used to form the corresponding acid or amide (Figure 1.2).

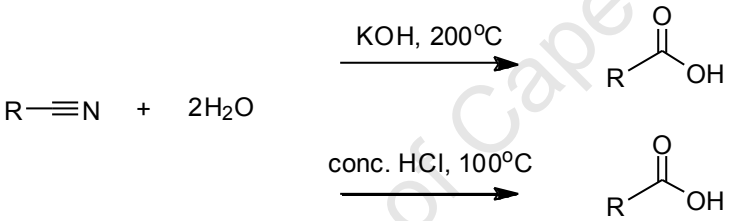


Figure 1.2: Alkaline and acid hydrolysis of nitriles.

However, stopping these reactions at the amide is not necessarily simple and the extreme conditions are not compatible with retention of acid or base labile functional groups (such as esters). Moreover, enantioselective hydrolysis is generally not possible. Although chemical hydration under neutral conditions is possible, it involves expensive reagents such as platinum, palladium or cobalt complexes and high reaction temperatures (Rey *et al.* 2004).

An alternative is to use biocatalysis. Biocatalysts can function at mild reaction conditions, physiological pH and temperature. They also demonstrate chemo-, regio- and stereo-selectivity, with their specificity circumventing the loss of labile groups that can occur with acid or base nitrile hydrolysis reactions.

Biologically we have the advantage of having two distinct enzyme systems, one of which converts nitriles directly to the carboxylic acid, the other providing the intermediate amide

(Figure 1.3). Among aerobic soil bacteria, *Rhodococcus*, *Bacillus* and *Pseudomonas* species are the most common culturable that can hydrolyse nitriles (Brady *et al.* 2006a).

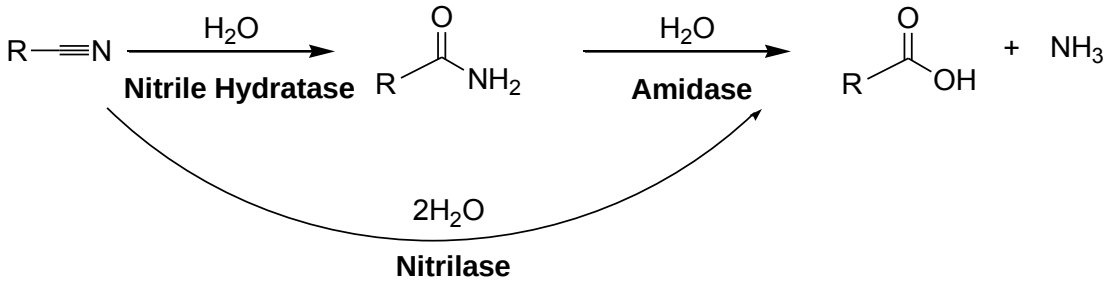


Figure 1.3: Nitrile hydrolysis catalysed by direct nitrilase conversion to the corresponding acid, or via a nitrile hydratase / amidase two-enzyme reaction.

Chemical transformations of nitriles require relatively severe conditions (elevated temperatures and pressures) and often generate toxic by-products (Crosby *et al.* 1994; Nagasawa and Yamada 1990; Wieser and Nagasawa 2000). Nitrilase and nitrile hydratase enzymes are, however, able to convert a wide range of nitriles, under milder conditions. For example, applications of nitrile-hydrolysing enzymes include the production of amides and acids such as acrylamide, nicotinic acid and lactic acid from their corresponding nitriles (Wyatt and Linton 1988) and the conversion of α–aminonitriles to optically active amino acids (Bhalla *et al.* 1992). Production of acrylamide from acrylonitrile is an ongoing, commercially successful, biotransformation using nitrile hydratase (Kobayashi *et al.* 1992a), as is the production of nicotinic acid from 3-cyanopyridine using nitrilase (Kobayashi and Shimizu 1994). Enantioselective hydrolysis of α-hydroxynitriles (cyanohydrins) (Faber 1995) and the production of optically active 2-arylpropionic acids, ibuprofen (Yamamoto *et al.* 1990) and naproxen (Effenberger and Böhme 1994; Layh *et al.* 1994), have also been achieved using nitrile-converting biocatalysts. Biosynthesis of the plant hormone indole-3-acetic acid from indole-3-acetonitrile is another useful nitrile-converting activity that has found a footing in industry (Kobayashi *et al.* 1993a).

In particular Rhodococci have been generally found to be excellent nitrile biocatalysts. *Rhodococcus* sp. are used in the manufacture of acrylamide (Kobayashi *et al.* 1992a), nicotinamide, and nicotinic acid (Kobayashi and Shimizu 1994).. *Rhodococcus* sp. R312 and

Rhodococcus erythropolis NCIMB 11540 whole cells were used to transform cyclic β-amino nitriles to cyclic β-amino amides/acids (Preiml *et al.* 2004), which have important pharmacological functions due to their antibiotic (Czernecki *et al.* 1997; Davies *et al.* 1999; Garner and Ramakanth 1986), antifungal (Davies *et al.* 1994; Theil and Ballschuh 1996) and cytotoxic (Sone *et al.* 1993) properties. *Rhodococcus equi* A4, now classified as *Rhodococcus erythropolis* (Vejvoda *et al.* 2007), is an efficient biocatalyst for chemoselective and diastereoselective transformations (Martínková and Kren 2002).

The remediation of highly toxic nitrile-containing wastes and the metabolism of nitrile-containing herbicides have proven useful in environmental management (Shaw *et al.* 2003). Synthetic fibre production results in large amounts of propionitrile as waste, which was successfully converted by cells of *Rhodococcus erythropolis* to ammonium propionate, which is useful as a feed supplement (Christian *et al.* 2001). Given the range of chemistries afforded by nitrile conversion, and the diverse products derived there from, the exploration of nitrile biocatalysis in microbial strains was pursued.

1.1.2. Whole cell profiling and identifying nitrile converting candidates

A selection of nitriles from aliphatic to aromatic, and ranging in bulk size and chain length, was chosen as a representative nitrile library to test the conversion abilities of various isolates from industrial soil samples. The *Rhodococcus* strains proved to be the most competent nitrile-converting strains. Five of the most promising rhodococcal strains from the culture collection, which grew well on nitrile substrates, were further tested and rated by relative growth on agar substituted with various nitriles as the sole nitrogen source (Table 1.1) (Brady *et al.* 2004b). In particular, *Rhodococcus* 4 (Table 1.1) had a diverse nitrile tolerance, and was observed to grow on both aliphatic and aromatic nitriles.

They were then tested for nitrile conversion along with some commercial nitrilases and whole cell biocatalysts and measured by HPLC to test more specific nitrile transformations. The relative activities against different nitriles are summarised in Table 1.2 (Brady *et al.*

2004b). When *Rhodococcus* 4 was cultured using various nitriles as the sole nitrogen source, it was found to grow on 13 of the 17 growth substrates. Of these substrates, the organism showed good or exceptionally good growth on at least 11 compounds. Its ability to utilise benzamide, adipamide and acetoxy-phenylacetamide as sole nitrogen sources suggested it contained amidase activity, while nitrilase and/or nitrile hydratase activity was suggested by its ability to convert 100% of the benzonitrile, chlorobenzonitrile and naproxen nitrile. *Rhodococcus* 4 also displayed fair activity towards phenylglycinonitrile, isocyanomethane, 3-hydroxy-3-phenylpropionitrile and benzonitrile.

The conversion of both aliphatic and aromatic nitriles by *Rhodococcus* 4 (initially designated soil sample isolate (I) morphology: small, or "Ismall") suggested a favourably versatile biocatalytic candidate. Also, the conversion of naproxen nitrile was an encouraging lead for potential manufacture of pharmaceutically important drugs using green chemistry; a product of conversion is the precursor to Naproxen, an aromatic propionic acid derivative in the class of aryl acetic acid group of non-steroidal anti-inflammatory drugs, or NSAIDs. An obviously worthwhile candidate for further exploration of biotechnology applications, *Rhodococcus* 4 was subsequently submitted to the American Type Culture Collection and later designated the title *Rhodococcus rhodochrous* ATCC BAA-870. The organism was subsequently studied further to evaluate its potential synthetic applications, in particular its capacity to enantioselectively hydrolyse racemic compounds.

Table 1.1: Relative growth of some *R. rhodochrous* strains on different sole nitrogen sources

	Substrate																
	Acrylonitrile	Adipamide	Adiponitrile	Acetoxyphenyl-propionitrile	Acetoxyphenyl-acetamide	α-Methylbenzyl-cyanide	Benzylidene malononitrile	Benzonitrile	Benzamide	4-Cyanopyridine	Diamino-malenonitrile	Fumaronitrile	Isobutyronitrile	Malononitrile	Phenylglycinonitrile	Propionitrile	Propionamide
Biocatalyst																	
Rhodococcus 1	4	4	4	1	3	1	0	1	5	1	1	1	4	2	2	4	5
Rhodococcus 2	4	4	3	4	2	0	0	0	4	0	1	0	4	2	1	3	5
Rhodococcus 3	4	4	3	5	0	3	0	2	4	2	0	0	5	3	5	3	5
Rhodococcus 4 (ATCC BAA-870)	5	5	4	3	4	0	0	1	5	1	0	0	5	4	4	5	5
Rhodococcus 5	0	0	4	0	0	0	0	0	0	0	0	0	0	0	4	4	0

Growth of the organism on each substrate is designated by 5 – exceptional, 4 – strong, 3 – good, 2 – fair, 1 – poor, and 0 – no growth. Screening of organisms was performed on minimal medium agar plates containing 5 mM of each substrate as the sole nitrogen source. Plates were incubated at 30 °C and the growth observed. Table adapted from results obtained by Brady *et al.* (Brady *et al.* 2006a).

Table 1.2: Percentage substrate conversion of selected biocatalysts

	Nitrile Substrate															
	Acetoxy-phenylacetonitrile	Benzylnitrile	α-Methylbenzyl-cyanide	2-Phenyl-glycinonitrile	Mandelonitrile	3-Hydroxy-3-phenylpropionitrile	2-Phenyl-butyronitrile	Benzonitrile	N-Phenyl-glycinonitrile	Isocyanomethane	Chlorobenzonitrile	p-Tolunitrile	3,4-Dimethoxy-benzonitrile	3-Phenyl-propionitrile	Naproxen nitrile	Phenylglycinonitrile
Biocatalyst	% Conversion															
Rhodococcus 4 (ATCC BAA-870)	9	14	1	1	2	24	3	100	5	33	100	5	5	6	100	18
Rhodococcus ABFGs	ND	22	0	0	1	2	0	ND	30	0	ND	16	3	ND	ND	ND
Rhodococcus MAWA (C6)	ND	100	46	11	6	76	3	ND	2	32	ND	100	44	ND	ND	ND
Rhodococcus NOVO SP361	ND	102	15	4	2	30	11	ND	41	19	ND	32	24	ND	ND	ND
Pseudomonas alcaligenes 1	0	ND	ND	ND	ND	ND	ND	0	ND	ND	0	ND	ND	ND	0	7
Pseudomonas fluorescens Nitrilase	ND	97	21	89	64	0	1	ND	0	14	ND	0	0	ND	ND	ND
Microbacterium 1	0	ND	ND	ND	ND	ND	ND	0	ND	ND	0	ND	ND	ND	40	9
Alcaligenes faecalis 1	0	ND	ND	ND	ND	ND	ND	0	ND	ND	0	ND	ND	ND	2	12

Conversion of each substrate is a percentage conversion. ND – not determined. Reactions were analysed by HPLC using a Phenomenex Luna C18 column (from Separations) and a Hewlett Packard Series 1100 Chromatograph with a Photo Diode Array detector (PDA). Gradient elution of compounds was done according to the method developed by N. Wilde (2001, personal communication) and modified by G. Kupi (2003, personal communication), using a 1 ml/min flow rate, 5 µl injection volume, and 20 minute run time. Table reproduced from results obtained by Brady and colleagues (Brady *et al.* 2004b) and N. Dube (in house research, not published).

1.1.3. Testing enantioselectivity of the organism

1.1.3.1. Naproxen nitrile conversion

Discovering efficient and practical routes to the synthesis of pharmaceutically active compounds is an increasingly important aspect of biocatalysis. Major factors in the environmentally-friendly synthetic process of drugs, or the intermediates thereof, are reduced waste production and solvent volume with an increase in yield and purity. Cost of manufacture is reduced as the number of steps required for the synthetic route is reduced. For example, it is difficult to chemically synthesise naproxen in an efficient and environmentally friendly manner due to the multiple chemical steps involved and resulting poor yield (Steenkamp and Brady 2003). Also, since naproxen is the optically active (S) form of 2-(6-Methoxy-naphthyl)-propanoic acid, it must be marketed as a composition free of the (R) stereoisomer (Brady *et al.* 2004a; Steenkamp and Brady 2008). Certain efficient and effective chemical routes of synthesis for naproxen have been proposed, but to date have not been used on a commercial scale owing to the high costs involved in producing the precursor, naproxen nitrile. An efficient, cost-effective, environmentally-friendly process for making 2-(6-Methoxy-naphthyl)-propanoic acid is therefore needed.

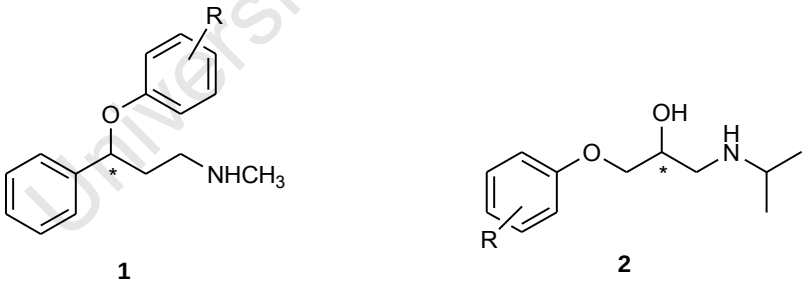
Rhodococcus butanica (ATCC 21197) which exhibited nitrile hydratase/amidase activity was used by Effenberger *et al.* (1994) to hydrolyse enantioselectively racemic naproxen nitrile (R/S)-naproxen to (S)-naproxen in enantiomeric excess (ee) (Effenberger and Böhme 1994). However, racemic naproxen amide was not a good substrate for this strain. Another rhodococcal species (sp. C311) was used by the same group to hydrolyse enantioselectively racemic (R/S)-naproxen nitrile and (R/S)-naproxen amide to produce (S)-naproxen in 99% enantiomeric excess with good chemical yields in aqueous medium as well as in a biphasic hexane/phosphate buffer system (Effenberger and Böhme 1994). Whole cells of the rhodococcal strains MP50 and C311 have been used to convert naproxen nitrile to naproxen *via* the amide (Layh *et al.* 1994). Strain C311 was also shown by Layh *et al.* (1994) to convert racemic naproxen nitrile to (S)-naproxen amide, and subsequently to (S)-naproxen, while racemic naproxen amide was hydrolysed to (S)-naproxen. The MP50 strain predominantly formed (R)-naproxen amide from racemic naproxen nitrile, while (S)-naproxen was formed

from racemic naproxen amide. Both strains converted the racemic naproxen amide to (S)-naproxen with an enantiomeric excess of 99% (Layh *et al.* 1994).

The hydrolysis of naproxen nitrile by *R. rhodochrous* ATCC-BAA870 (Table 1.2) would therefore provide a potential candidate as a tool in green pharmaceutical chemistry but also highlighted the possibility that the strain harboured other useful enantioselective capabilities (Brady *et al.* 2004a).

1.1.3.2. Enantioselective hydrolysis of beta-hydroxy nitriles

β-Adrenergic pharmaceutical compounds, or beta-blockers, (Figure 1.4) comprise a group of drugs used for the treatment of cardiovascular disorders such as hypertension, cardiac arrhythmia, or ischemic heart disease and psychiatric disorders. Each of these drugs possesses at least one chiral centre, and an inherent high degree of enantioselectivity in binding to the β-adrenergic receptor. Due to the difficulty of separation, or the lack of serious known side-effects attributed to the other enantiomer, most β-adrenergic blocking agents are currently marketed as racemates. However, this implies that half of the compound is inactive "enantiomeric ballast".



1 2

Figure 1.4: General structure of β-adrenergic blocking agents.

A common chemical structural feature of beta blockers includes at least one aromatic ring attached to the alkyl side chain with a secondary hydroxyl and amine functional group. The asterisk denotes the stereogenic centre. Reproduced from Kinfe *et al.* (Kinfe *et al.* 2009).

Exploring the preparation of optically active precursors for β-adrenergic blocking agents has become an attractive field, and the resolution of β-hydroxy nitriles for the stereoselective synthesis of β-adrenergic blocking agents has been increasingly researched (Hammond *et al.* 2007; Kamal and Ramesh Khanna 2001; Kamal *et al.* 2002; Kamal *et al.* 2005b; Kamal *et al.*

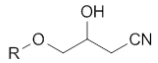
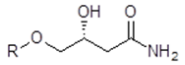
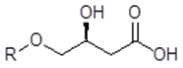
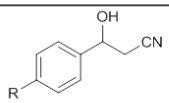
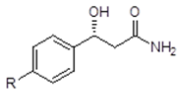
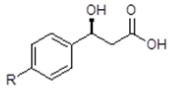
2005a; Wünsche *et al.* 1996). The chemistry of α-hydroxy nitriles has been explored, but there are relatively few reports detailing the asymmetric hydrolysis of β-hydroxy nitriles using microbial or enzymatic transformations, apart from that of 3-substituted glutaronitrile derivatives (Crosby *et al.* 1992; Kakeya *et al.* 1991; Maddrell *et al.* 1996).

Li and Wu reported the hydrolysis of 3-hydroxy-4-aryloxybutanenitriles using *Rhodococcus* sp. CGMCC 0497 (Wu and Li 2003), demonstrating that β-hydroxy nitriles could be hydrolysed to their corresponding amides and acids. However, unlike the hydrolysis of α-hydroxy nitriles, they did not obtain an excellent ee for hydrolysis of β-hydroxy nitriles. With the limited number of efficient approaches for the resolution of β-hydroxy nitriles using nitrile degrading enzymes, we felt there was scope for exploring alternative biocatalysts which can provide excellent ee. Using *R. rhodochrous* ATCC BAA-870, the asymmetric hydrolysis of aromatic β-hydroxy nitriles could be achieved with excellent e.e. and moderate to high yields (Kinfe *et al.* 2009). The highest enantioselectivity was found in the conversions of 4-(4-methoxyphenoxy)-3-hydroxybutanenitrile and 4-(benzyloxy)-3-hydroxybutanenitrile to their respective amides, with an ee of more than 99% (Table 1.3).

From the results obtained, it was deduced that the nitrile-metabolising enzyme system of *R. rhodochrous* ATCC BAA-870 consists of an (*R*)-specific nitrile hydratase and a (*S*)-specific amidase. The enantioselectivity depended on the nature of the substituent attached to the phenyl ring.

This was the first report to demonstrate hydrolysis of β-hydroxy nitriles using nitrile degrading enzymes with excellent enantiomeric excess. The work also highlighted a possible synthetic route to enantiomerically pure β-adrenergic agents such as the chiral intermediate used in the production of the cholesterol lowering drug Lipitor® (Atorvastatin calcium). *R. rhodochrous* ATCC BAA-870 is, therefore, a potentially useful tool in the synthesis of enantiomerically pure β-adrenergic agents.

Table 1.3. Resolution of β-hydroxy nitriles by *R. rhodochrous* ATCC BAA-870

Substrate	R	Amide-**a**	Acid-**b**
		ee (%)	ee (%)
	 **3 - 6**	 **3 - 6a**	 **3 - 6b**
3	Ph	>99	57
4	Bn	>99	60
5	*p*-OMe Ph	96	70
6	*p*-Cl Ph	82	72
	 **7** R = H **8** R = Me	 **7a** R = H **8a** R = Me	 **7b** R = H **8b** R = Me
7	H	51	51
8	Me	65	78

Incubation of various 3-hydroxy-4-aryloxybutanenitriles provided efficient and selective access to (R)-amide and (S)-Acid except in the case of the benzyloxy substrate 4, where (S)-amide and (R)-acid was obtained. In all cases complete conversion of the starting nitrile was achieved and the percentage ee and the products were identified after esterification using chiral HPLC and ^{1}H and ^{13}C NMR spectroscopy, respectively, after chromatographic separation.

The resolution of β-hydroxy nitriles to enantiomeric amides and acids also raises the possibility of production of non-proteinogenic amino acids, which are precursors for a range of biologically active compounds (Eichhorn *et al.* 1997). Viral protease inhibition strategies include substrate-based inhibitors such as those based on peptide-like products/peptidomimetics or non-cleavable substrates such as the novel non-covalent phenylethyl amide inhibitor of HCV NS3 protease (Colarusso *et al.* 2003).

1.1.3.3. Enantioselective hydrolysis of beta-amino nitriles

Beta-amino amides and acids have many applications in pharmaceutical industries and drug discovery. For example, beta amino acids are used in the synthesis of several drugs such as the antifungal antibiotic Cispentacin (Davies *et al.* 1994), antitumour drug Taxol (Holton *et al.* 1994), antidiabetic drug Sitagliptin (Steinhuebel *et al.* 2009), and other biologically active

peptides and small molecules (Hsiao *et al.* 2004). They may also be used in the manufacture of peptidomimetics for viral protease inhibitors (Aguilar *et al.* 2007; Colarusso *et al.* 2003; Gademann *et al.* 1999; Steer *et al.* 2002). Recently, the enantiomeric resolution of β-amino amides and acids by *R. rhodochrous* ATCC BAA-870 was achieved through relatively mild enzymatic hydrolysis, and offers a green biocatalytic method for the production of racemic β-amino acids (Chhiba *et al.* 2012).

Chhiba *et al.* used *R. rhodochrous* ATCC BAA-870 to achieve the asymmetric hydrolysis of unprotected β-aminonitriles to the corresponding amides (Figure 1.5 and Figure 1.6) with ee values of product and substrate of over 50% (Chhiba *et al.* 2012). The nitrile hydratase showed enantioselectivity towards substrates with substitutions on the β-aromatic moiety.

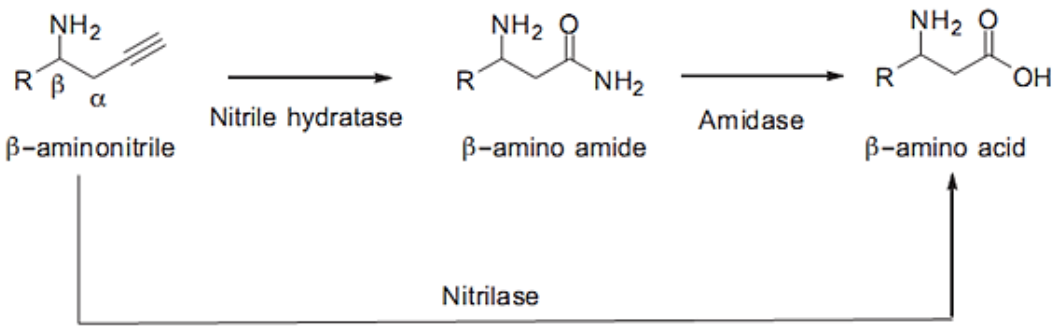


Figure 1.5: General scheme for the biocatalytic conversion of β-aminonitriles to β-amino amides and acids.

Scheme reproduced from Chhiba *et al.* (Chhiba *et al.* 2012).

A range of β-aminonitriles (3-amino-3-phenylpropanenitrile and derivatives) were hydrolysed to the corresponding amides using BAA-870 as biocatalyst. Results showed that the nitrile hydratase enzyme was in particular enantioselective for 3-amino-3-*p*-tolylpropanenitrile and 3-amino-3-(4-methoxyphenyl)propanenitrile and the corresponding amides (up to 85% in one case).

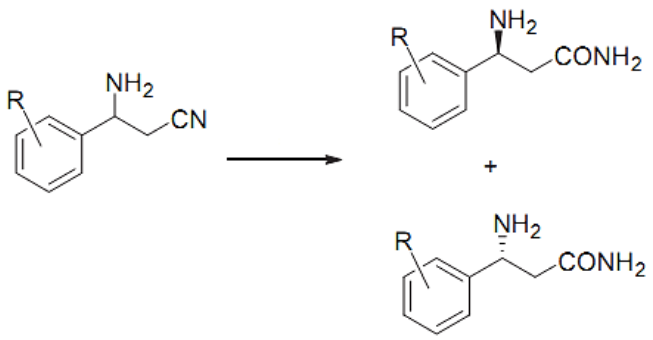


Figure 1.6. Synthesis of β-amino amides by *R. rhodochrous* ATCC BAA-870.

Scheme reproduced from Chhiba *et al.* (Chhiba *et al.* 2012).

The enantioselective hydrolysis of the nitrile group of β-amino substituted nitrile compounds was achieved with the *N*-protected β-substituted amino derivatives. *N*-tosyl and *N*-Boc protected compounds were not hydrolysed, possibly because these groups were too large to be accommodated in the nitrile hydratase (NHase) active site. The smaller *N*-acetyl compounds were hydrolysed, but the unprotected compounds were not, which did not support the active site size theory. It was surmised by Chhiba *et al.* that the lack of hydrolysis of unprotected compounds was due rather to protonation of the amino group at a pH less than 9, since the reaction proceeded when performed at pH 9 (Chhiba *et al.* 2012). The biocatalytic hydrolysis of β-amino nitriles to β-amino amides by BAA-870 lends further support to the use of the organism as a promising candidate in a green manufacturing pharmaceutical chemistry.

1.2. Targeting specific nitrile converters: Isolating the nitrilase and nitrile hydratase for biocatalysis

1.2.1. Nitrilase and Nitrile hydratase

1.2.1.1. Introduction

Hydration of nitriles in plants and microbes may be catalysed by one of two distinct pathways (Figure 1.3). Generally, aliphatic nitriles are converted to amides by nitrile hydratases and then hydrolysed by amidases to carboxylic acids. Aromatic nitriles, on the other hand, are metabolised to the corresponding acid in a single-step reaction by nitrilases. There are many exceptions to this including, for example, the nitrile hydratases of a *Rhodococcus* sp. (Hjort *et al.* 1990) and *Bacillus smithii* SCJ05-1 (Takashima 1995), and the nitrilase of *R. rhodochrous* K22 (Kobayashi *et al.* 1990b), which are capable of hydrolysing both aliphatic and aromatic nitriles.

Nitrilases and nitrile hydratases are attracting increasing attention in the field of biotransformation and biocatalysis. Their ability to produce a carboxylic acid from a corresponding nitrile, with nitrile hydratase producing an amide as an intermediate, has provided great potential for the use of these enzymes as nitrile catalysts functioning under relatively mild reaction conditions. One of the most well-known commercial examples of nitrile bioconversion by a nitrile hydratase is that of the manufacture of acrylamide from acrylonitrile (Kobayashi *et al.* 1992a). This application for the synthesis of acrylamide started in 1985, with Mitsubishi Rayon first producing over 30,000 tons of acrylamide per year (Faber 1995, 1997). Acrylamide bioproduction now exceeds 100,000 tons per year (Liese *et al.* 2000), with the worldwide biocatalytic production of acrylamide estimated to be more than 400 000 tons per annum (Zheng *et al.* 2010). This was the first bioconversion process to be successfully applied to the petrochemical industry and implemented for the industrial production of a commodity chemical. Most nitrile bioconversions are still achieved through whole-cell treatment of substrates using industrially useful microbial strains such as *R.*

rhodochrous. An acrylamide producing plant has recently been commissioned in South Africa by Senmin (Pty) Ltd, using biocatalysis technology developed by the world's leading chemical company BASF, in partnership with Ciba Speciality Chemicals Inc. (Milmo 2000). Production of acrylamide using a whole cell *Rhodococcus* at this plant is estimated at 20 000 tons.

Although several nitrile-hydrolysing enzymes have been isolated from natural sources, these enzymes were generally too unstable for commercial use in the isolated state, and most applications of these enzymes has consequently been as whole-cell biocatalysts. However, applying whole cell biocatalysts can present disadvantages in that some small aliphatic nitriles, and those with hydroxy and amino groups, can be utilized by the organism itself as carbon sources (Meth-Cohn and Wang 1997). Also, the decreased yield of amides obtained when whole cell systems are used can be avoided when using purified enzymes since they are not utilized by the organism. Despite the large number of nitrilases and nitrile hydratases that have been purified there is a scarcity of suitable and/or well-characterised commercial nitrile-converting biocatalysts.

Rhodococcal nitrilases and nitrile hydratases are the best characterised of the nitrile-degrading enzymes. Although many Pseudomonal nitrile-metabolizing enzymes have been studied, they generally have narrower substrate ranges than the rhodococcal enzymes (Dhillon *et al.* 1999; Layh *et al.* 1998).

Nitrile converting enzymes can be found in various *Rhodococcus* strains isolated from geographically diverse locations, including shallow marine sediments (Langdahl *et al.* 1996), deep-sea sediments (Heald *et al.* 2001) and various soil types, including garden soil (Kato *et al.* 1998; Layh *et al.* 1994; Layh *et al.* 1997), subtropical rain forest soil (Brandão *et al.* 2002), mangrove mud (Brandão *et al.* 2002) and contaminated river bank soil and sludges (Blakey *et al.* 1995; Scholtz *et al.* 1987). The suitability of the organism to diverse habitats accounts for the numerous types of nitrile metabolizing enzymes found within the species, including mesophilic and thermophilic enzymes with wide pH and temperature profiles (Cowan *et al.* 1998; Cramp and Cowan 1999). The majority of rhodococcal nitrilases, nitrile hydratases and amidases referred to in literature are inducible enzymes (Cowan *et al.* 1998; Kashiwagi *et al.*

2004; Kato *et al.* 1999; Kaul *et al.* 2004; Kobayashi *et al.* 1992c; Kobayashi *et al.* 1992b). Nitrilases and nitrile hydratases are functionally and structurally distinct enzymes.

1.2.2. Nitrilase

Nitrilase enzymes (E.C. 3.5.5.1) catalyse the direct hydrolysis of aliphatic and aromatic nitriles into the corresponding carboxylic acid and ammonia (Equation 1).

$$R\text{-}CN + 2H_2O \rightarrow R\text{-}COOH + NH_3 \qquad \text{Equation 1}$$

Nitrilases are widely distributed in plants, where they catalyse the conversion of indole-3-acetonitrile into the plant hormone indole-3-acetic acid. They are also found in fungi and bacteria that can metabolise nitrile-containing herbicides (Kaplan *et al.* 2011a; Parekh *et al.* 1994; Stalker *et al.* 1988; Veselá *et al.* 2010).

The enzymatic synthesis of enantiopure carboxylic acids using nitrilases reached a major milestone when Yamamato and co-workers produced (R)-mandelic acid from mandelonitrile using *Alcaligenes faecalis* whole cells (Yamamoto *et al.* 1991). However, due to the limited availability and stability of nitrilase enzymes, work on enantiopure compounds did not progress much during the 1990s. Since then, nitrilases have been used to stereoselectively synthesize amino- and hydro-carboxylic acids in high enantiomeric excess (Schmid *et al.* 2001). Enzymatic conversions of substrates to chiral α-hydroxy- or α-aminocarboxylic acids have been successfully achieved at BASF (Germany), and (R)-mandelic acid is now manufactured at the multi-ton scale. Applications of nitrilases have advanced in recent years due to ongoing studies detailing their structural (O'Reilly and Turner 2003; Thuku *et al.* 2009) and catalytic properties (Martínková *et al.* 2008; Martínková and Kren 2010).

1.2.3. Nitrile hydratase

Nitrile hydratases (NHases) catalyse the direct hydrolytic conversion of aromatic and aliphatic nitriles into the corresponding amide (Equation 2).

$$R\text{-}CN + 2H_2O \rightarrow R\text{-}CONH_2 \qquad \text{Equation 2}$$

The expression of nitrile hydratase and amidase enzymes are often associated in the same organism, and the structural genes can be found to be adjacent on the same operon under the control of the same activator protein (Lu *et al.* 2003). Amidases, associated with nitrile hydratases, are being increasingly used in the biotechnology arena for production and/or conversion of valuable intermediates. For example, immobilized penicillin acylase (amidase) (E.C. 3.5.1.11) is widely used in the production of 6-aminopenicillanic acid from penicillin (Rolf and Lim 1985). *Klebsiella oxytoca* contains an (R)-specific amidase, and converts (R,S)-amide to (R)-acid while the (S)-amide remains unconverted (Shaw *et al.* 2002). This enantiomeric reaction has been developed for large-scale biotransformation by Lonza (Shaw *et al.* 2003). Amidases are used by the Dutch chemical company DSM to produce non-proteinogenic L-amino acids by kinetic resolution of racemic amino-acid amides. These L-amino acids are building blocks for synthesis of pharmaceuticals. Several hundred tons of these products are used in the production of a broad spectrum of L- and D-amino acids.

1.2.4. Nitrilase and nitrile hydratase in industry

Due to the hydrophobic nature of most nitriles, organic solvents are usually added to the aqueous reaction medium during nitrile-conversion by nitrilases and serves to enhance the bioavailability of the substrate. Moreover, some solvent systems significantly enhance the enantioselectivity of certain enzymes. For example, hexane, acetone and α-cyclodextrin improved the enantioselectivity of the asymmetrization of 3-arylglutaronitriles (Nagasawa *et al.* 1993). Unfortunately, many of these nitrile-hydrolysing enzymes cannot be used in the presence of organic solvents, especially at high concentrations, which hamper their

applicability to situations where the organonitrile substrate is very poorly soluble in water. Water-miscible solvents such as methanol and ethanol have been shown to be suitable co-solvents for purified enzyme only in concentrations of up to 20% (v/v), above which enzyme inactivation occurs. However, a purified nitrile hydratase from *Rhodococcus equi* displayed activity in the presence of high concentrations of water-immiscible hydrocarbons, for example up to 90% (v/v) isooctane (Prepechalová *et al.* 2001). Purified high molecular mass nitrile hydratase from *R. rhodochrous* J1 is unusually stable, retaining 89 – 100% of its original activity in alcohols, and showing appreciable activities in 50% (v/v) ethylene glycol, dimethyl sulfoxide and acetone (Nagasawa *et al.* 1993). This enzyme may be unusually tolerant to organic solvents due to its complex quaternary structure, the holoenzyme being composed of 20 subunits (Nagasawa *et al.* 1993).

1.3. Building tool boxes for nitrile biocatalysis

The organism *R. rhodochrous* BAA-870 is a potentially useful nitrile biocatalyst. However, initial results suggest it contains at least two different nitrile hydrolyzing enzyme systems consisting of nitrilase and nitrile hydratase/amidase (Brady *et al.* 2004b; Chhiba *et al.* 2012; Kinfe *et al.* 2009). We do not know if there are multiple isoenzymes or what the copy number is. Moreover, due to poor mole balances observed in prior experiments (not shown) there may be enzymes that are responsible for the further metabolism of the amide and acid products of nitrile enzymatic hydrolysis (Meth-Cohn and Wang 1997). Hence it would be advisable to know more about the overall genome of the organism. Furthermore, if the nitrilase and nitrile hydratase are to be of general use as biocatalysts, it would be preferable to isolate the responsible genes and express the enzymes independently for further characterization studies. This would also facilitate structural and mechanistic studies, as well as permit future modification of the enzymes at the genetic level.

1.4. Objectives

1. To determine the *R. rhodochrous* BAA-870 genome sequence, annotate it, identify the genes encoding nitrile hydrolyzing enzymes, and possibly identify the genes for enzymes that may of value industrial biocatalysis.
2. To isolate the gene for nitrilase activity, express it heterologously, and characterize the enzyme.
3. To isolate the gene for nitrile hydratase activity, express it heterologously, and characterize the enzyme.

Chapter Two

2. Sequencing the genome of *Rhodococcus rhodochrous* ATCC BAA-870

2.1. Introduction

2.1.1. Rhodococcus rhodochrous

2.1.1.1. The Rhodoccocus genus

Rhodococcus is a bacterial genus that can be isolated from range of sources including soil, marine sediments, rocks, groundwater, boreholes, animal dung, insect guts, and healthy and diseased plants and animals (Goodfellow 1989a; Ivshina *et al.* 1994). Some interesting sources from which *Rhodococcus* have been isolated include a high-level nuclear waste plume (Fredrickson *et al.* 2004) and even a medieval grave (Takeuchi *et al.* 2002). The *Rhodococcus* genus is a gram positive, aerobic, non-motile, mycolate-containing, nocardioform actinomycetes (Goodfellow 1989b). Nocardioform refers to the mycelial growth-like characteristic; and fragmentation of the growing bacterial cells into coccoid and rod-shaped elements (Lechavalier 1989).

Rhodococcus exhibits a large metabolic diversity and as a result can degrade a variety of pollutants which affords environmental and industrial biotechnologies a useful tool in transformation and degradation chemistries. Biotechnology using *Rhodococcus* has been used in the manufacture of, for example, surfactants, polymers, amides and flocculants (Bell *et al.* 1998). The commercial potential of these biotechnologies is reflected in the number of increasing patents involving *Rhodococcus* (see Appendix 7.1) for some results of recent patent searches). *Rhodococcus* has useful phenotypic traits, and the commercial value of the organism in technology has increased due to improved cloning and genetic manipulation technology (de Carvalho and da Fonseca 2005; Rzeznicka *et al.* 2010; Schumacher and Fakoussa 1999a; Thomas *et al.* 2002).

2.1.1.2. Rhodococcus classification according to numerical, chemo- and molecular taxonomy

The actinomycetales order (Figure 2.1) includes some of the world's most important organisms: *Mycobacterium tuberculosis* is responsible for the largest number of human deaths by bacterial infection, and the streptomycetes produce most of the antibiotics used today. *Rhodococcus* is arguably the most industrially important actinomycetes genus (Van der Geize and Dijkhuizen 2004) owing to its success as a microbial biocatalyst (Banerjee *et al.* 2002) and applications in a wide range of uses, from bioactive steroid production (Yam *et al.* 2010), to fossil fuel desulphurization (Gray *et al.* 1996), to the production of kilotons of commodity chemicals (Kobayashi *et al.* 1992a).

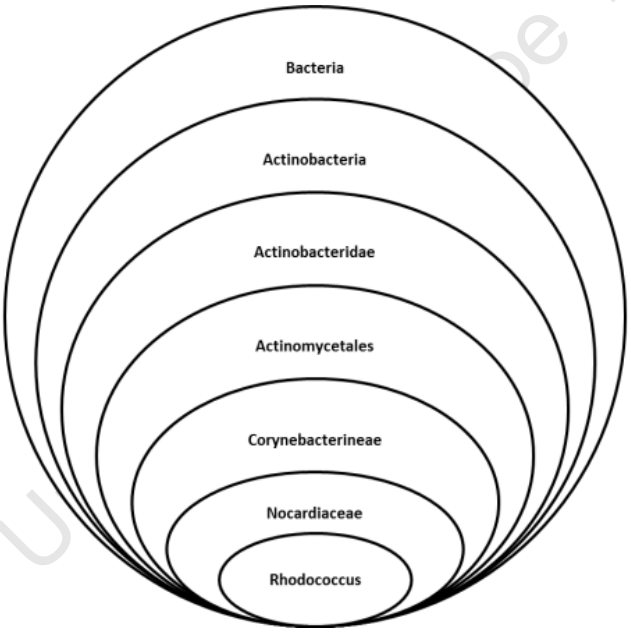


Figure 2.1: The full lineage of *Rhodococcus.*

A nocardioform actinomycetes bacteria, *Rhodococcus* includes cellular organisms; Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae; Actinomycetales; Corynebacterineae; Nocardiaceae; *Rhodococcus*. Information was obtained from NCBI Genome Data Bank.

The synonyms commonly used for the strain are many, and include *Rhodococcus roseus*, *Bacterium mycoides roseum, Staphylococcus rhodochrous* and *Gordona* (sic) *rosea* (Table 2.1). The name 'J1' is listed as a common misnomer as it was often used loosely to describe any *Rhodococcus*, which likely came about as a result of the extensive reference to *R.*

rhodochrous J1, largely seen as the most well studied model *Rhodochrous* species (Kobayashi *et al.* 1989; Kobayashi *et al.* 1991; Kobayashi *et al.* 1992c; Kobayashi *et al.* 1993b; Komeda *et al.* 1996b; Nagasawa *et al.* 1988; Nagasawa *et al.* 1993).

Table 2.1: *R. rhodochrous* taxonomy and historical synonyms and misnomers associated with the species

***R. rhodochrous* Taxonomy**	
synonym:	*Rhodococcus* roseus
synonym:	Bacterium mycoides roseum
synonym:	Staphylococcus *rhodochrous*
synonym:	"Staphylococcus *rhodochrous*" Zopf 1891
synonym:	*Rhodococcus rhodochrous* (Zopf 1891) Tsukamura 1974
includes:	*Rhodococcus rhodochrous* isolate DY6
includes:	*Rhodococcus rhodochrous* isolate DY5
includes:	*Rhodococcus rhodochrous* str. J1
misnomer:	J1

There are currently 12 established *Rhodococcus* species: *R. equi*, *fascians*, *erythropolis*, *opacus*, *coprophilus*, *rhodnii*, *ruber*, *marinonascens*, *percolates*, *zopfii*, *globerulus* and *rhodochrous*. The following strains also belong to established *Rhodococcus* species: *Arthrobacterpicolinophilus*, *Corynebacterium hoagie*, *Nocardiarestricta* and *Nocardiacalcarea*. *Rhodococcus* is a phylogenetically coherent clade – members are exclusively mycolate-containing genera, and are the only genera to possess mycolic acid in the cell wall (Embley and Stackebrandt 1994). The genera Corynebacterium, Mycobacterium and Nocardia were coined the “CMN” group, but include the genera *Rhodococcus*, Gordona, Tsukamurella and Dietza (Ruimy *et al.* 1994).

The *Rhodococcus* genus was originally proposed on the basis of a study by Goodfellow and Alderson (1977) which examined 92 unit characters of test strain actinomycetes, and grouped them according to percentage similarity (Goodfellow and Alderson 1977). These characteristics included staining, morphology, resistance to various antibiotics, tolerance to inhibitory compounds, ability to grow under various conditions of pH and temperature,

nutritional tests and chemical analyses. Fluorogenic tests to identify specific enzyme activities was also eventually included in similar numerical studies (Goodfellow *et al.* 1990).

Rhodococci are also defined by chemical characteristics which were outlined by Goodfellow (1989) and later Finnerty (1992) (Finnerty 1992; Goodfellow 1989b). The cell walls are described as having a chemotype-IV: this chemotype is characterised by arabinose and galactose as major cell wall sugars and peptidoglycan with one diamino acid, namely *meso*-diaminopimelic acid (Lechavalier and Lechavalier 1970). Other "CMN" group members have a chemotype-IV cell wall, but *Rhodococcus* is separated from them on the basis of a combination of chemotaxonomic characteristics that have been described by Rainey and colleagues (Rainey *et al.* 1995). The key considerations are the presence of tuberculostearic acid, mycolic acids with a length of 34 – 64 carbon atoms, and the fact that dehydrogenated menaquinones with eight isoprenoid units form the major menaquinone type. However, the menaquinones lack the cyclic element which is a characteristic of *Nocardia*. Relative proportions of lipids, including mycolates and fatty acids can be a useful differentiation of species within the genus (Klatte *et al.* 1994).

The evaluation and organization of the *Rhodococcus* taxonomy has in more recent years been helped by analysis of the sequence of the 16S rRNA gene, and gives an indication of how closely related the organisms are by the degree of variation in sequence. Bacteria with similarities in their 16S rRNA sequences of 97% and above can be considered the same species (Stackebrandt and Goebel 1994). DNA-DNA re-association values can add strength to species identification – a similarity value of 70% is considered a minimum for species identity (Wayne *et al.* 1987). A study by Rainey *et al.* 1995, showed *Rhodococcus* to be a relatively divergent species with at least 5 distinct groups: one group contains *R. rhodochrous, ruber* and *coprophilus*, another contains *R. erythropolis, globerulus, opacus* and *marinonascens*, while *R. rhodnii, fascians* and *equi* each lie within their own distinct phylogenetic branch (Rainey *et al.* 1995). The study also concluded that *Nocardia* evolved from within *Rhodococcus* rather than alongside it. Ochi analysed sequences of ribosomal AT-L30 proteins and found good agreement with phylogenies obtained from 16S rDNA analyses, and found evidence that the genus *Nocardia* forms a clade within *Rhodococcus* (Ochi 1995).

2.1.2. Problems caused by Rhodococci

2.1.2.1. Disease and infection

Interestingly, rhodococcal diseases are evident in plants, animals and humans. Infection of humans by *Rhodococcus* is increasing, both in occurrence and recognition of infection (Finnerty 1992; McNeil and Brown 1994). *Rhodococcus* has been a recognised equine pathogen for over 70 years, with *Rhodococcus equi* infections presenting in horses, particularly foals, and other animals (Prescott 1991). *R. equi* infection in humans is possible, and has shown increased frequency as a human pathogen, especially with the rise in immunosuppressed individuals. *R. equi* infection is associated with high mortality rates. Avirulent *R. equi* strains are widespread in soil and sand in Japan (Takai *et al.* 1996).

Rhodococcus rhodnii exists as a symbiont in the gut of the insect *Rhodnius prolixus*, an important vector of Chagas' disease (American trypanosomiasis) – growth and development of the insect is supposedly dependent on the growth of *R. rhodnii* and is a suggested target for biological control of the disease (Beard *et al.* 1992; Ben-Yakhir 1987). *R. fascians* infection in plants causes growth abnormalities. The virulence of different strains varies considerably, and susceptibility of even very closely related plants differs (Eason *et al.* 1995). Infection by *R. fascians* has several unique features, and no other work has been done on a gram-positive organism infecting plants. It has economic significance since it can infect crop plants such as tobacco and peas (Eason *et al.* 1995; Stange *et al.* 1996; Vereecke *et al.* 1997).

2.1.2.2. Foam formation in activated sludge plants

An increasingly common problem in water treatment plants is the formation of thick surface foams during activated waste-water treatment – these foams cause operational problems (Blackall 1994; de los Reyes *et al.* 1997). These surface foams reduce oxygen transfer, cause higher effluent biochemical oxygen demand, lead to greater amounts of suspended solids including pathogens in effluents, and provide a medium for wind-assisted dispersal of pathogens (Goodfellow *et al.* 1997; Schuppler *et al.* 1995; Tipping 1995). The significance of different nocardioforms in foaming is varied, but *Rhodococci* have been found to be prevalent in some cases (Blackall 1994; Fougias and Forster 1994; Soddell and Seviour 1995).

2.2. The significance of *Rhodococcus* in biotechnology

The significance of *Rhodococcus* in biotechnology is evident from the range of applications in which it has been used, including biodegradation of pollutants and bioremediation, biosurfactants and bioflocculants, desulphurization of fossil fuels, and even oil prospecting. Rhodococci can degrade chemical pollutants ranging from simple hydrocarbons through to aromatic and chlorinated hydrocarbons, nitroaromatics and polycyclic aromatics (including polychlorinated biphenyls, or PCBs, which are recalcitrant to degradation, hazardous, and persistent pollutants in soil and groundwater (Häggblom *et al.* 1994)). Various PCBs have been shown to be degraded by Rhodococci, including *R. rhodochrous, R. erythropolis*, and *R. globerulus* and other unclassified *Rhodococcus* strains (Asturias and Timmis 1993; Boyle *et al.* 1992; Maeda *et al.* 1995; Seto *et al.* 1995). Other toxic pollutants degraded by Rhodococci include pesticides (Parekh *et al.* 1994), sulphonated azo dyes (Heiss *et al.* 1992), *s*-triazines (Mulbry 1994b) and *n*-methyl carbamates (Behki *et al.* 1994).

The presence of multiple homologs of catabolic genes in *Rhodococcus* suggests that they may provide a comprehensive biocatalytic substrate profile (Van der Geize and Dijkhuizen 2004). Rhodococci are well known to have the capacity to degrade hydrophobic compounds, including polychlorinated biphenyls (PCBs), that is mediated, in for example *Rhodococcus jostii* by a cytochrome P450 system consisting of 29 genes which is involved in catabolism of hydrocarbons (Grogan 2011; Rosłoniec *et al.* 2009; Xiong *et al.* 2011). Cytochrome P450 oxygenase is often found fused with a reductase, as in *Rhodococcus* sp. NCIMB 9784 (Li *et al.* 2009). Biphenyl/PCB degradative genes are found in multiple sites on the *Rhodococcus* sp. RHA1 genome, both on the chromosome as well as on linear plasmids (Van der Geize and Dijkhuizen 2004). *Rhodococcus jostii* RHA1 was found to show lignin-degrading activity, possibly based on the same oxidative capacity as that used to degrade biphenyl compounds (Bugg *et al.* 2011).

2.2.1. Applied industrial uses

2.2.1.1. Industrial synthesis and transformation

The commercial success of industrial bioproduction of acrylamide by *R. rhodochrous* J1 at a large scale, is still the most cited example of successful application of bacteria in industry (Yamada and Kobayashi 1996). Acrylamide is a commodity chemical used extensively in coagulators, soil conditioners, adhesives, paints and petroleum recovering agents. Rhodococcal nitrilases and nitrile hydratases can be used to produce acrylic acid, various high value amides, the vitamins nicotinamide and *p*-aminobenzoic acid, antimycobacterial agents isonicotinic acid hydrazide and pyrazinamide, and α-hydroxy acids (Kobayashi and Shimizu 1994; Yam *et al.* 2011; Yamada and Kobayashi 1996; Zhou *et al.* 2010), in high yields and high specificities.

The production of poly(3-hydroxyalkanoic) acids by *R. ruber* has been used in the production of bio-plastic (Pieper and Steinbuchel 1992). The co-polyester of 3-hydroxybutyrate and 3-hydroxyvalerate (3HB-co3HV) is produced commercially as a biodegradable plastic (BIOPOL) by Zeneca Bio Products. However, *R. ruber* can produce 3HB-3HV with a different proportion of monomers while growing on a cheaper substrate than that used for BIOPOL production (Lee 1996; Pieper and Steinbuchel 1992).

Gaseous alkenes have been transformed into epoxides by *Rhodococcus* (Woods and Murrell 1990) and the epoxides subsequently used in ferric electric liquid crystals (Kieslich 1991) while the medically valuable compound, sec-cedranol, has been produced through efficient biotransformation by *Rhodococcus* (Takigawa *et al.* 1993). *Rhodococcus* enzymes have been used for the efficient production of amino acids L-leucine and L-phenylalanine (Bhalla *et al.* 1992; Hummel *et al.* 1987; Hummel 1997). A variety of enzymes from *Rhodococcus* are successfully used in biotechnology; cholesterol oxidases from *Rhodococcus* could have applications in food industry or steroid drug production (Christodoulou *et al.* 1994; Finnerty 1992; Kreit *et al.* 1994), and carbonyl reductases give a range of compounds that can be used for pharmaceuticals and agrochemicals (Peters *et al.* 1993). Food industry applications

include using *R. fascians* in bioreactors to improve the flavour of bitter fruit juices (Iborra *et al.* 1994; Marwaha *et al.* 1994) by degrading limonin, a bitter tasting compound.

2.2.1.2. Hydrocarbon degradation and heavy metal remediation

A naturally occurring consortium of bacteria was removed from cyanobacterial mats found floating in oil-polluted Arabian Gulf waters, and used to inoculate oil-contaminated soil by Sorkhoh *et al.* (1995). It resulted in increased removal of oil from the sand, and *Rhodococci* were noted to be the predominant bacteria in the microbial population (Sorkhoh *et al.* 1995). The rate of pentachlorophenol (PCP) degradation was enhanced by inoculation of sandy soil with *R. chlororaphis* (now *Mycobacterium*) (Briglia *et al.* 1994; Miethling and Karlson 1996). When composted soil contaminated with crude oil was inoculated with *R. ruber*, increased counts and persistence of hydrocarbon-oxidizing bacteria was noted (Christofi *et al.* 1998).

R. erythropolis, a hydrocarbon-degrading strain, was introduced to artificially contaminated soil plots, and shown to result in increased hydrocarbon degrading bacterial populations and increased hydrocarbon degradation overall (Koronelli 1996). *R. percolates* was first isolated from a chlorophenol-fed bioreactor (Briglia *et al.* 1996) and could be used in a bioreactor to degrade toxic compounds in industrial effluents. A muconic acid producing *R. rhodochrous* strain was used by Warhurst *et al.* for the assimilation of benzoate from soil (Warhurst *et al.* 1994).

The possible bioaccumulation by *Rhodococcus* of heavy metal ions, including radioactive heavy metal ions, has been explored. Two *Rhodococcus* strains were shown to accumulate high levels of caesium ions to 5% and 9% of their dry biomass (Tomioka *et al.* 1994).

2.2.1.3. Surfactants

Rhodococcus species have been shown to have the ability to produce surfactants (Finnerty 1992). Biosurfactants are relevant to biodegradation. Cellular surfactants, such as the mycolic acids present in *rhodococcal* cell walls, cause adherence of *Rhodococci* to hydrophobic phases in a two-phase system (Neu 1996). Surfactants lower the interfacial

tension between phases, making it easier for hydrophobic compounds to enter the cell (Fiechter 1992), while extracellular surfactants disperse hydrophobic compounds which increases the surface area for microbial attack (Finnerty 1994). Some biosurfactants have been reported to be more effective and efficient than existing synthetic surfactants, and may be more biodegradable and less toxic (Finnerty 1994; Philp and Bell 1995). Biosurfactants may be produced at high enough levels by *Rhodococci* to allow the use of whole broths in some oil industry applications (Abu-Ruwaida *et al.* 1991). Surfactants from a *R. ruber* strain enhanced removal of oil from soil in laboratory soil washing tests (Christofi *et al.* 1998).

2.2.1.4. Fossil fuel desulphurization

Removal of organic sulphur from fuel is difficult, and it has been suggested that microbial desulphurisation of petroleum and coal could be a way to prevent sulphurous emission from combustion, and consequently, reduce the associated problem of acid rain, while at the same time increasing the value of the fuel (Gray *et al.* 1996; Kilbane 1989). The potential of *Rhodococcus* spp. in desulphurization of fuel has been researched by many groups in the USA and Japan, including Energy Biosystems Corporation (USA) which has focused on the genetics and enzymology of *Rhodococcus* desulphurization (Bozdemir *et al.* 1996; Denome *et al.* 1994; Izumi *et al.* 1994; Kayser *et al.* 1993; Wang and Krawiec 1994). One unique selectivity of *Rhodococcus*, in particular, makes it a significant and interesting microbe for the desulphurization of fuel, namely the ability to cleave carbon-sulphur bonds while leaving carbon-carbon bonds intact (for example in the model compound dibenzothiopene). This selectivity allows for desulphurization of fuel without affecting its calorific value.

2.2.1.5. Biosensors and other applications

There are various reports on using rhodococcal cells or rhodococcal enzyme incorporation in biosensors for compounds such as substituted phenols and hydrocarbons (Riedel *et al.* 1993; Beyersdorfradeck *et al.* 1994; Peter *et al.* 1996). The possibility of *Rhodococcus* phenylalanine dehydrogenase used in a biosensor to screen for phenylketonuria was proposed by Hummel in 1997 (Hummel 1997). A biosensor for heroin detection was explored 17 years ago: heroin esterase from a rhodococcal strain was studied by Cameron

et al. 1994, and consequently research into the development of a heroin detection biosensor was funded by UK Customs and Excise (Coghlan 1994).

2.3. Sequencing the genome of *R. rhodochrous* ATCC BAA-870

2.3.1. The value of a microbial genome

As commented on by Galperin and Koonin in a recent review (2010), the beauty of the genome, or "the value of the sequence information, is in the eye of the beholder" (Galperin and Koonin 2010). Specific uses of genome sequence information can be sought after by the research group spearheading the genome project, such as the environmental remediation goals of the sequencers of *Rhodococcus jostii* RHA1 mentioned above. See Appendix 7.2.1 for a list of South African genome projects (excluding human genome projects) and Appendix 7.2.2 for a summary of the cost motivation behind sequencing the BAA-870 genome.

However, regardless of the specific aims, the genome of an organism generally can be used for:

- cloning and expression of any gene with relative ease
- reconstruction of metabolic pathways and signalling networks
- studying the adaptations to environmental and ecological stimuli and mutational patterns in a quantitative manner
- identifying genes and traits not obvious in the organism
- determining virulence factors in the case of pathogens
- creating microarrays to analyse gene expression
- studying the evolutionary traits of an organism on a small scale
- adding to the knowledge of biological diversity
- indirectly improving human and animal health.

2.3.1.1. Harnessing national biodiversity

Genomic resources and microbial diversity in South Africa is virtually unexplored. It is estimated that more than 90% of the world's microbial population is uncultured (DeLong 2009; Hugenholtz 2002), and the estimation for South African microbial resources is even higher at an estimated 99% (Burton and Cowan 2002). Given the immense geographical and micro-habitat diversity in the country, it can be said that the local biodiversity is a major pool of potential drug discovery, and the exploration and development of new organisms, enzymes and genes for the foods and environmental sectors, for example, can be expected to contribute a large share of the global annual microbial technologies market.

However, the potential for successful drug discovery will only be realized if there are effective mechanisms in place to "take the discoveries from the forest to factory in a sustainable manner" (Burton and Cowan 2002). Successful applications in biotechnology require investment in primary research and establishing the infrastructure for technology transfer of later stage research.

2.3.2. The genome as a resource for biocatalysis: The reverse genetics approach

Generally, the traditional route for obtaining cloned enzymes for biocatalysis is followed. Screening of a large microbial culture collection provides the 'hits' that are desired for further development into commercial bioprocesses. The whole organism approach has several disadvantages despite the success obtained in many cases. The strategy is an empirical one and the screening process has to be repeated for each substrate/reaction pair. The process is labour-intensive and many false or negative 'hits' are obtained since whole cells often catalyse side-reactions in addition to the desired conversion, or degrade the reaction product. The enzyme of interest makes up a relatively tiny proportion of the total protein content of the organism, and its level of activity can be too low to outcompete that of other enzymes to the substrate.

Once the enzyme which is responsible for the desired activity is isolated, its amino acid sequence can be used to construct primers for amplification and sequencing of the gene, which can then overexpressed in a heterologous host. This approach has been highly successful and remains extremely useful. However, in the case of enzymes present at naturally low levels, the initial purification is difficult, and amplification of genes from organisms with difficult template (high GC content and multiple closely related genes) can make the process difficult. Ultimately, the intrinsic properties of the enzymes available will determine how well biocatalysis enables novel routes to target molecules. In most cases the target enzymes did not specifically evolve for the reactions that they are used to catalyse. In identifying biocatalysts using the whole organism screening approach, one must identify the best enzyme for a specific task in a phase of bioprocess development which is time consuming and expends a significant proportion of the total effort used.

The sequencing of multiple genomes has expanded the inventory of putative protein sequences and enabled a 'reverse-genetics' approach to biocatalyst discovery, wherein prior knowledge of the gene itself expedites the screening process. Using bioinformatics methods, it is possible to identify classes of genes which can be cloned, expressed, purified and evaluated for their activity. The advent of high-throughput expression and purification methods has made the process for bulk screening even more feasible. The power of the reverse-genetics screening method increases in direct proportion to the number of sequenced genomes (Stewart 2006).

The strength of the reverse-genetics approach was demonstrated by Stewart 2006, who used it to build a library of synthetically useful dehydrogenases from *Saccharomyces cerevisiae* (baker's yeast). Baker's yeast was the first genome to be completely sequenced, and has more than a hundred year history of use in the reduction of a broad array of aldehydes and ketones. It is arguably the only living organism accepted as a routine and normal reagent in organic chemistry due to its low cost, lack of toxicity and availability. There are reports that individual yeast dehydrogenases may be highly stereoselective, although the overall bioconversions by the whole cell do not reflect this. Only a few yeast reductases have been identified using the traditional purification and gene cloning route, and therefore Stewart (2006) used the reverse-genetics approach to demonstrate its

feasibility in obtaining a broad range of biocatalytically useful enzymes. They identified 49 yeast sequences of interest, targeted 23 ORFs for expression after eliminating gene sequences of enzymes already well-characterised, and used 19 of these for testing bioconversion abilities (Stewart 2006). The reduction of a range of unsubstituted and α-substituted β-keto esters and ethyl acetoacetate substrates were tested. Most of their targets had unknown properties, and several targeted ORFs were hypothetical proteins. All their target proteins were expressed using fusion tags for ease of purification and to keep a universal strategy for all their targets. Their genome library approach provided reductases that performed a range of biocatalytic reductions, some with high stereoselectivities, and they demonstrated the performance of a library of dehydrogenases that could give multiple stereoisomers from a common starting material.

Similarly, a bacterial glycosyltransferase gene toolbox was generated by Erb *et al.* 2009, for natural product biosynthesis (Erb *et al.* 2009). More than 70 bacterial glycosyltransferases are included in their tool box, which provide a base for establishing detailed knowledge of the functions and mechanisms of these enzymes which will enable rational design of novel glycosylated compounds. A toolbox of genes is an enticing way to exploit the full metabolic potential of the enzymes therein, and sequencing the genome provides a large pool of information for selective additions to the box.

2.3.2.1. Directed gene cloning

The value of the complete sequence of *R. rhodochrous* ATCC BAA-870 for biocatalysis applications is enormous. Application of whole cell biocatalysts is problematic, especially in the case of our target activities where, for example, small aliphatic nitriles, and those with hydroxy and amino groups, can be utilized by the organism itself as carbon sources (Meth-Cohn and Wang 1997). Few synthetically useful pure enzymes were available until recently, and varieties of lipases and other hydrolases were the first class of enzymes to be packaged into enzyme kits that could be screened by bench chemists with relatively little or no biological training. Dehydrogenases have similarly been targeted (Zhu *et al.* 2005), but to date there is no kit containing nitrile converting enzymes. A library of enzymes allows multiple stereoisomers from a single starting compound. Biocatalyst targets of interest from

this organism include the high molecular weight nitrile hydratase, signature amidases, epoxide hydrolases, and lignocellulosic enzymes, to name just a few. Given the full genome sequence information, we can more easily 'mine' these enzymes with a directed approach.

2.3.3. Sequenced *Rhodococcus* genomes

In 2006 there was only one active full genome sequencing project involving *Rhodococcus* (Project ID: 13693) submitted to the National Center for Biotechnology Information (NCBI) database, that of *Rhodococcus* sp. RHA1, by the University of British Columbia, Canada/ Genome Sciences Centre, Canada. *Rhodococcus* sp. RHA1 was isolated from soil in Japan which was contaminated with a toxic insecticide, lindane (or γ-hexachlorocyclohexane) (Seto *et al.* 1995). It was found to be capable of degrading a range of polychlorinated biphenyls (PCBs), which are toxic compounds produced by industry (Masai *et al.* 1995). Its full genome is 9.7 Mbp, inclusive of one 7.8 Mbp chromosome and 3 plasmids (pRHL1, 2 and 3). The goal of that genome project was to better understand metabolic pathways and gene functions in the organism, and to apply the organism to bioremediation of contaminated environmental sites.

During 2009, an additional 3 *Rhodococcus* genome sequences were submitted to the NCBI, including *R. erythropolis* PR4 and SK121, and *R. opacus* B4. In 2010, two strains from the pathogenic *equi* subtype were submitted, *R. equi* 103S and ATCC 33707 (Table 2.2). There are numerous rhodococcal plasmid sequences submitted to the NCBI, including a 7637 nt *R. rhodochrous* plasmid pNC500 which was completed in 2007 (GenBank accession number AB266604). There is currently no full genome sequence for a *R. rhodochrous* strain in the NCBI.

Table 2.2: *Rhodococcus* genome sequencing projects currently submitted to the NCBI

Rhodococcal Genome	Sequence completed	NCBI Reference sequence status	Group/s responsible
R. jostii RHA1	2006	Provisional	University of British Columbia, Canada, and Nagaoka University of Technology, Japan
R. erythropolis PR4	2009	Provisional	National Institute of Technology and Evaluation, Japan
R. erythropolis SK121	2009	Draft	J. Craig Venter Institute
R. opacus B4	2009	Provisional	Hiroshima University, Department of Molecular Biotechnology, Japan
R. equi 103S	2010	Provisional	IREC (International *Rhodococcus equi* Genome Consortium)
R. equi ATCC 33707	2010	Draft	BCM (Baylor College of Medicine), Texas

2.4. Objectives

Rhodococcus species isolated from soil are known to have diverse catabolism, and their genomes hold the key to survival in complex chemical environments. This characteristic gives them genomes rich with potential synthetic targets and novel chemistries. The aims of this study were to sequence the genome of a biotechnologically relevant organism, *R. rhodochrous* ATCC BAA-870, in order to use the genome as a resource for gene fishing and specific gene identification, and as a resource for future potential biocatalysis. The optimal assembly and annotation of the genome is explored. The genome content is summarized with respect to the protein complement and genome organization of BAA-870, and compared to other rhodococcal genomes. This genome represents the first *rhodochrous* subtype sequenced, and will add to the knowledge of *Rhodococcus* phylogeny.

Specifically, the nitrile converting genes in the genome of BAA-870 will be summarized and compared with other rhodococcal genomes. Using the information extracted from the genome of BAA-870, potential targets of interest for future biotechnology applications will be identified. An outline of the benefits of having the genome sequence of *R. rhodochrous* ATCC BAA-870 is also presented herein.

2.5. Materials and Methods

2.5.1. Strain and culture conditions

R. rhodochrous ATCC BAA-870, isolated from soil in Johannesburg, was grown routinely on Tryptone soya agar (TSA) medium (Brady *et al.* 2004b). For genomic DNA preparation, the strain was grown in 50 mL Tryptone Soya Broth overnight at 37°C. Cells were centrifuged and the DNA purified using a using a Wizard® Genomic DNA Purification Kit (Cat. # TM050) from Promega (USA). DNA concentrations were measured spectrophotometrically by absorbance readings at 260 nm using a NanoDrop-1000.

2.5.2. Sizing of the *R. rhodochrous* ATCC BAA-870 genome using pulsed-field gel electrophoresis

2.5.2.1. Bacterial culturing

The previously characterized strain, *Rhodococcus* sp. R312 (Arnaud 1976, Briand 1994) was used as a genomic DNA size and restriction analysis control (Bigey *et al.* 1995). *R. rhodochrous* ATCC BAA-870 and *Rhodococcus* sp. R312 were grown in Luria Bertani (LB) medium and *Nitriliruptor alkaliphilus* was grown in minimal medium substituted with isobutyramide. *N. alkaphilus* is a new Actinobacterium closely related to, but distinct from, the *Rhodococcus* species, and recently characterised by collaborators at TU Delft, where the experiment was performed (Sorokin *et al.* 2009). Bacterial growth was measured using absorbance at 420 nm (Bigey *et al.* 1995).

2.5.2.2. Genomic DNA plug preparation

Bacterial cells were weakened by addition of 2% (w/v) glycine one hour before harvesting. Cells were centrifuged at 4 000 x *g* for 5 minutes and washed twice in Pett IV buffer (10 mM Tris-HCl, 1 M NaCl, pH 7.6) (Smith *et al.* 1988). Cells were suspended in Pett IV buffer to A_{420}

of 15 – 20. DNA-containing plugs were made by mixing 325 µl of cell suspension with 975 µl low melting point agarose (maintained at 40 °C) in plug molds as per the instructions in the BioRad Pulsed-field gel electrophoresis (PFGE) kit. Plugs were incubated at 4 °C for 30 minutes, then removed and incubated at 50 °C for 48 hours in 2 volumes of cell lysis buffer with one buffer change (10 mM Tris-Cl pH 7.5, 1 M NaCl, 100 mM EDTA, 0.5% (w/v) n-laurylsarcosine, 0.2% (w/v) sodium deoxycholate, 0.5% (w/v) Brij-58, 5 mg/ml lysozyme and 1 mg/ml proteinase K). The plugs were washed for 2 hours with gentle agitation at 37 °C in 8 volumes of Tris-EDTA (TE) buffer (10 mM Tris-Cl pH 8.0, 1 mM EDTA) containing 2 mM Pefabloc SC, followed by a wash 4 times with 8 volumes of TE buffer for 1 hour at room temperature. Plugs were stored in 0.5 M EDTA at 4 °C.

2.5.2.3. Restriction analysis

Agarose plug slices were washed 3 times with TE buffer for 30 minutes and then equilibrated twice by incubation in appropriate restriction buffer for 30 minutes on ice. Restriction enzyme (15 to 20 Units) was added and incubated for 16 hours at 4 °C, and then 4 hours at 37 °C. *R. rhodochrous* ATCC BAA-870, *Rhodococcus* sp. R312 and *Nitriliruptor alkaliphilus* were digested separately with *Ase*I, *Dra*I, *Ssp*I, *Swa*I, *Xba*I and *Hin*dIII.

2.5.2.4. Electrophoresis

Megabase agarose gels, 1 % in Tris-borate-EDTA (TBE) buffer, were run in a clamped homogeneous electric fields (CHEF) system (CHEF DR-II apparatus, Bio-Rad, U.S.A.). The electric field was maintained at 6 V/cm at a temperature of 14 °C, and run at different parameters depending on the desired fragment size separation (Table 2.3). Midrange PFG marker I agarose plug (Bio-Rad) was used as a size marker. No restriction digest controls were used. Gels were stained in ethidium bromide Safestain (Invitrogen) for 1 hour, and destained using deionised water.

Table 2.3: Electrophoresis parameters for separation of different fragment sizes

Fragment size (kb)	Run time (hours)	Ramp pulse time (seconds)	Field voltage (V/cm^{-})
600-900	48	50-60	6
350 – 650	40	30 – 45	6
240 – 485	40	15 – 30	6
60 – 240	30	5 – 15	6
48-1000	20	40-55	6
5 – 100	16	0.5 – 5	6
Up to 3 Mb*	96	700 – 800	2

*For separation of very large fragments, megabase agarose was used at 0.8%.

2.5.3. Whole genome shotgun sequencing with Illumina Genome Technology

R. rhodochrous BAA-870 genomic DNA was sequenced according to the manufacturer's protocol using Solexa technology. Briefly, genomic DNA (~5 mg) was fragmented by nebulization into fragments of less than 800 nt. The fragments were end-repaired and a single A nucleotide was added using Klenow enzyme (Epicentre Biotechnologies, Wisconsin). Adapters needed to hybridize the fragments to the flow cell were then ligated to the fragmented genomic DNA. The ligation products were separated and purified by agarose gel electrophoresis to recover 150-250 bp products. These fragments were then PCR amplified using primers complementary to the now attached adapter sequences. After a final purification step (QIAquick PCR purification kit, Qiagen), quality control was performed as suggested by the manufacturer, measuring the OD_{260} and the OD_{260}/OD_{280} ratio. An aliquot of the sample (7 pmol) was used to generate DNA clusters in one channel of a sequencing flow-cell. The flow-cell was subjected to 36 cycles of sequencing reaction on the Illumina Genome Analyzer. The Solexa sequencing output file data was analysed using the Solexa Data Analysis Pipeline v0.2.2.5 software and quality filtration using standard parameters was performed. A total of 5,643,967 reads, 36-bp in length were obtained (4,273,289 unique sequences).

2.5.3.1. Assembly and annotation of Solexa Sequencing data

The genome sequence data was assembled using a variety to programmes to assess the best assembly. These included Edena (Hernandez *et al.* 2008), Velvet (Zerbino and Birney 2008), Staden (Staden 1977), and a combination thereof. The best assembly data set was then annotated automatically using dCAS (Guo *et al.* 2009) and BASys (Van Domselaar *et al.* 2005), with annotations checked manually. Edena (Exact DE Novo Assembler) programme, version 2.1.1.1 for Windows (Hernandez *et al.* 2008), was downloaded from www.genomic.ch/edena (Genomic Research Laboratory, Geneva University Hospitals, Switzerland). Edena was run on an AMD Opteron CPU 2.8 GHz supplied with 4 Gb of RAM. To validate the assembly, we looked for a number of key genes, known to be present. As some of these could not be located, it was assumed there was insufficient coverage. Hence we performed mate-paired sequencing to improve the completeness of the data.

2.5.3.2. Assembly and annotation of mate-paired sequencing data

Purified genomic DNA from *R. rhodochrous* ATCC BAA-870 was submitted to BaseClear® BV in Leiden, the Netherlands. The original Illumina sequence reads obtained from the *ab initio* sequencing of *R. rhodochrous* ATCC BAA-870 were provided to the company, along with two sets of assembled contigs (assemblies of sequence data from *R. rhodochrous* ATCC BAA-870 using different parameters). Baseclear applied quality filtrations to all Illumina sequence reads by removing paired reads containing N's as nucleotides, and removed duplicated paired reads using the program SSPACE version 2 (Boetzer *et al.* 2011), which orders and orientates the contigs and combines them into so-called scaffolds. Baseclear scaffolded the quality filtered pre-assembled contigs into a number of scaffold sequences. The final scaffolds were produced by SSPACE in FASTA format and a statistical summary of the results provided by the company. The average size distance between mates was ~4500 bp. For assembling two BAA-870 sequence libraries, a shotgun library and a mate pair library, were used. The shotgun library with read length of 36 bp were used as input for contig building and the shotgun library, with read length 50 bp and a 5kb insert size, for scaffolding the assembled contigs. The IDBA assembler (Peng *et al.* 2010) was used for contig building. By using the mate pair library in SSPACE (Boetzer *et al.* 2011) the produced contigs were

ordered into scaffolds (Table 1). Edena (Exact DE Novo Assembler) programme, version 2.1.1.1 for Windows (Hernandez *et al.* 2008) was used for assembly.

2.5.3.3. Computer resources and software

Edena (Exact DE Novo Assembler) programme, version 2.1.1.1 for Windows (Hernandez *et al.* 2008), was downloaded from www.genomic.ch/edena (Genomic Research Laboratory, Geneva University Hospitals, Switzerland). Edena was run on an AMD Opteron CPU 2.8 GHz supplied with 4 Gb of RAM.

2.6. Results

2.6.1. Sizing of the *R. rhodochrous* ATCC BAA-870 genome using pulsed-field gel electrophoresis (PFGE)

The ideal restriction enzyme for sizing the genome should be specific for a recognition sequence which does not occur too often in the genome in order to obtain a small number of fragments which can be conveniently separated. Since the genome of BAA-870 is GC-rich, the restriction enzymes used should be AT-rich specific, contain nucleotide combinations which are relatively rare and be at least 6- or 8-bp cutters. The CTAG combination was found to be relatively rare in GC-rich bacteria (McClelland *et al.* 1987). For this reason, restriction enzymes containing the CTAG nucleotide combination were used. The restriction enzymes suitable for this purpose are severely limited, however, and only a few are commercially available. An initial PFGE experiment determined that *Hin*dIII was not suitable for size determination as it cut too often (left smeared DNA fragments, and too many small fragments which ran off the gel). The PFGE experiment was repeated without *Hin*dIII (Figure 2.2).

The restriction enzymes *Ase*I, *Ssp*I and *Xba*I gave good ranges of DNA size fragments after restriction (Table 2.4), however, there were many small fragments which were slightly smeared and difficult to quantify. A different set of parameters is required to separate fragments optimally in the smaller size range. The presence of a ~450 kb DNA fragment with no enzyme treatment, in both PFGE experiments, suggests the organism harbours at least one extra-chromosomal plasmid. Additionally, this plasmid is presumed to be a linear plasmid, given that there is no evidence of the typical 'tri-banded' pattern of circular plasmid DNA run through agarose gel: the coiled and supercoiled plasmid DNA pattern.

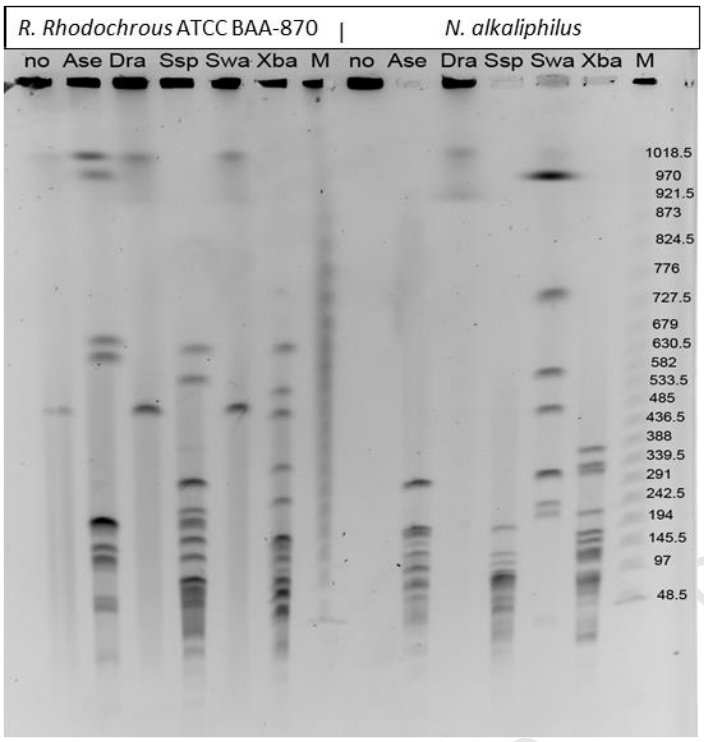


Figure 2.2: PFGE of restricted DNA from *R. rhodochrous* ATCC BAA-870 and a comparative control, *N. alkaliphilus*.

Pulse ramps were 40 – 55 s for 20 hours at 6 V/cm. Midrange PFG marker I (Bio-Rad) was used as a Kb size marker, with fragment sizes ranging from 1018.5 to 48.5 Kb. Restriction digestions no, Ase, Dra, SSp, Swa and Xba refer to the no digestion control, and digestions with *Ase*I, *Dra*I, *Ssp*I, *Swa*I and *Xba*I, respectively.

*Dra*I and *Swa*I gave similar digestion fragments sizes, despite their recognition sequences being different. Their total restriction fragment sizes are also similar at ~1.9 Mb. A successful PFGE genome sizing experiment should give total restriction fragment sizes of comparable values. The smallest size indicated by the restriction enzymes used is ~1.9 Mb with *Dra*I and *Swa*I, but ~4.1 Mb using *Ase*I. The total sizing is therefore not accurate using the current set of restriction enzymes, and a mean size cannot be determined. However, the experiment was a useful determinant of the presence of plasmids and set a lower limit on the size of the genome.

Table 2.4: DNA fragment sizes of *R. rhodochrous* ATCC BAA-870 as determined by enzyme restriction and PFGE analysis

Fragment size (kb) after digestion with:						
Fragment	No enzyme	*Ase*I	*Dra*I	*Ssp*I	*Swa*I	*Xba*I
1	453	1033	1033	597[2]	1033	649
2		988	**463**[2]	530	**459**[2]	505
3		610		**296**[2]		436
4		578		232		285
5		**206**[3]		206		202
6		148		193		**137**[2]
7		116		164		122
8		30		84		107
9		14		**75**[2]		89
10				49		**80**[2]
11				23		**66**[2]
12						**55**[2]
13						32
Total	453	4134	1956	2819	1951	3105

Bold numbers represent high intensity bands assumed to be doublets (superscript, 2) or triplets (superscript, 3).

Ideally, complete digestion of the genomic DNA plug should be achieved, as indicated by no presence of DNA staining in the well after the fragments are separated by pulsed field electrophoresis. This is noted in the digested DNA fragments for the sample *N. alkaliphilus*, used as a control, where the enzymes *Ase*I, *Ssp*I, *Swa*I and *Xba*I have completely digested the DNA in the plugs which were inserted into the sample wells. This suggests that the BAA-870 genome may not be entirely digestible by the restriction enzymes used, as the same concentration of DNA is present for samples *N. alkaliphilus* and BAA-870. The undigested genome material is proposed to be bigger than 1000 kb, making the lower size limit for the genome greater than 5.1 Mb.

2.6.2. Taxonomy and lineage of *R. rhodochrous* ATCC BAA-870

The 16S sequence of BAA-870 was compared with several other *Rhodococcus* 16S sequences, and a phylogenetic tree constructed (Figure 2.3).

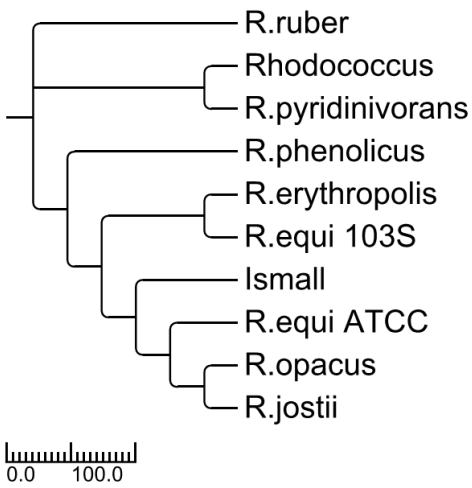


Figure 2.3: Phylogenetic tree of *Rhodococcus* 16S sequences

The figure was created using BioEdit sequence alignment Editor (Hall 1999) and ClustalW (Thompson *et al.* 1994). The strain Ismall refers to *R. rhodochrous* ATCC BAA-870.

2.6.3. Assembly of Solexa Illumina SRS data

For a description on the use of the Edena programme to assemble short read sequences, and the overall assembly statistics for reads from *R. rhodochrous* BAA-870, refer to Appendix 7.2.3. Initial sequence assembly involved changing the contig minimum output size, the minimum sequence overlap length and adjusting the programme mode between strict and non-strict assembly to find the optimal sequence assembly parameters. The overall sequence assembly statistics for *R. rhodochrous* BAA-870 are shown in Table 2.5.

A total of 5.65 Mbp of sequence was obtained from initial Illumina sequence reads (36 bp), with 4.27 Mbp assumed to be unique. In strict mode, the minimum contig size (*m*) that gave the highest N50 (2425) and percentage coverage (72.4 %), and the lowest number of contigs (4740), was 22 bp. The allowed minimum contig size was tested to compare the minimum number of base pairs that gave the optimum in non-strict mode. In non-strict mode, the minimum contig size that gave the highest N50 (5230) and percentage coverage (73.6 %), and the lowest number of contigs (2206), was 20 bp. Changing *m* to 18 bp made no difference to the data obtained since the default for the programme keeps the overlap length (*M*) as 20 bp. To test that the optimum data was obtained for *m* at 20 bp, *M* was

lowered to 16 to allow data generation at the lower values (Appendix 7.2.3). Overall, non-strict mode gave a higher percentage of reads assembled than strict mode (73.6 % vs 72.4 %), and a higher N50 value (5230 vs 2425). The lowest number of contigs was obtained using non-strict mode (2206 vs 4740).

Table 2.5: Overall assembly statistics for Solexa Illumina sequence reads from *R. rhodochrous* BAA-870

Fasta entries	5651222 bp
Sequence discarded	7255 bp
Sequence ok	5643967 bp
Unique sequence	4273289 bp
Reads length	36 bp
No. of nodes	4273289
No. of edges	4167859
Minimum overlap size	20 bp

In the case where the minimum overlap size in the overlap phase was changed to 16 bp with a minimum contig size of 19 for assembly, the smallest number of contigs obtained in this scenario was 2082, with an N50 value of 5540. In this case the percentage of reads assembled was 73.7 %. Overall assembly results should indicate whether all regions of the genome are represented if the reads are mapped against a reference genome sequence. Regions which appear to be weakly represented tend to contain long single base repeat sections which form secondary structures, as well as large complex moieties that do not replicate efficiently during sequencing steps.

2.6.4. Apparent sequence coverage by Solexa Illumina sequencing

After quality filtration of the Solexa sequencing output data using standard parameters was performed (Materials and Methods), a total of 5,643,967 reads, 36-bp in length were obtained (4,273,289 unique sequences).

2.6.4.1. Raw coverage depth

The raw coverage depth is given by the number of unambiguous reads multiplied by the read length, divided by the reference genomic DNA size:

$$(4.273 \times 10^6 \text{ X } (35)/ 5.643 \times 10^6)$$

The raw coverage depth is therefore 26.5X.

The sequence coverage of the BAA-870 genome using initial Solexa Illumina sequencing is therefore 74%, with an apparent raw coverage depth of 27X, if the total genome size is 5.8 Mbp.

2.6.5. Paired-end sequencing and mate-paired assembly

In order to improve the genome quality and coverage, the assemblies were subjected to a further round of sequencing by the so-called mate-paired assembly route, using general algorithm parameters (Table 2.6). After mate-paired sequencing and re-assembly, the final sequence coverage of the BAA-870 genome was 135X.

Table 2.6: General algorithm parameters used for mate-paired assembly of both pre-assembled data sets

Parameter	Value
Minimal number of read pairs required to connect contigs	5
Minimal number of overlapping nucleotides to merge contigs	15

The statistical summary of scaffolded and quality filtered pre-assembled contigs produced by SSPACE in FASTA format is highlighted in Table 2.7.

Table 2.7: SSPACE statistics of quality filtered sequence reads for two sets of mate-paired sequences

	Data Set 2 (Rdc_cap3_assembly.fa)	**Data Set 1 (Edena Assembly; Contig.fa)**
	Count	
Inserted pairs	7 452 293	7 452 293
Pairs containing N	19 583	19 583
Remaining pairs	7 432 710	7 432 710
Numbers of pairs used for scaffolding	87 308	111 326
Satisfied pairs	77 393	106 439
Unsatisfied pairs	9 915	4 887

Most of the filtered reads could successfully be assembled and yielded a number of scaffold sequences (Table 2.8).

Table 2.8: Statistics summary of mate-paired data sets before (pre-assembled contigs) and after (scaffolds) mate-paired assembly

	Data Set 2 Rdc_cap3_assembly.fa		**Data Set 1 Contig.fa**	
	Pre-assembled contigs	Mate-paired Scaffolds	Pre-assembled contigs	Mate-paired Scaffolds
Number of sequences	1 549	774	4 701	3 335
Sum (bp)	5 828 896	6 715 483	5 737 117	9 504 788
Max sequence size	44 038	164 783	30 378	84 870
Min sequence size	61	61	101	101
Average sequence size	3 763	8 676	1 220	2 850
N50	6 874	40 260	2 427	11 904

The scaffolded pre-assembled sequence data labelled 'rdc_cap3_assembly.fa' yielded a higher N50 value (40 260 as opposed to 11 904) with an average sequence size of 8 676 bp rather than 2 850 bp for 'contig.fa'. However, scaffolding of the contig.fa data set, showed a higher total bp sum of 9 504 788, or ~9.5 Mbp, whereas the cap3 scaffold gave a total of

~6.7 Mbp. This shows the extreme differences that can be obtained in sequence summary information according to the method of pre-treatment of sequence data.

The sequence library and assembly statistics are shown in Table 2.9 and Table 2.10 respectively.

Table 2.9: Sequence libraries statistics

	Single	**5kb-raw**
Read length	36	50
# of Reads	5.651.222	14.904.586
Total bp	203.443.992	745.229.300
Coverage assuming 5.5MB genome	~36.9X	~135.5X

Table 2.10: Assembly statistics

	IDBA assembly	**Filtering contigs >= 500 bp**	**Scaffolded with SSPACE**
# of contigs	1199	861	257
N50 (kb)	11.895	11.957	158.090
Max contig (kb)	103.966	103.966	445.186
Total sequence (Mb)	5.837.060	5.760.205	6.065.913

2.6.6. Annotation

The assembled genome sequence of *R. rhodochrous* ATCC BAA-870 was submitted to the Bacterial Annotation System web server, BASys (http://BASys.ca/BASys/cgi/submit.pl), for automated, in-depth annotation of the chromosomal sequence (Van Domselaar *et al.* 2005). The raw sequence data was submitted to the server which uses over 30 programs to identify almost 60 annotation subfields for each gene, including the gene and protein name,

function, possible paralogues and orthologues, molecular weight, isoelectric point, operon structure, subcellular localization, signal peptides, transmembrane regions, reactions, and pathways. The BASys annotation was done for *R. rhodochrous* ATCC BAA-870 raw sequence data with a genome length of 5 828 896 bp length, in which 7000 genes were identified and 6999 genes annotated (Figure 2.4). The output data was obtained from the generated hyperlink for the project, and used to construct images of the annotation summaries and details.

COG functional protein categories (Clusters of Orthologous Groups of proteins) are based on phylogenetoc clusters of proteins from complete genomes. COGs were delineated by comparing protein sequences encoded in complete genomes representing major phylogenetic lineages. Each COG consists of individual proteins or groups of paralogs from at least 3 lineages and thus corresponds to an ancient conserved domain (Tatusov *et al.* 1997; Tatusov *et al.* 2003). The COG function and identity assigned to proteins from the BASys annotation is transitively applied from a similarity Basic Local Alignment Search Tool (BLAST) search against the COG database of orthologous groups of proteins (Tatusov *et al.* 1997).

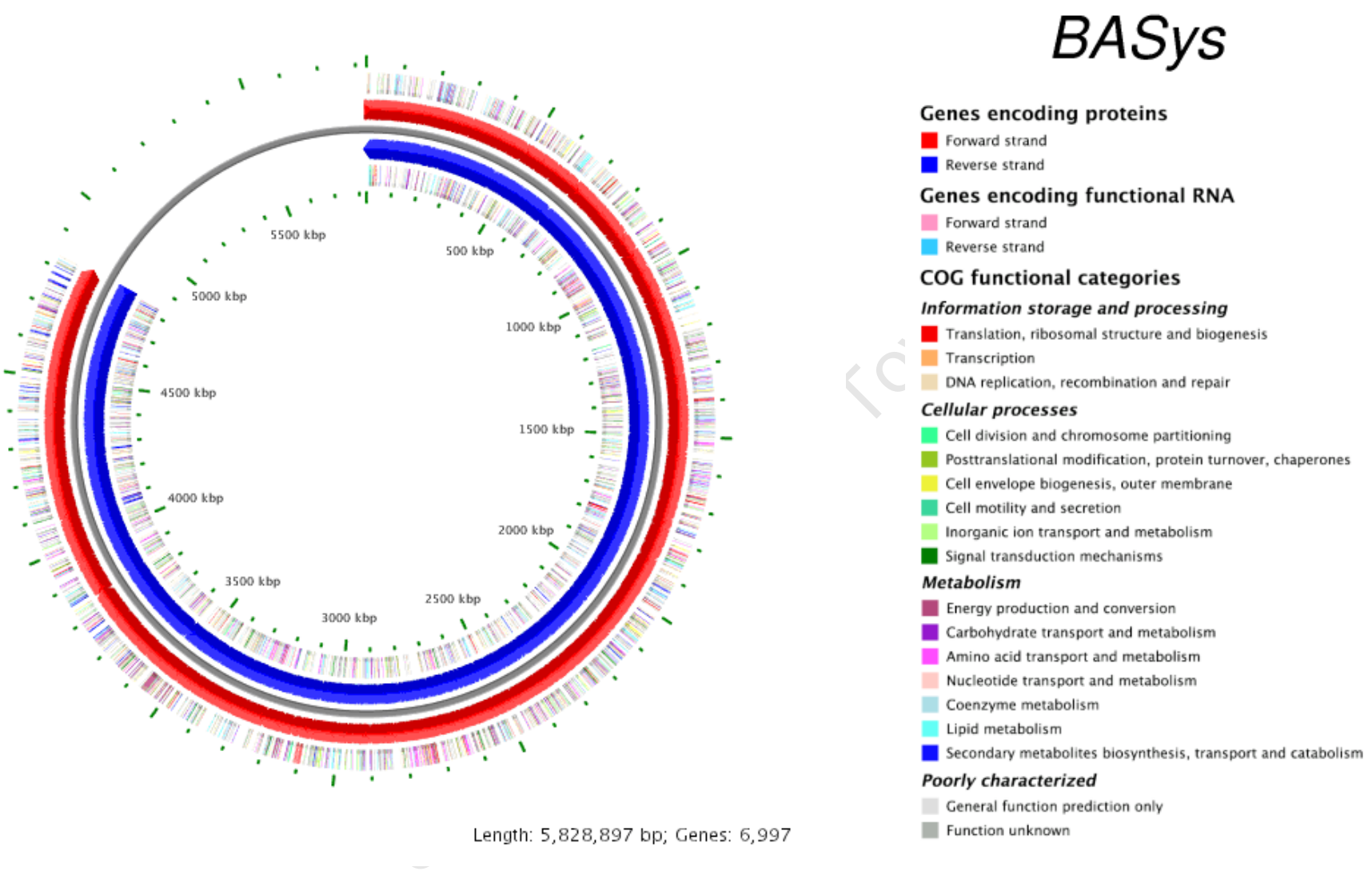


Figure 2.4: BASys bacterial annotation view of the genome of *R. rhodochrous* ATCC BAA-870.

The reverse and forward protein and RNA coding strands are shown, and annotated genes are grouped according to function. The region between 4800 Kbp and 5800 Kbp is 'unannotated' genome. The BASys annotation is based on Data Set 1, the 5.8 Mbp sized genome containing 6997 annotated genes.

2.7. The genome content and protein complement of *R. rhodochrous* ATCC BAA-870

2.7.1. Protein coding sequences

The genomic content of BAA-870 is outlined in Table 2.11 and compared to other rhodococcal genomes. BAA-870 contains 6997 predicted protein-coding sequences (CDS). Of this 0.9% encodes proteins of unknown function. There are 318 conserved hypothetical proteins and 604 hypothetical proteins (Figure 2.5). Additionally, 49 proteins belong to known protein families, but their function is unknown. In total, 971 genes (13.8% of the genome) have hypothetical or unknown function.

The sequenced genomes of various rhodococcal strains show a large difference in genome size. The genome of RHA1, 9.7 Mbp, is the biggest rhodococcal genome sequenced to date. The pathogenic *R. equi* genomes are the smallest at ~5-5.3 Mbp. The apparent genome size of *R. rhodochrous* ATCC BAA-870, 5.83 Mbp, is comparable to other rhodococcal genomes. All *Rhodococcus* have a high GC content, ranging from 62 – 68%. RHA1 has the lowest % coding DNA (87%), which is predictable given its large overall genome size.

Annotation of the protein complement of the genome is done by BLAST match to known sequences in the NCBI database. A large proportion of genes are labelled 'hypothetical' based on sequence similarity and/or the presence of known signature sequences of protein families. The proportions of hypothetical and predicted annotations of the proteins in BAA-870 are shown in Figure 2.5. A predicted protein is one which has no match to similar sequences in the protein database, but likely has a recognised upstream rbs.

Table 2.11: Summary of the genomic content of *R. rhodochrous* ATCC BAA-870 and other sequenced *Rhodococcus* species ranked by completion date

Organism	#Date Completed	Group	Reference	Chromosome (Mbp)	Plasmid (nt)	Total Size, Mbp	G + C %	Coding %	Protein coding genes
R. jostii RHA1	2006	Genome British Columbia, Vancouver	McLeod *et al.* 2006	7.8	1,123,075; 442,536; 332,361	9.7	67	87	9 145
R. erythropolis PR4	2009	Sequencing center: National Institute of Technology and Evaluation, Japan	Sekine *et al.* 2006	6.5	271,577; 104,014; 3,637	6.5	62	91	6030
R. opacus B4	2009	National Institute of Technology and Evaluation, Japan	Not published	7.9	558,192; 244,997. 111,160; 4,367;2,773	7.9	67	91	7246
R. erythropolis SK121	2009	J. Craig Venter Institute	Not published	6.79	None determined	6.79	62	91	6713
R. equi 103S	2010	IREC (International *Rhodococcus equi* Genome Consortium)	Letek *et al.* 2008; 2010	5.04	None determined	5.04	68	90	4512
R. equi ATCC 33707	2010	BCM (Baylor College of Medicine), Texas	Not published	5.26	None determined	5.26	68	91	5030
R. rhodochrous ATCC BAA-870	2011	Protein Technologies, CSIR	This research	5.7	0.45	5.83	65	98	6874*
R. opacus PD630	2011	Massachusetts Institute of Technology and The Broad Institute	Holder *et al.* 2011 (not submitted to the NCBI Genome database)	-	-	-	-	-	-

#Date completed refers to genome sequence completion; plasmids may have been completed at another time. Total genome size comprises the chromosome and the plasmid sequence. *The total number of genes found by BASys is 6997, with 6874 genes annotated. Genome information of strains other than BAA-870 is obtained from the NCBI database.

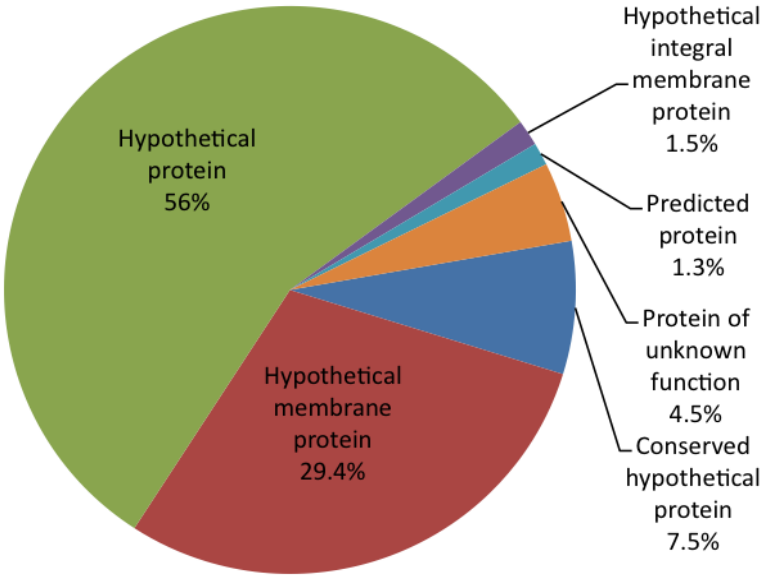


Figure 2.5: Proportions of unknown, predicted and hypothetical annotations of the genome of *R. rhodochrous* ATCC BAA-870.

The proportions of non-specific annotations in the genome (based on a total of 1083 proteins annotated with hypothetical or unknown function).

2.7.2. Transcriptional control

Transcriptional regulatory elements in BAA-870 include 29 sigma factors and at least 106 genes involved in signal transduction mechanisms, 242 genes encoding transcriptional regulators and 63 genes encoding two-component signal transduction systems. Streptomycetes of similar genome size have significantly more, presumably because if their sporulation capacity. There are 119 proteins implicated in translation, ribosomal structure and biogenesis.

The genome encodes all ribosomal proteins, with the exception of S21, as in other actinomycetes (Table 2.12). There is no rhodococcal ribosome reference in the Ribosomal Protein Gene database, http://ribosome.miyazaki-med.ac.jp (Nakao *et al.* 2004). The 52 tRNAs correspond to all 20 natural amino acids and includes two $tRNA^{fMet}$.

Table 2.12: List of ribosomal proteins found in the genome of *R. rhodochrous* ATCC BAA-870

Ribosome subunit	proteins associated
50S	L1, L2, L3, L4, L5, L6, L9, L10, L11, L12P, L13 (2), L14, L15, L16, L17, L18, L19, L20, L21, L22, L23, L24, L25, L27, L28, L29 (2), L30, L31, L31 type B, L32, L33 (2), L34, L35 L6P/L9E
30S	S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18 (2), S19, S20* S15P/S13E

*In the *E. coli* ribosome complement, one protein is found in both subunits (S20 and L26); L7 and L12 are acetylated and methylated forms of the same protein. L8 is a complex of L7/L12 and L10; L31 can exist in two forms (7.9 and 7 kDa).

The average GC% of the coding sequence of *R. rhodochrous* ATCC BAA-870 is 67.46%. The sliding nucleotide composition, or AT/GC fraction across the length of the genome, and direct and reverse strand nucleotide composition of the genome of *R. rhodochrous* ATCC BAA-870 is shown in Appendix 7.2.1 and 7.2.2 respectively.

The majority of the proteins in BAA-870 are between 100 and 400 amino acids in length (Figure 2.6). The amino acid composition of proteins making up the protein complement in *R. rhodochrous* ATCC BAA-870 is shown in Figure 2.7. The average protein lengths of the protein complement is shown in Figure 2.8.

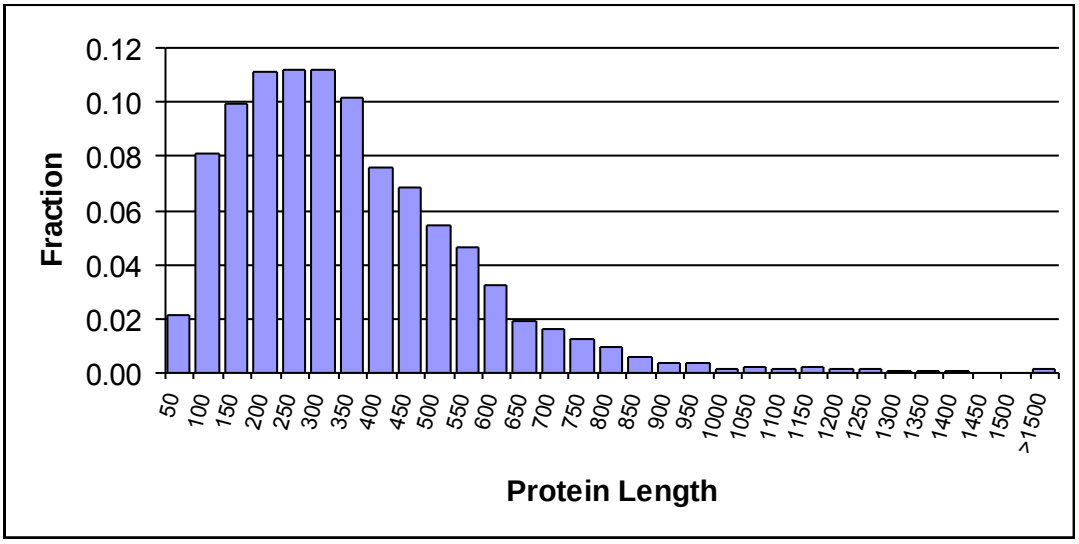


Figure 2.6: Protein lengths for *R. rhodochrous* ATCC BAA-870.

Each bar represents the fraction of the total protein sequences that have a length less than or equal to the given value, except for the last bar, which represents the fraction of the total sequences that are longer than the given length.

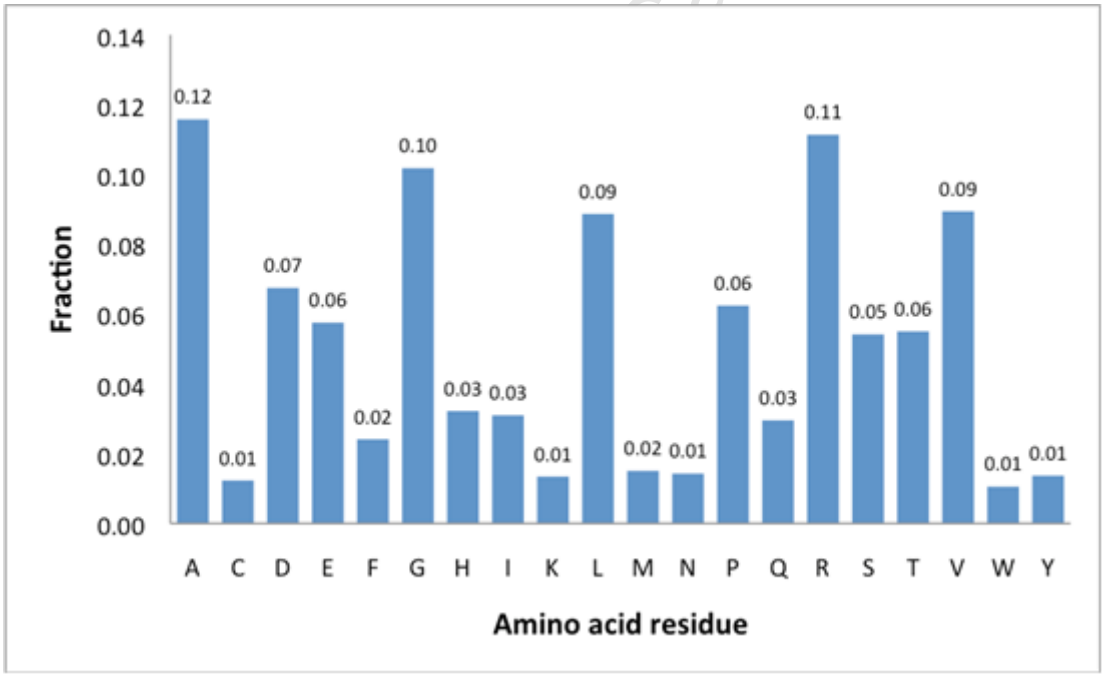


Figure 2.7: Relative proportions of *R. rhodochrous* ATCC BAA-870 protein amino acids.

Amino acid composition values were extracted from the BASys annotation obtained from NCBI. The standard one letter codes are used for amino acid residues. I, V, L = aliphatic; F, Y, W = aromatic. D, E = negative. K, H, R = positive. Each bar represents the fraction of the total amino acids matching the given residue.

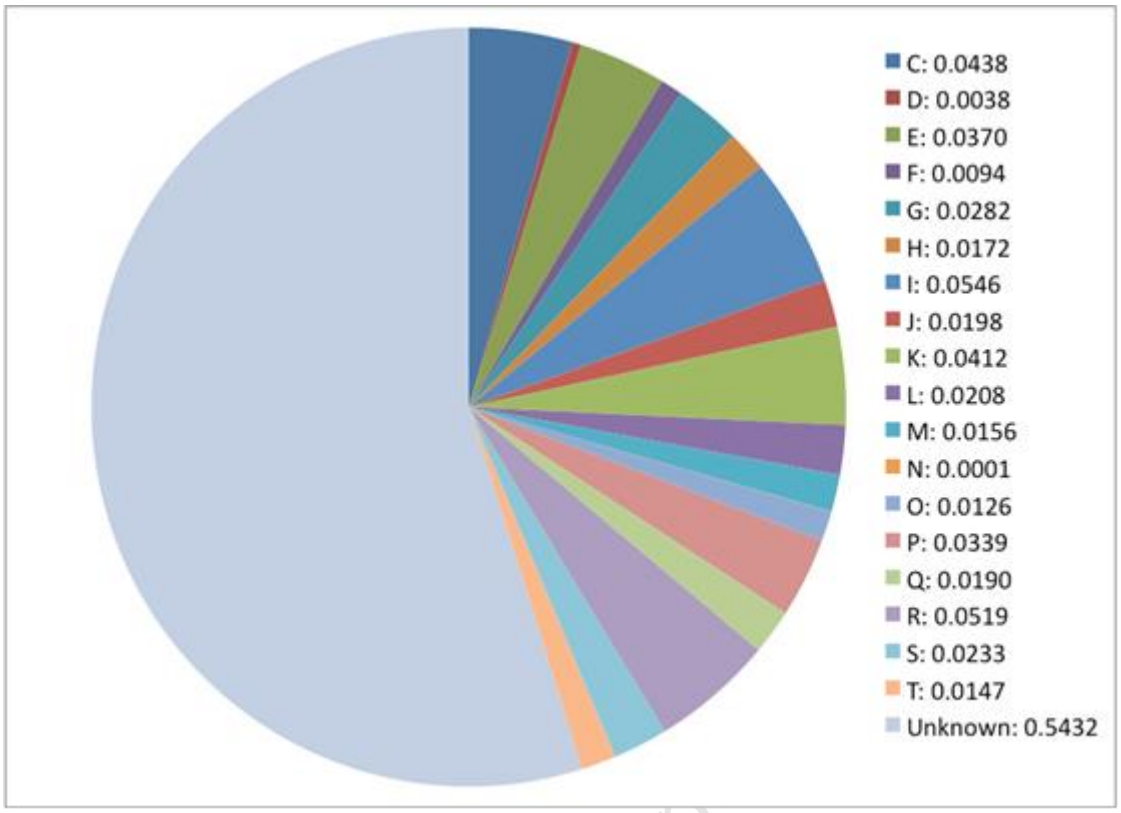


Figure 2.8: Fractions of protein function for *R. rhodochrous* ATCC BAA-870 according to known functional categories.

Unknown proteins form the majority (54.3 %) of proteins in the annotated genome. Letters refer to COG functional categories. C - Energy production and conversion; D – Cell division and chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme metabolism; I - Lipid metabolism;- Translation, ribosomal structure and biogenesis; K - Transcription; L - DNA replication, recombination and repair; M - Cell envelope biogenesis, outer membrane; N – Cell motility; O - Posttranslational modification, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - COG of unknown function; T - Signal transduction mechanisms.

2.7.3. Subcellular localisation

The relative distribution of proteins in BAA-870 is given in Figure 2.9. The majority of proteins are located in the cytoplasm (83.5%), while proteins located at the cellular membrane make up 15.4% of the total. The cytoplasmic proteins can be further subdivided into locations, including nucleoid or cell division sites, and proteins with shared periplasmic and cytoplasmic locations (Figure 2.10).

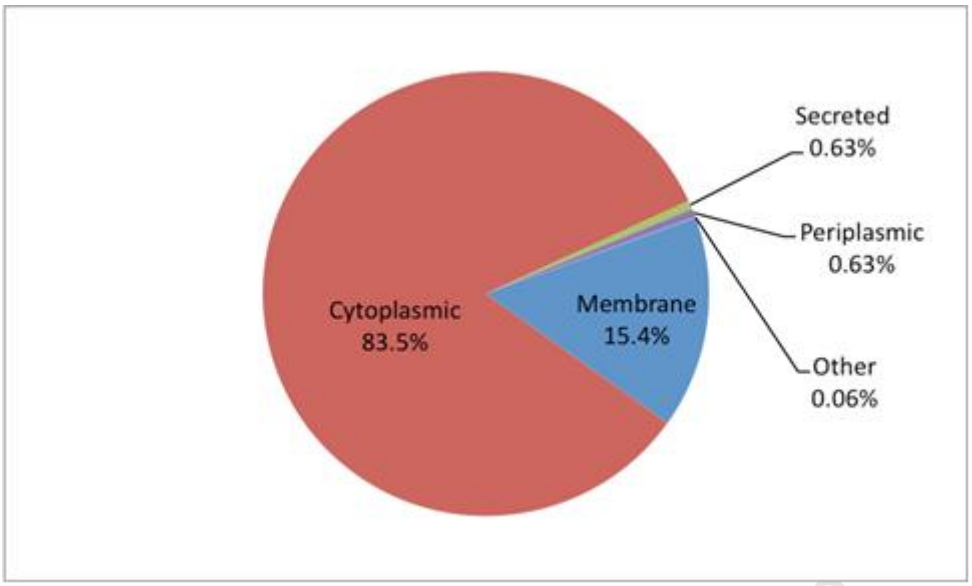


Figure 2.9: Overall protein location for *R. rhodochrous* ATCC BAA-870.

Cytoplasm proteins include potential and probable cytoplasm proteins, and those secreted into the cytoplasm. Membrane proteins can be subdivided into single pass, multi pass, lipid anchor and peripheral membrane proteins.

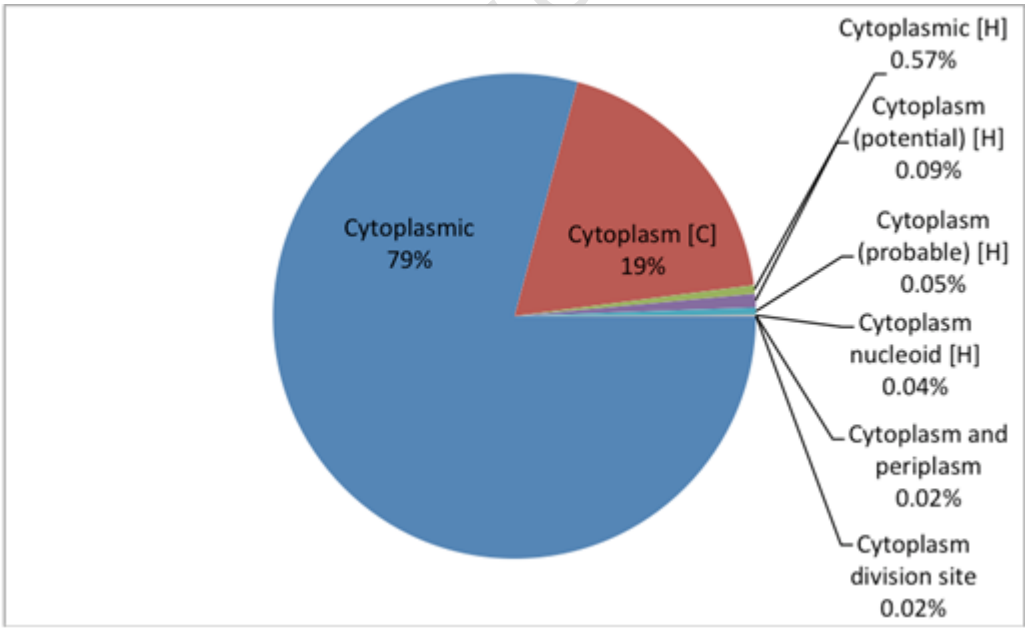


Figure 2.10: The ratio of cytoplasmic proteins in *R. rhodochrous* ATCC BAA-870.

[H]: Homology to a SwissProt Entry and [C]: homology to a CyberCell entry, as inferred by BASys annotation.

The fate of a membrane protein is more complicated with many potential locations and functions, given they are responsible for selective movement of solutes and signalling

molecules into and out of the cell. The proportions of cell membrane protein locations is shown in Figure 2.11.

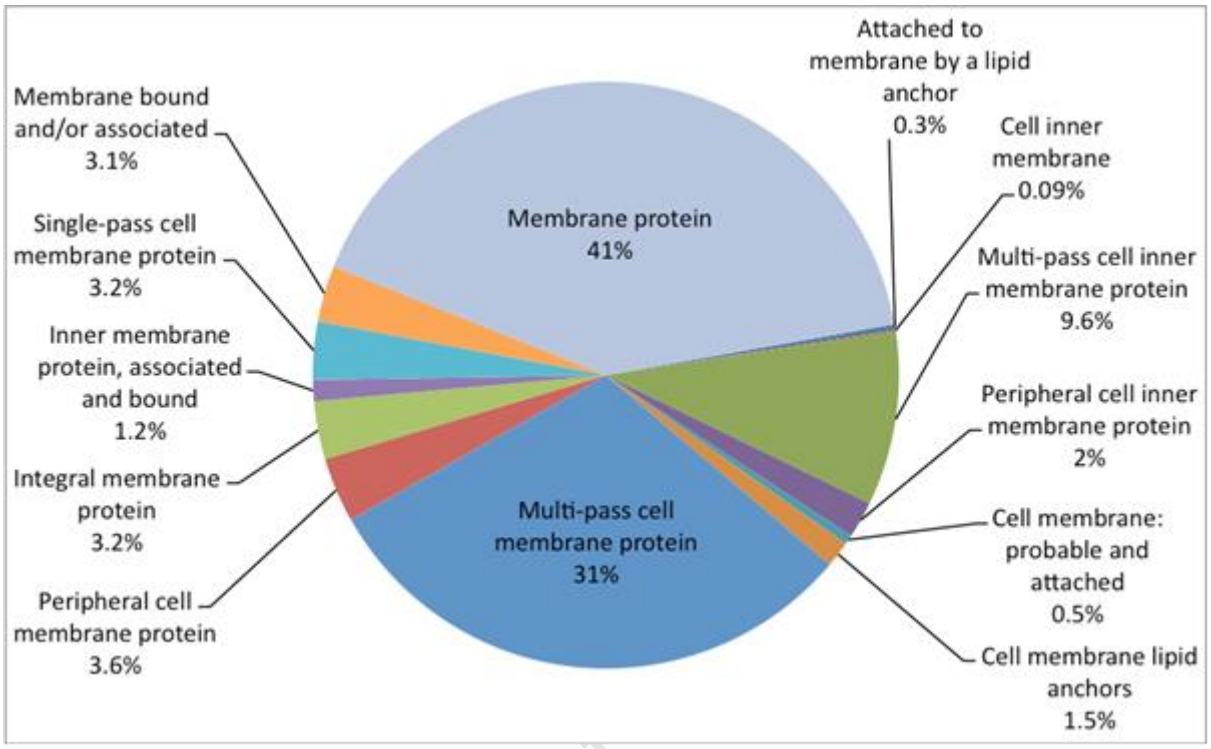


Figure 2.11: The fates of membrane proteins in *R. rhodochrous* ATCC BAA-870.

Proteins labeled probable and potential are included in their respective groups for clarity.

Cell membrane proteins include proteins which form part of lipid anchors, peripheral and integral cell membrane components, and proteins with single or multiple pass functions. Of the membrane proteins in BAA-870, 41% constitute regular membrane proteins, while 31% are multi-pass membrane proteins.

In BASys, the cell location is transitively applied from a similarity protein-protein BLAST (BLASTp) search of the translated query sequence against the SwissProt database (Altschul *et al.* 1990), if the SwissProt record contains subcellular location information. If there is no hit to SwissProt, or the SwissProt record does not contain subcellular location information, then BASys performs a keyword search on the Gene Ontology (GO) field. If the GO process, function or component show hydrolase activity, ribonuclease activity, nucleic acid binding or RNA binding activity, then BASys assigned the subcellular location as cytoplasmic. Similarly, if

the protein name is associated with transcriptional activity then the cell location is assigned cytoplasmic. If a subcellular location fails to be assigned using these searches, transmembrane region annotations are checked for an existing sequence match and assigned membrane. If the afore mentioned searches do not give a match, the subcellular location is predicted using the program PSORT-B version 2.0 (Gromiha and Selvaraj 1998), and if none of the searches match the protein is presumed to be cytoplasmic based on the observation that roughly 80% of all prokaryotic proteins are cytoplasmic. For further break downs of predicted periplasmic and secreted proteins, see Appendix 7.2.2.

2.7.4. Transport

A total 1382 genes are implicated with transport. Numerous components of the ubiquitous transporter families, the ATP-Binding Cassette (ABC) superfamily and the Major Facilitator Superfamily (MFS), are present in BAA-870. MFS transporters are single-polypeptide secondary carriers capable only of transporting small solutes in response to chemiosmotic ion gradients (Pao *et al.* 1998; Walmsley *et al.* 1998). BAA-870 has 84 members of the MF Superfamily, mostly from the phthalate permease and sugar transporter families. There are dozens of families within the ABC superfamily, and each family generally correlates with substrate specificity. BAA-870 transporters include at least 93 members of the ABC superfamily, which includes both uptake and efflux transport systems.

Out of 3071 genes assigned a COG function, 1361 (44.3%) are associated with transport. These include 178 carbohydrate, 237 amino acid, 105 coenzyme, 217 inorganic ion, 402 lipid and 48 nucleotide transport and metabolism gene functions, and 174 secondary metabolite biosynthesis, transport and catabolism genes (see Figure 2.8.).

2.7.5. Metabolism

The central metabolism of BAA-870 includes glycolysis, gluconeogenesis, the pentose phosphate pathway, and the tricarboxylic acid (TCA) cycle, a typical metabolic pathway for an aerobic organism. McLeod *et al.* reported that RHA1 contains genes for the Entner-

Doudoroff pathway (which requires 6-phosphogluconate dehydrase and 2-keto-3-deoxyphosphogluconate aldolase to create pyruvates from glucose). The Entner-Doudoroff pathway is, however, rare for Gram positive organisms which preferably use glycolysis for a richer ATP source, and there is no evidence of this pathway existing in BAA-870.

The complete biosynthetic pathways for all nucleotides, nucleosides and natural amino acids are also contained in the genome of BAA-870. For RHA1, oxidoreductases (1085) and ligases (192) disproportionately represent the major enzyme classes, and within the oxidoreductases the EC classes 1.1.-.-, 1.3.-.-, 1.13.-.-, and 1.14.-.-, (all oxygenases) are overrepresented. RHA1 has 203 oxygenases, including 19 cyclohexanone monooxygenases (EC 1.14.13.22). BAA-870 does not seem to possess a cyclohexanone monooxygenase, however, it does have two phenol 2-monooxygenases (EC 1.14.13.7), a methane monooxygenase (EC 1.14.13.25), three vanillate monooxygenases (EC 1.14.13.82), a precorrin-3B synthase (EC 1.14.13.83) and five phenylacetone monooxygenases (EC 1.14.13.92). Out of a total of 6874 genes annotated using BASys, 1342 genes were assigned an EC number. The proportions of assigned enzymes belonging to the six enzyme classes are shown in Figure 2.12.

In BAA-870 there are 501 oxidoreductases (37% of the enzymes). The oxidoreductases also comprise the major enzyme class; however, ligases (121) are not overrepresented. The oxidoreductases and transferases together make up 62% of the enzyme class fraction of BAA-870. BAA-870 has 59 annotated oxygenases, but the EC classification suggests there are at least 114 oxygenases present within the genome.

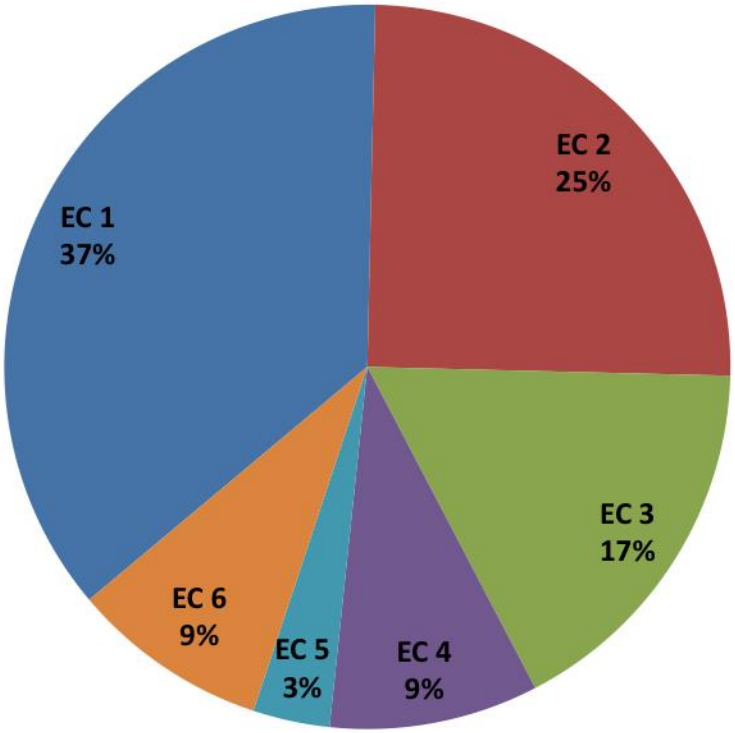


Figure 2.12: Assignment of enzyme class numbers to annotated genes in *R. rhodochrous* ATCC BAA-870.

EC numbers are transitively applied from a similarity (BLASTP) search of the translated query sequence against the translated KEGG GENES database of bacterial genomes (Kanehisa *et al.* 2004).

EC 1, oxidoreductases, make up the major enzyme class (37%), followed by EC 2 transferases (25%), and EC 3 hydrolases (17%). EC 4 and 6, lyases and ligases, both contribute 9% of the enzyme classes, while the EC 5 isomerases are the least at 3% of the total. RHA1 is similarly most rich in oxygenases, followed by transferases and hydrolases, with oxidoreductases making up 11.8% of the total protein coding sequence. The proportions of oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases are not necessarily related to genome size and can be disproportionately represented between related bacteria. While the enzyme class profiles are similar for the *Rhodoccoccus* strains BAA-870 and RHA1, other ecologically and phylogenetically related bacteria have very different enzyme class profiles. In some *M. tuberculosis* strains, for example, ligases are the predominant enzyme class, while for *B. xenovorans* LB400 (which occupies a similar ecological niche to RHA1), oxidoreductases are the most predominant enzyme classe, followed by hydrolases and transferases.

The number of cytochrome P450s in RHA1, 25, is typical of actinomycetes. There are 17 cytochrome P450s in ATCC BAA-870, which mirrors the smaller genome size and their prevalence reflects a fundamental aspect of rhodococcal physiology. RHA1 has a limited ability to grow on linear alkanes, and the predicted monooxygenases able to transform such compounds are few. Similarly, ATCC BAA-870 has one annotated monooxygenase but 46 annotated dioxygenases (59 oxygenases in total). It is unclear which oxygenases in BAA-870 are catabolic and which are involved in secondary metabolism, but their abundance is consistent with BAA-870's ability to degrade an exceptional range of aromatic compounds (oxygenases catalyse the hydroxylation and cleavage of these compounds). Their prevalence reflects a fundamental aspect of BAA-870's, and perhaps other, rhodococcal physiology.

2.7.6. Secondary metabolism

The analysis of the RHA1 genome by McLeod *et al.* suggested that Rhodococci harbour an extensive secondary metabolism. RHA1 contains 24 non ribosomal peptide synthase genes, six of which exceed 25 kbp, and seven polyketide synthase genes (McLeod *et al.* 2006). Environmental isolates, and particularly actinomycetes, contain high numbers of such secondary metabolic genes, which are involved in biosynthesis of siderophores, cell signalling molecules and antibiotics. However, for RHA1, the relatively high numbers of these genes was surprising given that only one secondary metabolite has actually been reported for a *Rhodococcus*. A unique mixed-type catecholate-hydroxamate siderophore, rhodochelin, has been isolated from *Rhodococcus jostii* RHA1 (Bosello *et al.* 2011; Bosello *et al.* 2012). Rhodochelin was revealed to be a tetrapeptide that contains an unusual ester bond between an L-δ-N-formyl-δ-N-hydroxyornithine moiety and the side chain of a threonine residue. The numbers of secondary metabolite genes for BAA-870 are comparably high and consistent with those found in RHA1.

Evidence of an extensive secondary metabolism in *Rhodococcus* ATCC BAA-870 is supported by the presence of several genes implicated in secondary metabolite biosynthesis, transport and catabolism. The ATCC BAA-870 genome has 4 polyketide synthase genes, and 4 regulators of polyketide synthase expression, including an isomerase involved in polyketide

biosynthesis. A total of 23 non ribosomal peptide synthase genes implicated in secondary metabolite synthesis can be found in ATCC BAA-870 (all located in the cytoplasm), and unlike the majority of those from RHA1, all are translated into predicted proteins of >25 kbp. Other secondary metabolite genes found in ATCC BAA-870 include a dihydroxybenzoic acid-activating enzyme (2,3-dihydroxybenzoate-AMP ligase), bacitracin synthase 3, four enterobactin synthase component F genes, ten linear gramicidin synthase subunit C genes, and one tyrocidine synthase 3. The 2,3-Dihydroxybenzoate-AMP ligase activates the carboxylate group of 2,3-dihydroxy-benzoate (DHB), via ATP-dependent PPi exchange reactions, to the acyladenylate. Bacitracin synthase induces peptide synthesis, activates and incorporates five amino acids, forms a thiazoline ring between the first two amino acids and incorporates a D-glutamine in the fourth position. The linear gramicidin synthase activates the 13th to the 16th (Trp, D-Leu, Trp and Gly) amino acids in linear gramicidin and catalyzes the formation of the peptide bond between them. This enzyme is also responsible for the epimerization of the 14th (D-Leu) amino acid. It also catalyzes the NAD(P)H-dependent reduction of the C-terminal glycine residue of the N-formylated 16-mer peptide, that binds to the peptidyl carrier domain of the terminal module of this protein, to form a peptidyl-aldehyde intermediate that is released from the enzyme complex. Tyrocidine A incorporates six amino acids, Asn, Gln, Tyr, Val, Orn, and Leu, in their L-configuration into the peptide product. These genes, along with the high number of transporters potentially involved in drug export, suggest that Rhodococci contain an extensive secondary metabolism.

2.7.7. Aromatic Catabolism

The RHA1 genome is suggested to encode at least 26 peripheral aromatic pathways and 8 central aromatic pathways. As deduced from the better characterized pseudomonads (Jiménez *et al.* 2004), a large number of 'peripheral aromatic' pathways funnel a broad range of natural and xenobiotic compounds into a restricted number of 'central aromatic' pathways.

Analysis of the catabolic potential of BAA-870 against a wide range of compounds suggested that the organism also hosted a number of central and peripheral catabolic pathways for the metabolism of aromatic compounds. Analysis of potential catabolic genes in the genome

suggest that at least four major pathways exist for the catabolism of central aromatic intermediates, comparable to the well-defined aromatic metabolism of *Pseudomonas putida* KT2440 strain (Jiménez *et al.* 2002). The central aromatic beta-ketoadipate pathway contains the protocatechuate (*pca* genes), and catechol (*cat* genes) branches, and the homogentisate (*hmg*/*fah*/*mai* genes) and the phenylacetate (*pha* genes) pathway. Peripheral catabolism pathways include genes encoding the catabolism of p-hydroxybenzoate (*pob*), benzoate (*ben*), quinate (*qui*), phenylpropenoid compounds (*fcs*, *ech*, *vdh*, *cal*, *van*, *acd* and *acs*), phenylalanine and tyrosine (*phh*, *hpd*) and n-phenylalkanoic acids (*fad*). Peripheral and central aromatic catabolism pathways (Table 2.13), unlike the characterized *Pseudomonad* pathways (Jiménez *et al.* 2004), are not well defined in *Rhodococcus*. BAA-870 may therefore serve as a model system for studying the genetic, biochemical and ecological aspects of the catabolism of aromatic compounds, and also facilitate the rational manipulation of the strain for improving biodegradation and biotransformation processes.

Table 2.13: Central and peripheral aromatic catabolism pathway genes encoded by the genome of *R. rhodochrous* BAA-870

Catabolism pathways	Gene name	Genes in BAA-870	Protein name (alternate name)	General function
Central aromatic				
Beta-ketoadipate pathway, Protocatechuate branch[1]	*pca*	pcaG, pcaH	Protocatechuate 3,4-dioxygenase alpha, Protocatechuate 3,4-dioxygenase beta	Plays an essential role in the utilization of numerous aromatic and hydroaromatic compounds via the beta-ketoadipate pathway
Beta-ketoadipate pathway, Catechol branch[2]	*cat*	catD, catI, catJ	3-oxoadipate enol-lactonase 2 (Beta-ketoadipate enol-lactone hydrolase II); 3-oxoadipate CoA-transferase subunit A and B	
		cbdA	2-halobenzoate 1,2-dioxygenase (2-chlorobenzoate 1,2-dioxygenase)	Component of the 2-halobenzoate dioxygenase multicomponent enzyme system; catalyzes the incorporation of both atoms of molecular oxygen into 2-halobenzoate to form catechol
Homogensitate pathway[3]	*hmg*	hmgA	Homogentisate 1,2-dioxygenase (Homogentisic acid oxidase; Homogentisicase)	Implicated in secondary metabolite biosynthesis, transport and catabolism
Phenylacetate pathway[4]	*pha*	hpaB	4-hydroxyphenylacetate 3-monooxygenase (4-HPA 3-hydroxylase)	Utilizes $FADH_2$ supplied by hpaC or another flavin reductase, to catalyze the hydroxylation of 4-hydroxyphenylacetic acid, leading to the production of 3,4-DHPA. Can also oxidize phenol to catechol, and hydroxylate other

				phenol derivatives
Peripheral aromatic pathways				
p-Hydroxy-benzoate	*pob*	pobA, pobB	P-hydroxybenzoate hydroxylase, Phenoxybenzoate dioxygenase	The dioxygenase degrades exclusively diarylether compounds having carboxyl groups in the 3- or 4-position. Yields a hemiacetal that spontaneously hydrolyzes to phenol and protocatechuate
Benzoate	*ben*	benA, benB, benK	Benzoate 1,2-dioxygenase subunit alpha and beta; Benzoate transport protein	Degradation of benzoate to 2-hydro-1,2-dihydroxybenzoate; benK: probable uptake of benzoate
Quinate	*qui*	aroA, aroB, aroD, aroE, aroQ	3-phosphoshikimate 1-carboxyvinyltransferase; 3-dehydroquinate synthase and dehydratase; Shikimate dehydrogenase	Aromatic amino acids biosynthesis, shikimate pathway; aroQ catalyzes a trans-dehydration via an enolate intermediate
Phenylpropenoid compounds	*ech*	echA6, echA8, echA12, echA14, echA17	Probable enoyl-CoA hydratase echA6, echA8, echA12, echA14 and echA17	Could possibly oxidize fatty acids using specific components
	vdh	vdh, ValDH	Vitamin D_3 25-hydroxylase; Valine dehydrogenase	Hydroxylates vitamin D_3 into 25-hydroxyvitamin D_3 and 1-alpha,25-dihydroxyvitamin D_3; Oxidative deamination of branched-chain amino acids
	van	vanA, vanB	Vanillate O-demethylase oxygenase (4-hydroxy-3-methoxybenzoate demethylase), Vanillate O-demethylase oxidoreductase (Vanillate degradation ferredoxin-like protein)	function unknown, reductase implicated in energy production and conversion
	acd	acdA	Acyl-CoA dehydrogenase	Involved in the degradation of long-chain fatty acids
	acs	acsA, acsA1	Acetyl-coenzyme A synthetase; Acetoacetyl-coenzyme A synthetase	Enables the cell to use acetate during aerobic growth alternate to the TCA cycle; Involved in poly-3-hydroxybutyrate degradation. Activates acetoacetate to acetoacetyl-CoA
n-Phenylalkanoic acids	*fad*	fadA, fadB, fadB1, fadB2, fadD, fadD11, fadE10, fadE12, fadE25, fadF, fadH, fadI, fadJ, fadK, fadR	3-ketoacyl-CoA thiolase and acyltransferase, 3-hydroxypropionyl-coenzyme A dehydratase, 3-hydroxybutyryl-CoA dehydrogenase, various long and short chain fatty acid CoA ligases and acyl-coA dehydrogenases, NADH oxidase, 2,4-dienoyl-CoA reductase, 3-ketoacyl-CoA thiolase	Various roles in fatty acid oxidation and esterification, role in autotrophic carbon fixation via the 3-hydroxypropionate/4-hydroxybutyrate cycle. Catalyzes the reversible dehydration of 3-hydroxypropionyl-CoA to form acryloyl-CoA, reversible dehydration of (S)-3-hydroxybutyryl-CoA to form crotonyl-CoA. Esterification Of Exogenous Long-Chain Fatty Acids Into Metabolically Active CoA Thioesters. Reduction of a range of alternative electron acceptors

[1-4]: the four main pathways for the catabolism of central aromatic intermediates.

2.8. Nitrile converting enzymes in rhodococcal genomes

The locations and numbers of nitrile converting enzymes in the available genomes of *Rhodococcus* were identified.

2.8.1. Nitrile converting enzymes in known rhodococcal genomes, and comparison of chromosomal size and organisation

BAA-870 was isolated on a nitrile as the sole source of nitrogen (Brady *et al.* 2004b) and was determined to have nitrile hydrolysis capabilities. However, the genes responsible for the conversions were not known. Now, with the sequence of the BAA-870 genome available, it is evident that there is one nitrilase and one nitrile hydratase, numerous amidases, and a number of regulator elements. The specific open reading frames of nitrile converting enzymes and related regulators and activators found in the BAA-870 genome are summarised in Table 2.14 This result and the nitrile converting enzymes from other sequenced rhodococcal genomes is compared in Table 2.15.

Table 2.14: Nitrile converting enzymes in *R. rhodochrous* ATCC BAA-870

***R. rhodochrous* ATCC BAA-870 Enzyme**	
Nitrilase	1
Nitrile hydratase (L-NHase)	alpha subunit beta subunit beta subunit homolog
Nitrile hydratase regulators/activators	regulator 2 (nhr2) putative regulator 1 putative regulator 1 activator
Amidase	8 putative amidases Amidase (amdA)
Amidase regulators	putative substrates transport protein (amiS) putative aliphatic amidase regulator (amiC)

Table 2.15: Comparison of Nitrile converting enzymes in *Rhodococcus* species

Organism	Nitrilase	Nitrile Hydratase	NHase regulators	Amidase	Amidase Regulators
R. rhodochrous ATCC BAA-870	1	1	4	9	2
R. erythropolis PR4	0	1	4	9	3
R. jostii RHA1	5	2	-	12 2 on plasmids	-
R. opacus B4	3	2	-	14	-
R. equi 103S	0	2	-	41	-
R. equi ATCC 33707	0	2	0	12	
R. erythropolis SK121	4	2	0	8	

None of the nitrile converting enzymes in *R. erythropolis* PR4 are located in plasmids (Sekine *et al.* 2006). Despite the slightly bigger genome size (6.5 Mbp) the numbers of nitrile hydratases (1), NHase regulators (4), amidases (9) and amidase-associated regulators (3) are similar to the numbers for BAA-870. The bigger *R. jostii* RHA1 genome (7.8 Mbp excluding plasmid sequence) contains more nitrilases (6), and in addition to NHase, a few other NHase related enzymes. However, surprisingly, no NHase regulators are annotated. The RHA1 genome contains an abundance of amidases (*ami*A E.C. 3.5.1.4, 9 probable amidases, 11 possible and putative amidases, 2 possible aliphatic amidases, 2 aliphatic amidases, 1 probable formamidase and 1 *N*-acetylmuramoyl-L-alanine amidase) but also no associated amidase regulators have been annotated (McLeod *et al.* 2006). Plasmids pRHL1 and pRHL2 contain amidases, a hypothetical amidase and an *ami*E (E. C. 3.5.1.4) respectively. The RHA1 genome plasmids are the only plasmids in the Rhodococcus genomes sequenced to contain amidases.

The 7.9 Mbp genome of *R. opacus* B4 contains 3 nitrilases, 1 NHase and also an abundance (25) of amidases (annotated as amidases, mycothiol-s-conjugate amidases, N-

acetylmuramoyl-L-alanine amidases, as well as multiple putative amidases). None of the nitrile-converting enzymes in *R. opacus* B4 are found on plasmids.

The two sequenced *R. equi* genomes each have 2 nitrile hydratases, with multiple amidases. The *R. equi* genomes have no nitrilases and no NHase activators and regulators. *R. equi* 103S has multiple amidases, consisting of AmdA amidases, putative amidases and 2 mycothiol-*s*-conjugate amidases, 4 *N*-acetylmuramoyl-L-alanine amidases and 1 aliphatic amidase (Letek *et al.* 2008 and 2010). Similarly, the *R. equi* ATCC 33707 genome contains 2 aliphatic amidases, 12 amidases, 2 mycothiol-s-conjugate amidases, 3 *N*-acetylmuramoyl-L-alanine amidases and 3 N-carbamoyl-putrescine amidases (not published). The unpublished *R. erythropolis* SK121 genome has 4 nitrilases, 2 NHases and a total of 16 amidases, including 2 *N*-acetylmuramoyl-L-alanine amidases and 2 mycothiol-*s*-conjugate amidases.

It is interesting that the genome of *R. opacus* B4 contains fewer amidases than the genome of *R. equi* 103S. The size and structure of the two genomes differ considerably: *R. opacus* B4, an 'environmental' *Rhodococcus*, has a covalently closed genome of 7.9 Mbp whereas the pathogenic *R. equi* 103S has a much smaller, linear 5.04 Mbp genome. The two very different host environments clearly have influenced their genomes: It has to be noted that *R. opacus* B4 was originally isolated from gasoline-contaminated soil and specifically cultured for its tolerance to hydrocarbons and organic solvents (Na *et al.* 2005). On the other hand, *R. equi* grows preferably in the intestine of the host animal (where it parasitizes macrophages) and in manure-rich soil. *R. equi* 103S, both a soil saprotroph and an intracellular parasite, has maintained a niche-adapted genome unlike the large and diverse genome typical of a soil *Rhodococcus* such as *R. opacus* B4.

As recently as 2010, the 103S genome was extensively studied and the genome reported in great detail (Letek *et al.* 2010). No genes with an obvious role in carbohydrate transport were identified in 103S (Letek *et al.* 2010), consistent with the reported inability of *R. equi* to utilize sugars (Bochner 2009; Quinn *et al.* 1994). By contrast, *R. opacus* B4 can grow on carbohydrates (Zaitsev *et al.* 1995) and the genome encodes multiple sugar transporters. The only carbon sources used by *R. equi* 103S were shown to be organic acids and fatty acids (Letek *et al.* 2010). The 103S genome encodes an extensive lipid metabolic network, thus, *R.*

equi seems to assimilate carbon principally through lipid metabolism. Their highly different environmental tolerances are therefore clearly reflected in their adapted metabolisms. More specifically, p*athogenic R. equi* grows in acidic host habitats such as the macrophage vacuole (with a pH ≤ 5.5), and the airways or the intestine of horse (with typical pH values of ~5.5 and ~6.5, respectively (Duz *et al.* 2009; Miyaji *et al.* 2008). It is suggested that multiple amidases may support the bacterium living in an acidic environment through the release of ammonia (van Vliet *et al.* 2003).

2.9. Evolutionary aspects of the genome

2.9.1. Horizontal gene transfer

RHA1 has primarily appeared to have gained its large genome and diverse metabolic capacity through ancient gene duplications and acquisitions. The chromosome has undergone relatively little recent genetic flux as supported by the presence of only two intact insertion sequences, relatively few transposase genes, and only one identified pseudogene. The ATCC BAA-870 genome contains 16 transposase and inactivated derivative genes, one of which is from the IS30 family, a ubiquitous mobile insertion element in prokaryotic genomes.

2.9.2. Evolutionary divergence and Specialization

The similarity of BAA-870 with *M. tuberculosis* may provide insights into the slow growing pathogen. Conserved genes of particular interest include transporters, regulators, catabolic enzymes, cell-wall proteins and 623 proteins of unknown function. Shared transporters include those encoded by the mammalian cell entry (*mce*) genes: both *M. tuberculosis* and *R. rhodochrous* RHA1 have four *mce* gene clusters. BAA-870 has seven *mce* gene clusters, including an extended *mce* gene cluster. In *M. tuberculosis*, MCE proteins are critical virulence factors thought to transport molecules between the bacterium and host.

M. tuberculosis and ATCC BAA-870 also share cell-wall proteins, regulators, and steroid catabolic genes (Yam *et al.* 2010), some of which are required for growth of *M. tuberculosis* in the macrophage.

The largest complete microbial genomes are soil heterotrophs. The complexity of soil is a result of soil biogenesis and decomposition, and the wide range of compounds produced by plants. Microbes are successfully competitive if they have the ability to use diverse compounds simultaneously, and this is likely the underlying cause of selective pressure to keep large genomes with numerous paralogous genes. The large genome of RHA1 is suggested to be the source of strength of catabolic diversity displayed by the organism, which has evolved primarily through ancient acquisition and duplication processes, since there is no evidence of recent gene duplication and it's genome is considered stable (McLeod *et al.* 2006).

The genome size and catabolic diversity of *Burkholderia xenovorans* LB400 is comparable to that of RHA1 (Novo *et al.* 2003). A comparison of the two genomes revealed that ecologically similar bacteria are able to evolve large genomes by different means. RHA1 was shown by different analyses that it has acquired fewer genes by horizontal transfer than the similarly sized LB400 genome, as well as most bacteria characterised to date (McLeod *et al.* 2006). RHA1 appears to have evolved its large genome for simultaneous catabolism of diverse compounds in an oxygen rich environment.

The smaller ~5Mb genomes of the *R. equi* species, compared to the BAA-870 and RHA1 environmental species, are not due to reductive evolutions, but rather due to genome expansion in non-pathogenic species, adapted for host colonisation and competition in a fatty acid-rich intestine of the herbivore host. Many of the potential virulence-associated genes identified in *R. equi* are conserved in environmental Rhodococci or have homologs in non-pathogenic Actinobacteria (Letek *et al.* 2008; Letek *et al.* 2010).

2.10. Identification of future gene targets for biocatalysis applications

Surprisingly, despite the benefits of biocatalysis highlighted, only an estimated 150 biocatalytic processes are currently being applied in industry (Panke and Wubbolts 2005; Zheng and Xu 2011). In addition to the array of biocatalytic enzymes which the analysis of a genome can provide, sequence information is also drastically increasing the scope of novel antibacterial targets. Previously, antibiotic discovery primarily focused on the improvement of existing classes of antimicrobial agents which had a limited set of targets, and these targets were mostly aimed at homologs of *E. coli*. Novel bacterial targets and new classes of antibiotics will be increasingly sought after as the emergence of resistant pathogens continues to rise. A review by Payne *et al.* demonstrated the use of genomics in facilitating characterisation of new antimicrobial targets (Payne *et al.* 2000). New drug discovery approaches facilitated by the analysis of genomes is being increasingly highlighted, and several reviews and articles on the genomics-based approach to drug discovery are available (Amini and Tavazoie 2011; Barrett 2007; Chalker and Lunsford 2002; Payne *et al.* 2001; Pucci 2006; Winter *et al.* 2011). Novel antibiotic target areas commonly highlighted include pathways involved in fatty acid biosynthesis, chorismate biosynthesis, and particularly the range of enzyme targets given by tRNA synthetases and those involved in two-component signal transduction systems. Table 2.16 highlights some of these targets which are found in the genome of BAA-870, and their possible uses in biotechnology. The 'reverse-genetics' approach to identifying novel enzymes, and genome mining of sequence information provides a fast and economical way to select activities of interest. For example, there are at least 35 and 23 various esterases and lipases in the genome of BAA-870, enzymes with immense potential for applications in industry. Rhodococcal genomes are important models for studying actinomycete physiology and can facilitate the exploitation of these industrially important microorganisms.

Table 2.16: Identification of enzymes of biocatalytic interest in *R. rhodochrous* ATCC BAA-870

Enzyme	Enzymatic activity	Applications and/or Example	EC number
Cyclohexanone monooxygenase (CHMO)	Flavoprotein that catalyses a wide range of oxidative reactions, including enantioselective Baeyer-Villiger reactions (Stewart 1998), sulfoxidations (Chen *et al.* 1999), amine oxidations (Ottolina *et al.* 1999) and epoxidations (Colonna *et al.* 2002)	Baeyer-Villiger oxidation of a variety of cyclic ketones into lactones; Two rhodococcal CHMO crystal structures were used by Mirza *et al.* to elucidate the enzyme mechanism and study the adjustable substrate binding pocket which accommodates a diverse range of substrates (Mirza *et al.* 2009)	EC 1.14.13.22
Cytochrome P450 oxygenase	Degradation of substituted aromatics, thiocarbamates and atrazine. Oxidation of a diverse range of organic compounds, including lipids, steroidal hormones and xenobiotics. Oxidation of difficult targets and catalyzes the insertion of oxygen into non-activated C-H bonds	Cytochrome P450 BM3 (CYP102A1) from the soil bacterium *Bacillus megaterium* catalyzes the NADPH-dependent hydroxylation of several long-chain fatty acids at the ω–1 through ω–3 positions (Fulco and Ruettinger 1987). Asymmetric sulfoxidation has been demonstrated by a *Rhodococcus* sp. (Zhang *et al.* 2010a). Uses in chemistry due to direct hydroxylation of aliphatic or aromatic C–H bonds (Chefson and Auclair 2006), activity on organic/ ionic solvents (Chefson and Auclair 2007), and replacement of natural cofactors (Chefson *et al.* 2006)	Monoxygenases EC 1.13.-.- and dioxygenases EC 1.14.-.-
Shikimate-pathway enzymes	3-Dehydroquinate dehydratase (Type II DHQase) and 3-dehydroquinase (Type I DHQase), chorismate synthase and isochorismate synthase	Potential targets for anti-bacterial, anti-fungal and anti-parasitic drugs	Various
Shikimate kinase	Phosphotransferase, produces shikimate-3-phosphate from shikimate, specifically transfers phosphorus-containing groups with an alcohol group as acceptor; catalyses the fifth step in the biosynthesis of aromatic amino acids from chorismate	Aromatic amino acid biosynthesis	EC 2.7.1.71
Shikimate dehydrogenase	Oxidoreductase, reduces dehydroquinase to shikimic acid (shikimate); acts on the CH-OH group of donor with NAD+ or NADP+ as acceptor	Enantiomeric conversion of racemic substrate analogues of 3-dehydroshikimic acid (Bugg *et al.* 1988)	EC 1.1.1.25

Aldose reductase (aldehyde reductase)	Catalyzes the reduction of a variety of aldehydes and carbonyls, including monosaccharides. Broad substrate specificity for both hydrophilic and hydrophobic aldehydes. Primarily known for catalyzing the reduction of glucose to sorbitol.	Inhibitors of aldose reductase are a target for treatment of diabetic neuropathy (Várkonyi and Kempler 2008). Enzymes used as bio-reducing agents in the synthesis of chiral hydroxy-β-lactams (Kayser *et al.* 2005)	EC 1.1.21
D-Hydroxyacid dehydrogenase	A zinc flavoprotein (FAD). Acts on a variety of (R)-2-hydroxy acids	*P. vulgaris* and *P. stutzeri* (Hao *et al.* 2007) used in pyruvate production, the former in an enzyme-membrane reactor	EC 1.1.99.6
Arabinose dehydrogenase	Catalyzes the oxidation of D- or L-arabinose by transferring hydrogen ion to an electron acceptor, such as NAD+, producing D- or L-arabinono-1,4-lactone, NADH, and H+	Production of L-ascorbic acid and secretion into the culture medium by co-overexpression of arabinose dehydrogenase and arabinone-1,4-lactone oxidase in *S. cerevisiae* and *Z. bailii* (Hancock *et al.* 2000)	EC 1.1.1.116 (D-arabinose dehydrogenase) and EC 1.1.1.46 (L-arabinose dehydrogenase)
Xylose reductase	Catalyses the initial reaction in the xylose utilization pathway, the NAD(P)H dependent reduction of xylose to xylitol	Site-directed mutagenesis of recombinant xylose reductase from *P. stipitis* improved ethanolic xylose fermentation by *S. cerevisiae* (Bengtsson *et al.* 2009)	EC 1.1.1.307 (D-xylose reductase)
Naphthalene cis-dihydrodiol dehydrogenase (cis-1,2-dihydro-1,2-dihydroxy-naphthalene dehydrogenase)	Broad substrate specificity toward polycyclic aromatic hydrocarbon dihydrodiols, also acts on cis-anthracene dihydrodiol and cis-phenanthrene dihydrodiol	For production of catechol-like metabolites; Various stereospecific conversions of halogenated dihydroxycyclohexanones have been demonstrated by *P. putida*; enantioselective toluene cis-dihydrodiol dehydrogenase from *P. putida* F1 was used to produce enantiomerically pure (-)-biphenyl cis-(3S,4R)-dihydrodiol and (-)-phenanthrene cis-(1S,2R)-dihydrodiol from biphenyl and phenanthrene, respectively (Parales *et al.* 2000)	EC 1.3.1.29
N-Carbamyl-D-amino acid amido-hydrolase (DCase)	DCase catalyzes the hydrolysis of N-carbamyl-D-amino acids to the corresponding D-amino acids, which are useful intermediates in the preparation of beta-lactam antibiotics	DCase from *Agrobacterium radiobacter* NRRL B11291is used for the industrial production Of D-amino acids (Grifantini *et al.* 1996b; Nakai *et al.* 2000)	EC 3.5.1.77 (3 genes in BAA-870)

Naphthalene dioxygenase	Relaxed substrate specificity; can oxidize almost 100 substrates, including the enantiospecific cis-dihydroxylation of polycyclic aromatic hydrocarbons and the olefin groups of benzocycloalkenes, benzylic hydroxylation, N- and O-dealkylation, sulfoxidation and desaturation reactions	Biotranformation by immobilized enzyme: oxidation of naphthalene to produce optically pure (+)-cis-(1R,2S)-1,2-napthalene dihydrodiol (McIver *et al.* 2008). Degradation of naphthalene by *Pseudomonas* sp. strain NCIB 9816-4. The enzyme has a broad substrate range and catalyzes several types of reactions including cis-dihydroxylation, monooxygenation, and desaturation (Parales *et al.* 2000). Bioconversion of indene to 1-indenol and cis-(1S,2R)-indandiol by *Rhodococcus* sp. I24 (Priefert *et al.* 2004)	EC 1.14.12.12
Epoxide hydrolase	Key vertebrate hepatic enzyme that is involved in the metabolism of numerous xenobiotics, such as 1,3-butadiene oxide, styrene oxide and the polycyclic aromatic hydrocarbon benzo-α-pyrene 4,5-oxide (Fretland and Omiecinski 2000). In a series of oxiranes with a lipophilic substituent of sufficient size (styrene oxides), monosubstituted as well as 1,1- and cis-1,2-disubstituted oxiranes serve as substrates or inhibitors of the enzyme. In vertebrates, five epoxide-hydrolase enzymes have been identified (EC 3.3.2.-)	Enantioselective enzyme from *Pseudomonas* sp. used in the production of chiral substances, e.g. production of (2R,3S)-ethyl 3-phenylglycidate with 95% enantiomeric excess (Li *et al.* 2003). Production of enantiopure (S)-styrene oxide by enzyme in batch kinetic resolution of racemic styrene oxide (Lee *et al.* 2007). Epoxide hydrolases and several isoenzymes are promising tools for generating enantiopure epoxides and diols (Widersten *et al.* 2010)	EC 3.3.2.9 (microsomal epoxide hydrolase) 4 different genes in BAA-870
s-Triazine hydrolase (TrzA)	Dechlorination of triazine compounds and deamination of the structurally related *s*-triazine (e.g. melamine) and pyrimidine compounds (e.g. 4-chloro-2,6-diaminopyrimidine) (Mulbry 1994a)	Dealkylating and Dechlorinating of the Herbicide Atrazine by *Rhodococcus* (Shao *et al.* 1995). Biodegradation of atrazine and related *s*-triazine compounds in a chemical spill (Wackett *et al.* 2002). General microbial degradation of *s*-triazine herbicides (Mandelbaum *et al.* 2008)	EC 3.8.1.-
1-Chloroalkane halidohydrolase (haloalkane dehalogenase)	Acts on a wide range of 1-haloalkanes, haloalcohols, haloalkenes and some haloaromatic compounds	Production of (R)-3-chloropropane-1,2-diol + 2-oxo-propionaldehyde, which are chiral synthons for various chiral pharmaceuticals, agrochemicals and ferro-electroliquid crystals (Dravis *et al.* 2001). Can be used for the biocatalyzation and bioremediation of haloalkanes	EC 3.8.1.5

4-Nitrophenol monooxygenase (npcB and npcA)	Two component enzyme: NphA1 (oxidase component) and NphA2 (flavin reductase component). NphA1 oxidizes 4-nitrophenol into 4-nitrocatechol in the presence of FAD, NADH and NphA2 reduces FAD in the presence of NADH	Degradation of 4-nitrophenol, and other by-products of the hydrolysis of organophosphorous pesticides, by *Rhodococcus* sp. strain PN1 (Yamamoto *et al.* 2011)	EC 1.14.13.29
3-Hydroxy-isobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases	Catalyzes the NAD+-dependent oxidation of 3-hydroxyisobutyrate to methylmalonate semialdehyde of the valine catabolism pathway. Oxidoreductase, specifically acting on the CH-OH group of a donor with NAD+ or NADP+ as acceptor	The enzymes from *P. putida* E23 is specific for the L-enantiomer of hydroxyisobutyrate, and could produce D-3-hydroxyisobutyrate from DL-isomers and was suggested to be useful for the synthesis of certain antihypertensive reagents (Chowdhury *et al.* 2003)	EC 1.1.1.31 (5 genes in BAA-870)
L-Aminopeptidase/D-esterase	Catalyzes the release of N-terminal D and L amino acids from peptide substrates. The group represents rare aminopeptidases that are not metalloenzymes	DmpA from *O. anthropi* hydrolyzes the chromogenic substrate H-d-Ala-pNA and short α-peptides composed of l-amino acids with good efficiencies. 3.4.11.19: D-stereospecific aminopeptidase (Geueke and Kohler 2007). Hydrolases; Acting on peptide bonds (Peptidases); Aminopeptidases; D-stereospecific aminopeptidase	EC 3.4.11.-
Phenoxy-benzoate dioxygenase	Catalyses conversion of 3- or 4-phenoxybenzoate to phenol and protocatechuate via a putative hemiacetal intermediate	Transformation of mono- and dichlorinated phenoxybenzoates (Halden *et al.* 2000)	EC 1.14.14.-
Dihydrofolate reductase	Reduces folate to 5,6,7,8-tetrahydrofolate	DHFR is a reasonable antimicrobial target for *B. anthracis* and that there is a class of inhibitors that possess sufficient potency and antibacterial activity to suggest further development (Beierlein *et al.* 2009; Joska and Anderson 2006). Valid drug target in drug development, especially essential enzyme is a target for development of specific inhibitors for treatment of the biodefense organism: antimicrobial target for *B. anthracis*, homology modelling of inhibitors and growth inhibition assays (Bennett *et al.* 2009)	EC 1.5.1.3 (6 genes in BAA-870))

Beta-xylosidase	Hydrolase activity, hydrolyzing O-glycosyl compounds	Beta-xylosidases increases efficiency of hydrolysis of plant raw materials with high hemicellulose content such as maize cobs by the enzymatic preparation Celloviridine G20x depleted of its own beta-xylosidase. *A. niger*: coexpression of enzyme with *Trichoderma reesei* xylanase II in *S. cerevisiae* allows for degradation of birchwood xylan to D-xylose. *Penicillium* sp. used for bioconversion of xylan-rich plant wastes to value-added products; bioconversion of xylan-rich plant wastes to value-added products (Bengtsson *et al.* 2009; Rahman *et al.* 2003)	EC 3.2.1.37 (xylan 1,4-beta-xylosidase)
4-Hydroxybenzoate polyprenyl-transferase and related prenyl-transferases	General reaction: polyprenyl diphosphate + 4-hydroxybenzoate = diphosphate + 4-hydroxy-3-polyprenylbenzoate Biosynthesis of ubiquinone, attaches a polyprenyl side chain to a 4-hydroxybenzoate ring, producing the first ubiquinone intermediate that is membrane bound. The number of isoprenoid subunits in the side chain varies in different species. The enzyme does not have any specificity concerning the length of the polyprenyl tail, and accepts tails of various lengths with similar efficiency	Aromatic prenylation reactions of 4-hydroxybenzoate by different 4-hydroxybenzoate polyprenyltrnasferases showed different specificities for the chain length of the isoprenoid precursors (Boehm *et al.* 1997)	EC 2.5.1.39
Panthothenate synthetase	Catalyses the ATP-dependent condensation of pantoate and β-alanine	Potential anti-tuberculosis target (Yang *et al.* 2011) Pantothenate (vitamin B5) biosynthesis	EC 6.3.2.1
Dihydroxy-naphthoic acid synthase	Biosynthesis of bicyclic aromatic compounds; synthesis of 1,4-dihydroxy-2-naphthoate, a branch point metabolite leading to the biosynthesis of menaquinone (vitamin K2, in bacteria), phylloquinone (vitamin K1 in plants), and many plant pigments	Synthesis of polynaphthyl derivatives of β-hydroxynaphthoic acid; environmentally friendly synthetic approach to the synthesis of hydroxynaphthoic acids (Furuya and Kino 2009)	EC 4.1.3.36 (2 genes in BAA-870)

Thioredoxin reductase	Only known enzymes to reduce thioredoxin	The enzyme is essential for cell growth and survival, it is a good target for anti-tumour therapy in cancer. For example, motexafin gadolinium (MGd) is a new chemotherapeutic agent that selectively targets tumour cells, leading to cell death and apoptosis via inhibition of thioredoxin reductase (Richards and Mehta 2007)	EC 1.8.1.9 (2 genes in BAA-870)
Superoxide dismutase	Catalyzes the dismutation of superoxide into oxygen and hydrogen peroxide.	The enzyme is a target for development of antiangiogenic and antitumour drugs (Ferrari *et al.* 1989; Omar *et al.* 1992) In agriculture, soaking fish larva in enzyme solution protects fish from 100 ppm paraquat-induced oxidative injury (Ken *et al.* 2003)	EC 1.15.1.1
Hydantoin utilization protein A	Converts D- and L-substituted hydantoins to corresponding N-carbamyl-amino acids. Converts 5-substituted hydantoins to corresponding L-amino acids.	Amino acid biosynthesis	EC 3.58.2.-
Xanthine dehydrogenase	Acts on a variety of purines and aldehydes, including hypoxanthine This enzyme can also be converted into EC 1.17.3.2 by EC 1.8.4.7, enzyme-thiol transhydrogenase, in the presence of glutathione disulfide	This enzyme can also be converted into EC 1.17.3.2 by EC 1.8.4.7, enzyme-thiol transhydrogenase (glutathione-disulfide) in the presence of glutathione disulfide. In other animal tissues, the enzyme exists almost entirely as EC 1.17.3.2, but can be converted into the dehydrogenase form by 1,4-dithioerythritol. Immobilized and stabilized xanthine dehydrogenase for use in organic synthesis (Tramper *et al.* 1979)	EC 1.2.1.37
Aromatic-Ring-Hydroxylating Dioxygenase	Multicomponent 1,2-dioxygenase complexes that convert closed-ring structures to non-aromatic cis-diols. The enzyme complex has both hydroxylase and electron transfer components	Plays a key role in the biodegradation of numerous environmental pollutants, both in the natural environment (via natural attenuation) and in the engineered bioremediation systems	EC 1.14.12.-

Biphenyl dioxygenase	The enzyme from *Burkholderia fungorum* LB400 is part of a multicomponent system composed of an NADH:ferredoxin oxidoreductase (FAD cofactor), a [2Fe-2S] Rieske-type ferredoxin, and a terminal oxygenase that contains a [2Fe-2S] Rieske-type iron-sulfur cluster and a catalytic mononuclear nonheme iron centre. Chlorine-substituted biphenyls can also act as substrates. Similar to the three-component enzyme systems EC 1.14.12.3 (benzene 1,2-dioxygenase) and EC 1.14.12.11 (toluene dioxygenase).	Catalyzes critical steps of the bacterial polychlorinated biphenyl degrading pathway. Biotransformation of flavones by biphenyl dioxygenase produced novel compounds (Kim *et al.* 2003)	EC 1.14.12.18
3-Chlorobenzoate-3,4-dioxygenase oxygenase	Benzoate dioxygenase; enzyme system, containing a reductase which is an iron-sulfur flavoprotein (FAD), and an iron-sulfur oxygenase. Requires Fe2+. catalyzes the dihydroxylation of benzoate to produce 1,2-dihydroxy-cyclohexa-3,5-diene-1-carboxylic acid	Production of (R,S)-1,2-dihydroxycatechol, which is used as a polymerisation monomer; Synthesis of useful chiral chemical products: biotransformation of benzoate to cis-diols (Sun *et al.* 2008)	EC 1.14.12.10
Fluoroacetate dehalogenase	Hydrolytic dehalogenation of fluoroacetate and other haloacetates	Biodegradation of heavily chlorinated compounds resistant to degradation (Pries *et al.* 1994; van Pée and Unversucht 2003)	EC 3.8.1.3
(*S*)-2-Haloacid dehalogenase 4A	Acts on acids of short chain lengths, C2 to C4, with inversion of configuration at C-2. [also EC 3.8.1.9 (R)-2-haloacid dehalogenase, EC 3.8.1.10 2-haloacid dehalogenase (configuration-inverting) and EC 3.8.1.11 2-haloacid dehalogenase (configuration-retaining)]	Detoxification of halogenated herbicides, solvents and other xenobiotic compounds by immobilized enzyme; production of optically active 2-hydroxyalkanoic acids and 2-haloalkanoic acids for chiral synthesis in industry (Barth *et al.* 1992; Diez *et al.* 1996)	EC 3.8.1.2
3-(3-Hydroxy-phenyl)-propionate/3-hydroxycinnamic acid hydroxylase	Catalyzes the insertion of one atom of molecular oxygen into position 2 of the phenyl ring of 3-(3-hydroxyphenyl)propionate (3-HPP) and hydroxycinnamic acid (3HCI)	Aromatic hydrocarbon catabolism	EC 1.14.13.n3
Pigment production hydroxylase	Pigment production acting as an hydroxylase that transforms indole to indoxyl, resulting in the formation of indigo	Flavonoid biosynthesis, novel indigo-producing oxygenase gene found in *Rhodococcus* sp. (Kwon *et al.* 2008)	5

Tryptophan synthase	A pyridoxal-phosphate protein. The alpha-subunit catalyses the conversion of 1-C-(indol-3-yl)glycerol 3-phosphate to indole and glyceraldehyde 3-phosphate. The indole then migrates to the beta-subunit where, with serine in the presence of pyridoxal 5'-phosphate, it is converted into tryptophan. Also catalyses the conversion of serine and indole into tryptophan and water, and of 1-C-(indol-3-yl)glycerol 3-phosphate into indole and glyceraldehyde phosphate (the latter reaction was listed formerly as EC 4.1.2.8). In some organisms, this enzyme is part of a multifunctional protein, together with one or more other components of the system for the biosynthesis of tryptophan [EC 2.4.2.18 (anthranilate phosphoribosyltransferase), EC 4.1.1.48 (indole-3-glycerol-phosphate synthase), EC 4.1.3.27 (anthranilate synthase) and EC 5.3.1.24 (phosphoribosylanthranilate isomerase)].	The beta subunit is responsible for the synthesis of L-tryptophan from indole and L-serine. The enzyme is a target for structure-based design of herbicides (Sachpatzidis *et al.* 1999)	EC 4.2.1.20
Penicillin V acylase/ Isopenicillin N synthase	The enzyme catalyzes the degradation of conjugated bile acids in the mammalian gut. Removes, in the presence of oxygen, 4 hydrogen atoms from delta-L-(alpha-aminoadipyl)-L-cysteinyl-D-valine (ACV) to form the azetidinone and thiazolidine rings of isopenicillin	Penicillin biosynthesis	3.5.1.24 and 1.21.3.1 1 and 3

Specific enzyme activities for most enzymes were inferred using information obtained from the BRENDA database (www.brenda-enzymes.org), the main collection of enzyme functional data available to the scientific community (Scheer *et al.* 2011; Schomburg *et al.* 2002).

2.11. A Baeyer-Villiger enzyme tool box and rhodococcal Shikimate enzymes for drug discovery

Several Baeyer-Villiger enzymes and Shikimate pathway enzymes are found in the genome of BAA-870 (see Table 2.15). Baeyer-Villiger enzymes are interesting biocatalysts due to their ability to catalyse several types of oxidation reactions, and particularly, Baeyer-Villiger monooxygenases even more so due to their unique ability to insert an oxygen into the carbon-carbon bond (Kamerbeek *et al.* 2003; Kayser 2009). A wide range of ester or lactone products are afforded from the oxidation of carbonylic substrates (Orru *et al.* 2011). Five annotated BVMOs are found in the genome of BAA-870, including three annotated as phenylacetone monooxygenases. Cyclohexanone monooxygenase and several other monooxygenase implicated in Baeyer-Villiger reactions are also present. The range of kinetic resolutions that can be achieved by Baeyer-Villiger monooxygenases (BVMOs), and the broad applicability of BVMOs in synthetic chemistry, provides an ideal enzymatic candidate for a BVMO tool box. Several shikimate pathway enzymes have been found in the BAA-870 genome. All of the enzymatic steps in the shikimate pathway have potential for drug design targets, including the chorismate converting enzymes chorismate synthase and salicylate synthase.

2.11.1. Baeyer-Villiger enzymes

Five annotated BVMOs are found in the genome of BAA-870, including three annotated as phenylacetone monooxygenases. Baeyer-Villiger oxygenases are known for the conversion of cyclic ketones into lactones. BVMOs, in particular, are flavoenzymes and belong to the oxidoreductase enzyme class (1.14.13.x). BVMOs also convert aromatic, linear or aryl alkyl ketones into esters, using molecular oxygen rather than the peracids which are typically used in chemical synthesis (Baeyer and Villiger 1899).

BVMOs were first isolated more than 30 years ago, however, detailed knowledge of these enzymes had been lacking until recently. Major breakthroughs were made in the structure-function relationship of BVMOs, assisted by the elucidation of two BVMO crystal structures

which were isolated from an environmental *Rhodococcus* strain, bound with FAD and NADP(+) in two distinct states (Mirza *et al.* 2009), as well as the identification of numerous novel BVMOs (Torres Pazmiño *et al.* 2010). This has led to an intensified exploration of biocatalytic properties of novel BVMOs and structure-inspired enzyme redesign. Baeyer–Villiger reactions afford high regio-, chemo-, and enantioselectivity, and the enzyme-mediated reactions give valuable advantages over metal-based catalysts used in chemical oxidations (Strukul *et al.* 1997). It was also previously shown that linear aliphatic ketones bearing an amino substituent in the β position are also accepted as substrates by BVMOs.

Rehdorf *et al.* recently exploited the regioselectivity of BVMOs for the formation of β-amino acids and β-amino alcohols (Rehdorf *et al.* 2010). The potential of biologically active β-peptides in particular has great potential for peptidomimetic drug development. Chemical methods for the preparation of β-amino acids have been developed, with a focus on asymmetric synthesis (Gutiérrez-García *et al.* 2001; Juaristi *et al.* 1999; Liu and Sibi 2002), however, the syntheses are often not as selective as enzymatic syntheses (Chhiba *et al.* 2012). The main enzymatic routes for obtaining highly enantiopure β-amino acids have been based on resolutions using enzymes acting on either C-N bonds (such as acylases (Gröger *et al.* 2004) amidases (Li *et al.* 2007; Soloshonok *et al.* 1995) and aminopeptidases (Heck *et al.* 2009), or C-O bonds (using hydrolytic enzymes such as lipases (Gedey *et al.* 2001; Solymár *et al.* 2002) and esterases (Liljeblad and Kanerva 2006)). Optically pure β-amino acids can also be obtained from α-amino acids by using aminomutases (Frey and Magnusson 2003; Wu *et al.* 2011) and by reductive amination of ketones with β-aminotransferases (Banerjee *et al.* 2005).

Rehdorf *et al.* investigated whether linear aliphatic ketones with an amino substituent in the β-position could also be accepted as substrates by BVMOs (Rehdorf *et al.* 2010). Interestingly, they found that certain BVMOs incorporated the oxygen on the less sterically hindered carbon atom to furnish the regioisomeric ester **2**, which yielded a β-amino acid after hydrolysis (Figure 2.13). Oxygen insertion at the higher substituted carbon center lead to the "normal" ester **3**, which could be hydrolyzed to a β-amino alcohol. The regioselective insertion of oxygen in BVMO catalyzed oxidations afforded the expected protected β-amino

alcohol, but could also be used to provide access to β-amino acids such as β-leucine. Synthesizing the "abnormal" ester by chemical methods remains challenging, and thus the implementation of a BVMO-mediated ketone oxidation offers a distinct advantage.

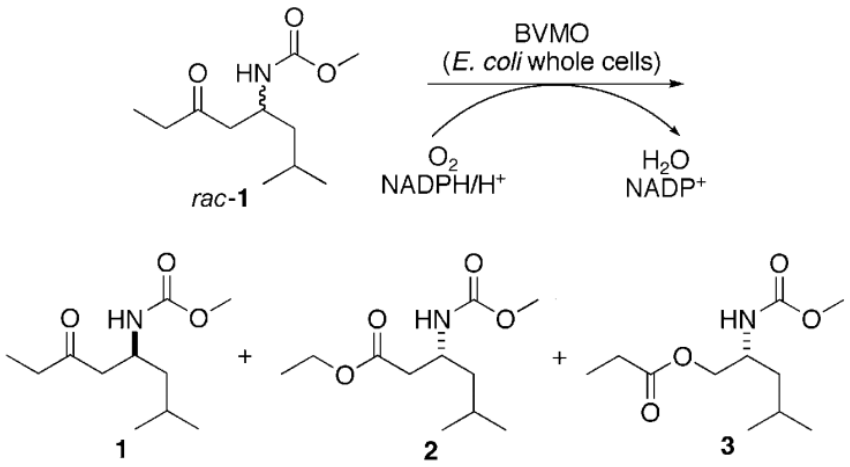


Figure 2.13: The BVMO-catalysed kinetic resolution of aliphatic N-protected β-amino ketone 1 using whole cells of *E. coli* expressing BVMO.

The reaction scheme, from Rehdorf *et al.* 2010, shows the production of the regioisomeric esters 1, 2 and 3 from a β-amino ketone using a BVMO. After hydrolysis, ester 2 yielded a β-amino acid (Rehdorf *et al.* 2010).

Four BVMOs tested by Rehdorf *et al.* were active against 5-amino-3-one **1** and showed high enantioselectivities. Fortuitously, the resulting nonprotected β-amino ketones were not further converted by any of the BVMOs tested.

The enzymatic Baeyer–Villiger reaction opens up new synthetic routes for the formation of β-amino acids, and this offers a useful alternative to already described enzymatic processes (Gedey *et al.* 2001; Gröger *et al.* 2004; Heck *et al.* 2009; Li *et al.* 2007; Liljeblad and Kanerva 2006; Soloshonok *et al.* 1995). The regioselectivity of BVMOs provide pharmaceutically valuable optically active synthons. Baeyer-Villiger monooxygenases can therefore be considered unique oxidative biocatalysts with future industrial applications (Torres Pazmiño *et al.* 2010).

2.11.2. Shikimate enzymology as targets for drug design

While mammals derive their aromatic compounds from their diets, plants and microorganisms produce all key aromatic compounds involved in primary metabolism, including the three aromatic amino acids found in proteins (tryptophan, tyrosine and phenylalanine), via the shikimate pathway (Figure 2.15). Shikimate pathway enzymes have therefore been a focus as targets for anti-microbial compounds, antiparasitic agents, fungicides, antibiotics, and non-toxic herbicides (inhibition of the shikimate pathway has potential herbicidal activity) (Lu *et al.* 2013). For example, the herbicide Glyphosate targets the shikimate pathway. Recently, Koma and colleagues metabolically engineered an *E. coli* shikimate pathway to produce six aromatic compounds used in several industries (Koma *et al.* 2012).

Shikimate (Figure 2.14) is the anionic form of shikimic acid, a metabolite of plants and microorganisms. Shikimic acid is a precursor, via the shikimate pathway, of the aromatic amino acids tyrosine and phenylalanine, and tryptophan amongst other indole derivative, as well as alkaloids, flavonoids, vitamin K and ubiquinone.

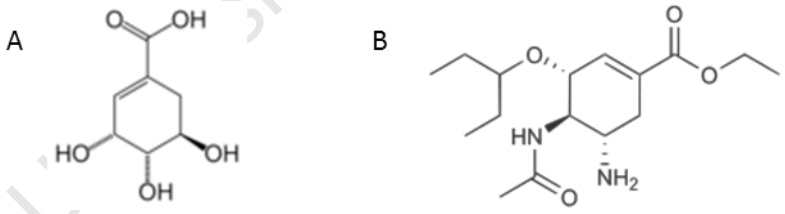


Figure 2.14: The structures of Shikimic acid (A) and oseltamivir (B), trade name Tamiflu®.

Shikimic acid was first isolated from the Japanese flower, Shikimi. In the pharmaceutical industry, shikimic acid is used as the base material in the production of oseltamivir, or the commercially named Tamiflu, an anti-influenza drug marketed by Hoffman-La Roche (Roche). Shikimic acid is extracted from plant material, mostly Chinese star anise, but also from the needles of certain pine tree strains and sweetgum fruit seeds. Dependence of shikimic acid extraction on plants has had major drawbacks, and the shortage of Chinese star anise in 2005 led to a global shortage of Tamiflu®. There is scope for developing shikimic acid

expressing systems using recombinant technology, either by modifying existing metabolic pathways in an organism or using recombinant shikimate enzyme cascades for bioconversions to the valuable shikimic acid

Dehydroquinase (DHQase) catalyses the dehydration of 3-dehydroquinic acid to 3-dehydroshikimic acid (dehydroshikimate). The occurrence of two distinct forms of DHQase offers the possibility of developing selective inhibitors for each type (Coggins *et al.* 2003). Dehydroshikimate is consequently converted to shikimic acid by shikimate dehydrogenase.

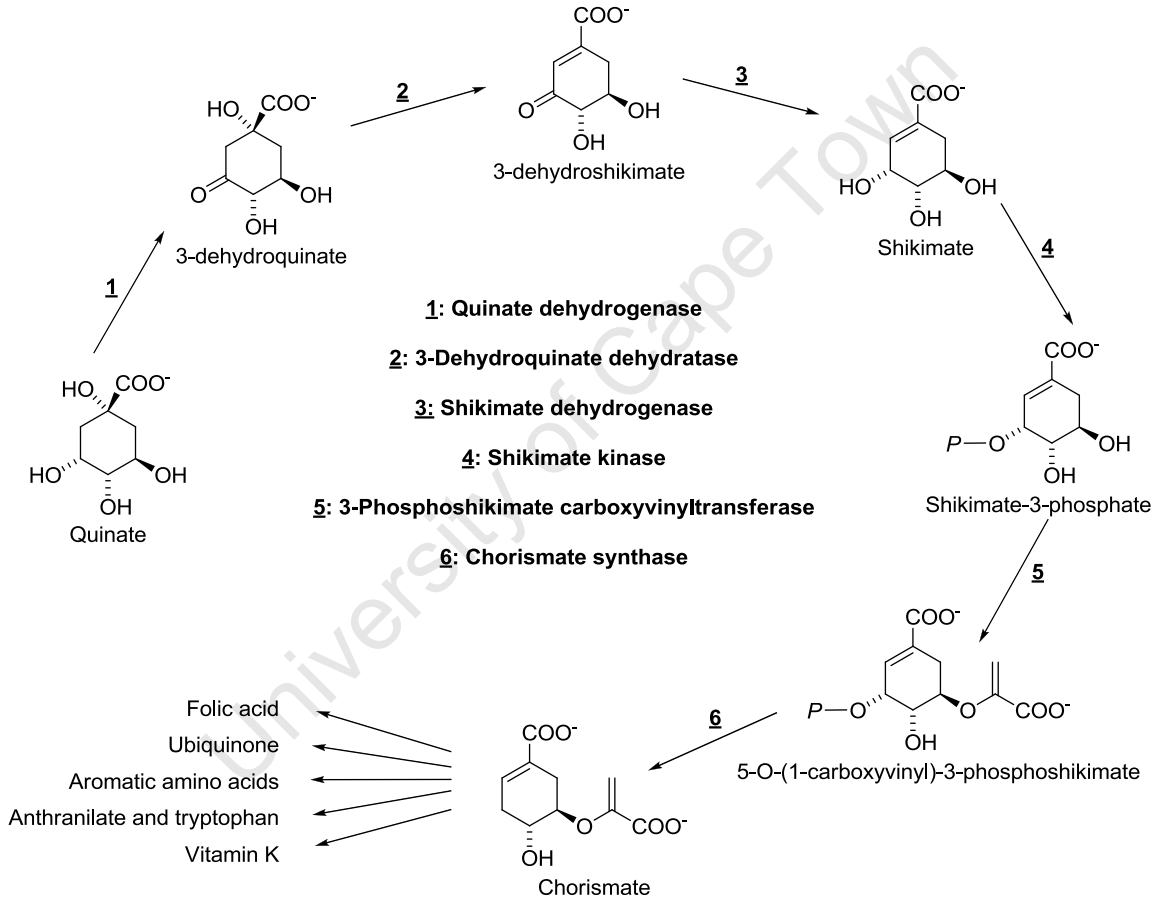


Figure 2.15: A simplified view of a part of the biosynthetic shikimate pathway.

NAD(P)/H and ATP/phosphate transfers have been left out for clarity. The pathway was drawn using ChemDraw® v8.0.

The nutrient *p*-aminobenzoic acid, a precursor of folate, appears to be critical in the biosynthesis of nucleotides by pathogenic bacteria (Payne *et al.* 2000) that rely on their own biosynthetic capabilities on host infection. Elucidation of structures of the shikimate

enzymes, and various enzyme-ligand complexes will allow further progression of rational inhibitor design.

2.12. Discussion

There are many reasons to sequence a genome. The most evident would appear to be the mining of genome information for the benefit of human health, since there has been a bias towards sequencing the genomes of pathogens. New sequencing technology has revolutionized the cost and pace of obtaining genome information, however, there has been a drive to sequence the genomes of organisms which have useful economic consequences, as well as those with environmental interest, and which indicate evolutionary relationships and/or ecological diversity (Martiny and Field 2005; Venter *et al.* 2004).

2.12.1. Pulsed field gel electrophoresis

Pulsed-field gel electrophoresis is recommended for identifying plasmids and sizing the genome, although sizing results vary greatly depending on the restriction enzymes used. PFGE confirmed the presence of a plasmid of ~ 450 kb in addition to the chromosomal genome of *R. rhodochrous* ATCC BAA-870. Sequencing data indicates the presence of genes for the parA protein, which is involved in plasmid partitioning. Pulsed-field gel electrophoresis equipment is not readily available, but the experimental determination of the genome size is a crucial step in confirming the total base pair reads when sequencing a genome *ab initio*, and can experimentally clarify the raw and effective coverage depth obtained from short read sequences. Since increasing the time of incubation with enzyme did not increase the digestion of the genomic DNA, it is presumed that the current available set of restriction enzymes is not optimal for restriction of the genomic DNA of BAA-870. As more restriction enzymes become available, and the sequence specificities of these enzymes become more varied, it will be possible in future to confirm a genome size with better accuracy. However, it is still useful for determining the presence of smaller plasmids.

2.12.2. Solexa Illumina sequencing

The basic steps in processing sequence data include assembling individual 'read' sequences (36 bp in length from Solexa Illumina sequencing) into longer contigs (contiguous sequences). These contigs constitute a 'draft' genome assembly, in which 'gaps' between contigs can be closed through iterative procedures. The draft and completed genomes can be annotated using automated pipelines which predict and assign potential genes or coding sequences, and using a variety of functional resources can determine the functional roles of the predicted genes. Information obtained from a genome dataset includes gene coordinates, locus identifiers and names, and functional annotations such as metabolic pathways, protein families and domains, cellular activities, structures and locations, and associated protein clusters.

Short read sequence (SRS) technology provides sequences faster and cheaper at the risk of fragmented assembly. However, traditional Sanger sequencing methods still provide better quality data that can be assembled into more complete genomes, and additionally, programs and algorithms cope better with longer sequence reads than shorter ones. An example from a recent review shows that as the read length is decreased, the assembly quality deteriorates (Chaisson *et al.* 2004; Whiteford *et al.* 2005). Whiteford *et al.* observed rapidly decreasing contig size for reads shorter than 50 bp (Whiteford *et al.* 2005). Using Sanger sequence reads of more than 750 bp, Chaisson *et al.* assembled 59 contigs from *Neisseria meningitides* sequence data, with 48 of the assembled contigs having a length of > 1 Kbp (Chaisson and Pevzner 2008). The same assembly using 70 bp reads (as in so-called short sequence read technology) consisted of > 1800 contigs, of which only a sixth were > 1 Kbp. Intermediate read lengths of 200 bp could be assembled into 296 contigs.

One can compensate for the fragmented assembly of short sequence reads, to a limited extent, by increasing the depth of coverage, particularly in relatively small bacterial genomes. However, repetitive sequences cause a break in linking information that no amount of coverage can truly compensate for (Pop *et al.* 2002). In essence, for a sequence repeat of length 'N' or longer, a read of length 'N' cannot bridge this repetition. For a short read length, the consequence is that every repeat in the genome causes an unavoidable

break in sequence assembly. The break in assembly does not improve for shorter genomes, even if the coverage is increased. Paired-end sequences such as the mate-paired sequencing employed in this study can improve the assembly by adding longer sequence reads which scaffold assemblies made from shorter reads, thereby reducing somewhat the complexity of sequence assembly. However, according to a recent review by Pop and Salzberg, mate-paired sequence strategies are still relatively new, and their effectiveness has yet to be determined (Pop and Salzberg 2008). The benefit of mate-paired sequencing, however, is that it produces pairs of sequence reads whose approximate distance within the genome is known. This mate-paired information helps sequencing programs to spatially order the contigs along the genome and correctly resolve some sequence repeats.

2.12.3. Genome Sequence Assembly

The assembly of a bacterial genome can be easily handled using desktop applications and fairly standard computer hardware and resources. The freeware assembly programmes Edena and dCAS can be used in conjunction or separately to assemble short read sequence data. Other programmes such as Velvet, SSAKE and SHARCGS can be used separately or in addition to Edena to assemble genome data sequence, however, they are far more demanding of computer resources. The programme, dCAS, a desktop application for cDNA sequence annotation, was used as a first attempt to annotate the assembled sequence, with a good overall annotation percentage. Genome annotation differs greatly when using cDNA sequence versus raw DNA sequence. Thereafter, a combination of dCAS, CLC Bio and Edena was used to assemble the sequences. The best result obtained using this combination was 2082 contigs, with an N50 value of 5540. In this case the percentage of reads assembled was 73.7 %. This was not necessarily the best means of assembling.

The alternative to assembly of a genome *ab initio* is to use the information of a sequenced genome that is already known. The closely related genome can be used as a guide, onto which the new sequence is assembled. However, as has been noted in this study, while helpful in determining genome organisation, it can be seen that this can also bias the genome organisation of the unknown genome, especially for rhodococcal genomes where

even the most closely related species information available can be drastically different from each other, with greatly differing genome size. Sequence repeats still remain a problem in this case, even if the species are very similar in their genome organisation, since the repeat in an assembly algorithm is treated as redundant data as opposed to real repeated sequence. Theoretical simulations of data assembly are useful in highlighting the shortfalls of assembling short read sequences. The assembly program AMOScmp, a comparative sequence assembly algorithm program, uses an existing genome as a backbone onto which the new sequence is assembled. However, it was shown that this comparative assembly strategy only works best when the two species are more than 90 % identical (Pop *et al.* 2004), as sequencing errors can be tolerated by AMOScmp better when a reference genome is used for assembly.

A 'perfect' assembly program should be able to detect all true overlaps in error-free data and assemble all genome regions for which there is a unique solution. Whiteford *et al.* highlighted the possibilities of a 'perfect' assembler by using the bacterial genome, *Escherichia coli*, as an example (Whiteford *et al.* 2005). Even with 30 bp reads, the approximate length of reads of SRS technology such as Illumina, theoretical assembly of the *E. coli* genome should give 75 % coverage with contigs of > 10 000 bp in length, where 96 % of the genes are successfully assembled within single contigs (and not split across multiple contigs). The longer eukaryotic *C. elegans* genome, however, would produce only a 51 % genome assembly with contigs of > 10 000 bp in length, even if 50 bp reads were used (instead of 30 bp). Short reads would leave significant proportions of the genome un-assembled for longer eukaryotic genome assemblies.

All sequence assembly methods assume that the DNA sequence reads are correct, so certain sequencing errors cause erroneous gene annotations, particularly when errors produce an in-frame stop codon. *Ab initio* sequence annotations are organized around open reading frames and are therefore highly sensitive to the positions of start and stop codons.

2.12.3.1. Paired-end sequence assembly

Mate-paired sequencing, in addition to the Solexa Illumina short read sequence information, provides long range positional information, allowing for highly precise alignment of reads. A genome size difference of nearly 3 Mbp obtained depending on the method of data pre-treatment clearly indicates a need for a standardised approach to genome sequencing and is open to interpretation when assessing the quality of sequence added to genome databases. Indeed, the assembly of the *ab initio* sequence of BAA-870 shows that significantly different sequence assembly results can be reported depending on the initial parameters used to define the assembly (using an unstrict mode and lower minimum overlap size of 18, for example, can show the genome consists of 2206 contigs of an average size 2623 bp, with 73.6% of the total 5 787 574 bp genome assembled). Contrast this with a minimum overlap size of 24 bp in strict mode, for example, and the reported sequence assembly statistics will change to 6014 contigs of an average size 939 bp, indicating that 71.9 % of the 5 647 805 total genome was assembled. The reporting of a genome assembly which is obviously open to interpretation, and addition of the genome information to electronic databases, may benefit from a set of guidelines and/or given parameters so as to standardise somewhat the information which is available to researchers. As the sequencing technology improves and genome sequencing becomes cheaper and faster, the addition of sequence information to public space will drastically increase. The reporting of genome sequence can be regarded as a compounding of potential problems: not only is the raw sequence and assembly thereof open to a range of manipulation which can result in extremes of perceived data quality, but so too is the annotation of that data. Annotation of genome sequence is a crucial step in providing genome information that is useful.

2.12.4. Genome Annotation

It has been assumed that the annotation of prokaryotic genomes is simpler than the intron containing genomes of eukaryotes. However, annotation has been shown to be problematic, with over- or under-prediction of small genes especially, where the criterion used to decide the size of an ORF can systematically exclude annotation of small proteins (Warren *et al.*

2010). Warren *et al.* 2010, used high performance computational methods to show that current annotated genomes are missing 1153 candidate genes which have been excluded from annotations based on their size. These genes do not show strong similarities to gene sequences in public databases indicating that they may belong to gene families which are not currently annotated in genomes. Furthermore, they uncovered ~38 895 intergenic ORFs, currently labelled as 'putative' genes only by similarity to annotated genes; in other words, the annotations are absent. Therefore, Prokaryotic gene finding and annotation programs do not accurately predict small genes, and are limited to the accuracy of existing database annotations. Hypothetical genes (genes without any functional assignment), genes that are assigned too generally to be of use, mis-annotated genes and undetected real genes remain the biggest challenges in assigning annotations to new genome data (Frishman 2007; Galperin and Koonin 2004; Roberts 2004).

On investigating the misannotation levels of publicly available protein sequence databases, Schnoes and colleagues, 2009, found that the manually curated UniProtKB/Swiss-Prot database had the least error in annotation of a model set of selected protein families compared to the primarily computationally annotated databases GenBank NR, UniProtKB/TrEMBL, and KEGG (Schnoes *et al.* 2009). It seems that even the level of mis-annotation in genomes is debateable depending on the analysis used. For example, on examining the annotation of small genomes, two groups suggested that at least 8 % of the functional annotation of these genomes were incorrect (Brenner 1999; Devos and Valencia 2001), while the rate was predicted to be much higher (at least 37 % incorrect) by Devos and Valencia, 2001, if the definition of the rate of error used was changed (Devos and Valencia 2001). Jones *et al.* 2007, modelled the annotation error of genes in the GO database, and estimated that up to 49 % of computationally annotated sequences could be misannotated (Jones *et al.* 2007), highlighting the dangers of error propagation when initial sequences are misannotated. Sequence databases may need to be significantly overhauled in order to correct misannotation and prevent the propagation of errors into more recent sequences.

Automatic annotation of genomes is subject to problems and is error-prone. Annotation can be based on the wrong gene models, including models containing fused genes, missed exons and/or non-existent genes. Annotations may give one spurious similarity hits due to the

transfer of annotation by homology or biased amino acid composition. Collaborative genome publishing is necessary in order to collectively pool resources for big annotation projects and to identify poor and/or incorrectly assigned BLAST results so as not to carry through 'bad' genome annotation.

There remains to be established a standard solution for assembly of sequence data in closely related species, and new genome sequences will continue to appear before the technology and software available to process the data is improved, or a standard data processing flow becomes available. Even with existing programs, the assembly of one set of sequence data can be changed significantly by changing only the input parameters for the program, and no guideline or quality measure is used for the submission of genomes to public databases such as NCBI.

2.12.5. Genomes and biodiversity

A single genome sequence cannot reflect the true biological diversity of a taxon. Several genomes from a wide-spread ocean surface water inhabitant cyanobacterium, *Prochlorococcus marinus*, were sequenced with the aim of elucidating the main difference between the high- and low-light ecotypes (Rocap and al 2003). However, the study was surprising: the different *P. marinus* strains had vastly different gene compliments and a high rate of gene acquisition and loss. The study showed that the core gene set shared by the strains was limited and shrinking, but that the pan genome, or total set of genes represented by a *P. marinus* isolate, is extremely large and expanding (Scanlan *et al.* 2009). The significance of the pan genome has therefore been highlighted, since organisms are assigned the same 'species' due to nearly identical 16S rRNA sequences, yet can have vastly different genomes contents. This is certainly true for the *Rhodococcus* bacteria.

Each new genome sequenced contributes to a richer and more complex collection of diverse microbial information. Even closely related genomes contribute unique information – relatively small differences in genome sequences across species allow detailed evolutionary research on even the smallest of time scales. It also will contribute to more accurate gene predictions and refine the definitions of gene functions assigned. It was calculated by

Tettelin *et al.* during a study on the genomes of seven *Streptococcus agalactiae*, that an average of 33 new genes per genome will be discovered as new genomes are sampled (Tettelin *et al.* 2005). Also, analysis of seven *Escherichia coli* genomes predicted that on average, 441 new genes could be provided by the sequencing of each new isolate, highlighting the power of comparative genomics to identify genes (Chen *et al.* 2006). These discoveries have contributed to the strengthening of the concept of the 'pan-genome' – the fact that the gene pool of many microbial species is far larger than the number of genes found in any single organism. Although a simple concept, comparative genomics is nevertheless an important discovery for the genomics era.

2.12.5.1. Implications of hypothetical genes, genes of unknown function and 'orphan' enzymes

In all sequenced genomes, a large fraction of predicted genes encode proteins of unknown biochemical function, and as much as 15% of 'known' function genes are mis-annotated (Kuznetsova *et al.* 2005). Genes are annotated on the basis of sequence similarity to proteins which have already been annotated (Altschul *et al.* 1997). It's a necessary bioinformatics technique which fails to assign 40-60% of new sequences (Green and Karp 2004). In the Protein Data Bank (PDB), as of July 2011, approximately 1887 three-dimensional protein structures are annotated as 'hypothetical', which amounts to roughly 1 in every 40 entries. Only 6 years ago, approximately 500 three-dimensional protein structures were annotated as 'hypothetical', which amounted to roughly 1 in every 50 entries in the PDB (Pazos and Sternberg 2004). There are more than 1525 enzyme activities which have an assigned Enzyme Commission (EC) numbers, but for which no sequence can be found in sequence databases. Such unassigned sequence enzymes are referred to as enzymes with 'orphan activities' (Karp 2004; Lespinet and Labedan 2006a).

Orphan enzyme activities correspond to the enzyme activities (EC numbers) defined by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), and which are not associated with any amino acid sequences in the major public databases (i.e. UniProt, Enzyme, PDB, IntEnz, KEGG, BRENDA) (Lespinet and Labedan 2006a). The statistics amount to more than 36% of enzymes with known functional EC

activities having no gene sequence. In a search of the Orenza database (orphan enzyme database), 6 out of 22 EC classifications of *R. rhodochrous* known enzymes in BRENDA are considered orphans. Lespinet *et al.* have suggested that orphan enzymes could be a major source of new drug discovery targets in the future (Lespinet and Labedan 2006b).

In any given prokaryotic genome, more than 30% of genes are annotated as 'function unknown'. This percentage for 'hypothetical' genes is even higher in eukaryotes where, for example more than 60% of the genes in *Plasmodium falciparum* are hypothetical (Gardner *et al.* 2002). Large percentages of genes have non-specific annotations, including 'hypothetical' or 'putative' functions, and even for the best characterized model organism *Escherichia coli*, more than 20% of gene functions are unknown (Liang *et al.* 2002).

Understanding the functional content of the genome bridges the gap between the explosive amount of genome data added to our knowledge, and the uses obtained there from. Just over 10 years ago, Peer Bork made the "70% hurdle" prediction, which still stands true today: on average, for approximately one-third of the genes in any given genome, the functions could not be predicted through traditional methods of genome analysis, and even more so, the accuracy of functional prediction was only 70% for the remaining genes (Bork 2000). It was hoped that high throughput analytical methods would improve this 70% barrier, but the high error rates endemic to these methods did not improve this figure much (Bork 2000). Using high throughput analysis of the gene and protein expressions, interactions and ligand binding can increase the amount of data for a gene in a model genome, but, the amount of data obtained for each gene still did not necessarily translate into better understanding of the gene function (Jensen *et al.* 2009).

Using the example of an oxidoreductase, a single enzyme in this class would use a variety of electron acceptors and substrates, making the functional assignment of this enzyme difficult. High throughput enzyme assays help to assign general biochemical functions to products of previously uncharacterized genes but are often unable to pinpoint the natural substrates for enzymes (Kuznetsova *et al.* 2005; Kuznetsova *et al.* 2006). The feasibility and merits of a high-throughput approach for identifying oxidoreductases and hydrolases was demonstrated by Kuznetsova *et al.* (2005): They used recombinant expression and affinity purification to

purify 96 proteins in three hours, providing enough protein (10-100 mg in 100-150 ml aliquots) for at least 10 enzymatic assays each (Kuznetsova *et al.* 2005). Also, many proteins lack known enzymatic activity and could function exclusively in protein-protein interactions. Functional assignments of the majority of genes are based only on their sequence similarity of their proteins to experimentally characterized proteins in only a few model organisms.

The abundant 'hypothetical' genes are likely to function in processes of translation, transcription and ribosome biogenesis given that most universally conserved genes function in these related roles (Koonin and Wolf 2008). Several 'conserved hypothetical' genes were characterized recently and shown to be involved in the transcriptional modification of tRNA (El Yacoubi *et al.* 2009; Phillips *et al.* 2010). Large scale screens were applied to a set of purified proteins from *Saccharomyces cerevisiae*, and several new enzymes involved on tRNA metabolism were identified (Grayhack and Phizicky 2001; Martzen *et al.* 1999; Phizicky *et al.* 2003; Xing *et al.* 2004).

Philips *et al.* (2010) discovered an amidotransferase involved in the modification of archaeal tRNA. The biosynthesis of the archaeal modified nucleoside is complex, and initially involves the production of 7-cyano-7-deazaguanine, an advanced precursor that is produced in the tRNA-independent portion of the biosynthesis. In this case, a nitrile has the direct addition of NH_3 rather than being hydrolysed to an amide, showing that the reaction was not via a nitrile hydratase as previously thought, but constitutes the previously unknown amidinotransferase enzyme chemistry (Phillips *et al.* 2010). This is the only example of the conversion of a nitrile to a formamidine known in biology, and a new class of amidotransferase chemistry. Given that the amidotransferase function experimentally assigned to a previously 'conserved hypothetical' gene by Philips *et al.* (2010) belongs to the signature amidase superfamily, it is likely that there are not only more signature amidases waiting to be discovered, but also more associated nitrile hydratase functions to be assigned in *R. rhodochrous* ATCC BAA-870.

The genomes of free-living organisms, as opposed to the more streamlined genomes of parasites, saprophytes and symbionts, contain many more 'hypothetical' genes (which are less common than 'conserved hypothetical' genes) (Kolker and al 2005). This is certainly true for the large soil-acclimatized bacterium *R. rhodochrous* ATCC BAA-870, which has adapted

to a huge range of soil conditions. Organisms with 'moonlighting' proteins, that is proteins with more than one function or enzymatic activity, also contribute to the mis-annotation of genes.

At least 75% of known enzymes in any genome, including those from human and *E. coli*, belong to only three E.C. classes: hydrolases, oxidoreductases and transferases (PEDANT database, (Frishman *et al.* 2003; Frishman 2007)). Oxidoreductases and hydrolases alone comprise 40-60% of known enzymes in various genomes (PEDANT database). In the PDB, 85% of the 74 601 structures to date are oxidoreductases, transferases and hydrolases. It is therefore highly likely that many of the unknown proteins in BAA-870 will also contribute to these activities.

It is plausible that a large proportion of remaining uncharacterized genes belong to functions of regulation of gene expression, complex secondary metabolism, and post-translational and post-transcriptional modifications, based on the observations that the fraction of metabolic and regulatory genes increases with genome size (Galperin 2005; Konstantinidis and Tiedje 2004; van Nimwegen 2003). Additionally, another class of uncharacterized gene function in free-living organisms includes detoxification (usually hydrolysis) of potentially hazardous substances and products of a wide variety of metabolic reactions (Galperin *et al.* 2006). This indeed could contribute to the size and retention of a large number of unknown genes in *R. rhodochrous* ATCC BAA-870. The 'house cleaning' genes have been shown to be specifically abundant on aerobic organisms which face the spontaneous oxidation amino acids, nucleotides and other critical cellular components. Cellular house cleaning reactions would mostly include methylation, acetylation and adenylylation, amongst others. Therefore, "It is probably no coincidence that many poorly characterized proteins appear to function as hydrolases" (Kuznetsova *et al.* 2005; Kuznetsova *et al.* 2006). A recently characterized gene which was assigned the 'conserved hypothetical' annotation has now been renamed *msrC*, and encodes a methionine-(R)-sulfoxide hydrolyzing enzyme (an enzyme for coping with a product of methionine oxidation) (Lin *et al.* 2007).

Not all unassigned or hypothetical genes may necessarily encode a functional protein or have a specific cellular function. Proportions of genetic material in microbes and other

organisms are remnants of viral genetic elements, or transient mobile genetic elements which pass through microbial genomes, and the rate of horizontal gene transfer in bacteria is very high (Kent *et al.* 2011). A sequenced genome should be viewed as a temporal snapshot, since the genome is dynamic and may include short-lived genomic elements that are not maintained by selection.

2.12.6. The Significance of *Rhodococcus* genomes

Rhodococcus genomes in particular encode for large numbers of oxygenases (Van der Geize and Dijkhuizen 2004). This includes multiple alkane monooxygenases (genes alkB1–alkB4) (Whyte *et al.* 2002) and alkane hydroxylase homologs (Van Beilen *et al.* 2002). RHA1 has 203 oxygenases, including 19 cyclohexanone monooxygenases (EC 1.14.13.22). *Rhodococcus* cyclohexanone monooxygenases can be used in the synthesis of industrially interesting compounds from cyclohexanol and cyclohexanone. These include adipic acid, caprolactone (for polyol polymers) and 6-hydroxyhexanoic acid (for coating applications) (Thomas *et al.* 2002). Chiral lactones can also be used as intermediates in the production of prostaglandins (Banerjee 2000). The same oxidative pathway is can be used to biotransform cyclododecanone to lauryl lactone, or 12-hydroxydodecanoic acid (Kostichka *et al.* 2001; Schumacher and Fakoussa 1999b). Cyclododecanone monooxygenase of *Rhodococcus* SC1 was used in the kinetic resolution 2-of substituted cycloketones for the synthesis of aroma lactones in good yields and high enantiomeric excess (Fink *et al.* 2011). RHA1 has a limited ability to grow on linear alkanes, and the predicted monooxygenases able to transform such compounds are few.

ATCC BAA-870 also has few monooxygenases predicted to transform linear alkanes. BAA-870 does not seem to possess a cyclohexanone monooxygenase; but it does have several other monooxygenases, including two phenol 2-monooxygenases (EC 1.14.13.7), a methane monooxygenase (EC 1.14.13.25), three vanillate monooxygenases (EC 1.14.13.82), a precorrin-3B synthase (EC 1.14.13.83) and five phenylacetone monooxygenases (EC 1.14.13.92). In BAA-870 there are 501 oxidoreductases (37% of the enzymes) including 46 annotated dioxygenases (59 oxygenases in total).

The actinomycetales order includes many economically and medically important organisms, and *Rhodococcus* is undoubtedly one of the most industrially important actinomycetes genus (Van der Geize and Dijkhuizen 2004) owing to its success as a microbial biocatalyst and a broad range of applications in bioproduction and bioremediation (Banerjee *et al.* 2002; Gray *et al.* 1996; Holder *et al.* 2011; Kobayashi *et al.* 1992a; Larkin *et al.* 2006; Martínková *et al.* 2009; Yam *et al.* 2010).

Since the sequencing of the RHA1 genome, numerous gene targets have proved to be of value in various biotechnology related areas. For example, the genes encoding vanillin dehydrogenase and vanillate O-demethylase in RHA1 were studied by Chen *et al.* to outline vanillin catabolism (Chen *et al.* 2012). Ahmed *et al.* identified two unannotated genes present in the genome of RHA1 which were similar to peroxidase genes of lignin degrading microbes (Ahmad *et al.* 2011). One of these genes was shown to have a significant role in lignin degradation, and could oxidise both polymeric lignin and well as model lignin compounds. Both studies have led to insights into lignin degradation by RHA1 and other actinomycetes, and in the latter case, provided the first detailed characterization of a recombinant bacterial lignin peroxidase. Lignin biodegradation is an important step in biofuel production, and new lignin degrading enzymes could be useful for better biofuel extraction.

The sequencing of several actinomycete and *Rhodococcus* genomes, have revealed that many of them have large and relatively rare linear chromosomal elements, while most other bacterial genomes are circular. The high proportion of linear chromosomes have made the actinomycete genomes interesting models for studying chromosome topology and evolution (Kirby 2011; Larkin *et al.* 2010).

2.13. Conclusions

By the end of June 2011, 1695 genomes representing 1582 bacteria, 113 archaea and 74 eukaryotes were completely sequenced, and deposited in the freely available public nucleotide sequence databases GenBank/EMBL/DDBJ. There are an astonishing 4876 further microbial genomes currently in progress. Genomes are being sequenced at an increasing rate due to advances in sequencing technology and reduced costs of sequencing. Genome information can be utilized for novel sequence mining, and as a resource for biocatalysis targets. The genome sequence of *R. rhodochrous* ATCC BAA-870 is the first *rhodochrous* subtype to be sequenced. It will be the seventh rhodococcal genome to be submitted to the NCBI, but the fifth subtype addition. The generally large and complex genomes of *Rhodococcus* afford a wide range of genes attributed to extensive secondary metabolic pathways which are presumably responsible for the array of biotransformations and bioremediations demonstrated. These secondary metabolic pathways have yet to be characterised, and offer numerous targets for drug design as well as synthetic chemistry applications. A number of potential genes which could be used for further biocatalysis have been identified in the genome of *R. rhodochrous* ATCC BAA-870, including epoxide hydrolases, enantioselective enzymes with a range of chemistries, Baeyer-Villiger enzymes, shikimate pathway enzymes, and even enzymes which may be useful in biofuel production. A substantial fraction of genes in the sequenced BAA-870 genome have unknown function, and these could be important reservoirs for novel gene and protein discovery.

Integrating data from multiple sources is required to analyse a microbial genome, and the genome sequence information can be continuously curated (reviewed, corrected and expanded) in order to optimize the annotation. Improving the functional annotation of 'hypothetical' and 'putative' genes will discover new enzymes. The genome of *R. rhodochrous* ATCC BAA-870 represents a microbe of considerable biotechnological interest. The BAA-870 genome could provide insights into the *M. tuberculosis* genome, an important pathogen. It may also prove a better streptomycetes experimental system due to its advantage of a faster cell growth and simpler developmental cycle.

The metabolic abilities of the *Rhodococcus* genus will continue attract attention for industrial uses, and further applications of the organism will be identified. Preventative and remediative biotechnologies will become increasingly popular as the demand for alternative means of curbing pollution increases and the need for new antimicrobial compounds and pharmaceuticals becomes a priority. The sequence of the *R. rhodochrous* ATCC BAA-870 genome will facilitate the further exploitation of Rhodococci for industrial uses and provide numerous sources for new biotechnology projects and applications, as well as enable further characterisation of a model biotechnologically relevant organism.

A number of potential genes which could be used for further biocatalysis have been identified in the genome of *R. rhodochrous* ATCC BAA-870, including epoxide hydrolases, enantioselective enzymes with a range of chemistries, Baeyer-Villiger enzymes, shikimate pathway enzymes, and even enzymes which may be useful in biofuel production. A substantial fraction of genes in the sequenced BAA-870 genome have unknown function, and these could be important reservoirs for novel gene and protein discovery.

Most of the biocatalytically useful classes of enzyme suggested by Pollard and Woodley (Pollard and Woodley 2007) are present on the genome: proteases, reductases, nitrilase/cyanohydralase/nitrile hydratase and amidase, transaminase, epoxide hydrolase, monooxygenase, cytochrome P450 and haloperoxidase. Only oxynitrilase (hydroxynitrile lyase) was not detected. Rhodococci are robust industrial biocatalysts, and the metabolic abilities of the *Rhodococcus* genus will continue to attract attention for industrial uses as further biodegradative (Martínková *et al.* 2009) and biopharmaceutical (Yam *et al.* 2011) applications of the organism are identified. Preventative and remediative biotechnologies will become increasingly popular as the demand for alternative means of curbing pollution increases and the need for new antimicrobial compounds and pharmaceuticals becomes a priority.

The sequence of the *R. rhodochrous* ATCC BAA-870 genome will facilitate the further exploitation of Rhodococci for industrial uses and provide numerous sources for new biotechnology projects and applications, as well as enable further characterisation of a model biotechnologically relevant organism. The genome sequence data has been deposited

to GenBank, and given the BioProject accession number PRJNA78009. The genome sequence of *R. rhodochrous* ATCC BAA-870 is the first *rhodochrous* subtype to be sequenced. It is the seventh rhodococcal genome to be submitted to the NCBI, and the fifth subtype addition.

While the limitations of current sequencing technology are known and debated, the point at which the genome assembly and annotation quality is acceptable has to be decided. There are currently no known guidelines existing for the presentation of bacterial genome sequences, and indeed, the focus on sequencing data quality guidelines for the human genome appears to remain a priority in the sequencing research field. The exponential increase in submission of new bacterial genomes means that there is also with it an increase in the number of poor quality and poorly annotated genomes submitted. At what level is a genome good enough to be submitted, or even acceptable to consider it finished? This point may be debated, but, for now it is safe to say that the submission of genome data can only be as good as the programs used to generate the assemblies and annotations. Assembly and annotation of genome sequence data with even the best algorithms will still require many hours of manual editing and human insight, at the risk of adding human error by incorrect assumption. We can rely on scores of data quality, such as the N50 value, and sequence coverage, as an estimate of the completion and quality of the genome sequence, but databases do not require the reporting of these values when sequence is submitted. Perhaps the real test of quality will be determined by the experimental success achieved when the genome sequence is used as the basis for further scientific study.

Chapter Three

3. Nitrilase

3.1. Introduction

3.1.1. Nitrilase reaction, superfamily, mechanism and known structures

Nitrilase enzymes form part of a large and diverse superfamily of thiol enzymes that are found in all animals, plants, fungi and bacteria (Brenner 2002). Nitrilase enzymes (E.C. 3.5.5.1) catalyse the direct hydrolysis of aliphatic and aromatic nitriles into the corresponding carboxylic acid and ammonia (Equation 3) using a Cys-Glu-Lys catalytic triad.

$$\text{R-CN} + 2\text{H}_2\text{O} \rightarrow \text{R-COOH} + \text{NH}_3 \qquad \textbf{Equation 3.}$$

Some organisms may have multiple nitrilases from more than one of the thirteen superfamily branches (Figure 3.1), and have been classified according to sequence similarity and domain presence (Brenner 2002; Pace and Brenner 2001). Figure 3.2 shows the four types of reactions carried out by the nitrilase superfamily branches.

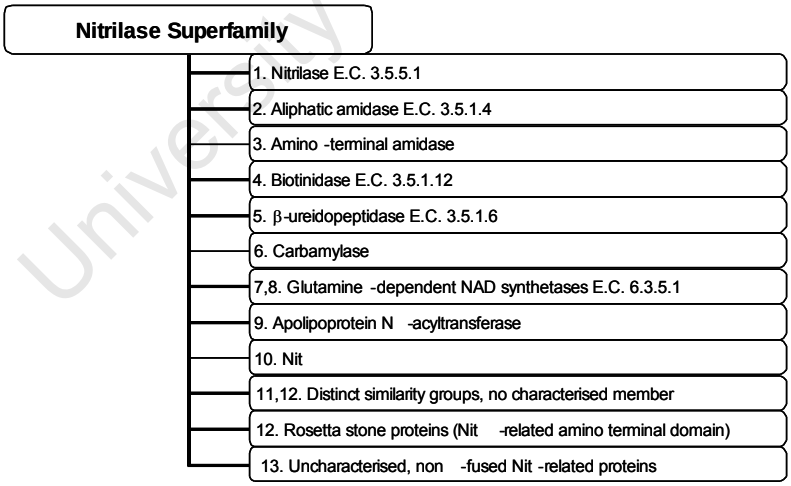


Figure 3.1: Overview of the nitrilase superfamily of enzymes, showing the 13 branches.

Nitrilases usually occur as homooligomers, and very infrequently monomers, and do not contain metal cofactors. The structures of other members of the superfamily, *N*-carbamyl-D-amino acid amidohydrolase from *Agrobacterium* and worm NitFhit, branch 6 and 10

enzymes respectively, were independently shown to consist of a tetramer of compact α-β-β-α sandwiches in which two sheets of six β-strands form layers between pairs of α-helices. The superfamily likely utilizes a novel glutamic acid, lysine and cysteine catalytic triad (Brenner 2002). Both the lysine and glutamic acid residues have been shown to be catalytically essential by construction of stable, inactive mutants with K134N and E59Q substitutions, respectively (Novo *et al.* 2002). *Rhodococcus* nitrilase characterisation has indicated the protein is a thiol enzyme that proceeds via a covalent intermediate (Stevenson *et al.* 1990), and mutagenesis has identified the cysteine nucleophile (Kobayashi *et al.* 1992c), as well as confirmed the essentiality of the corresponding cysteine residue in other nitrilase superfamily branches (Brenner 2002; Grifantini *et al.* 1996a).

The nitrilase superfamily enzymes serve diverse roles in nature, including synthesis of hormones and other signalling molecules, vitamin and coenzyme metabolism, protein post-translational modification, as well as roles in the detoxification of small molecules (Kobayashi *et al.* 1997; Kobayashi *et al.* 1998). There is significant sequence similarity between bacterial and plant nitrilases (Kobayashi *et al.* 1992b). Nitrilases in some plants catalyse the conversion of indole-3-acetonitrile into the plant hormone indole-3-acetic acid, and are involved in nitrogen recycling and catabolism of cyanogenic glycosides (Piotrowski 2008). Nitrilases have also been suggested to play a role in endogenous glucosinolate metabolism in plants (Janowitz *et al.* 2009), and may be found in fungi and bacteria that can metabolise nitrile-containing herbicides. Nitrilase 1 homologs were suggested to have evolved to function in glucosinate metabolism in the plant family Brassicaceae, since they display broad substrate specificities and can hydrolyze nitriles from the decomposition of typical glucosinolate secondary metabolites (Janowitz *et al.* 2009). The nitrilase 4 homolog is typically involved in cyanide detoxification, converting beta-cyanoalanine into asparagine, aspartic acid and ammonia.

Nitrilase (*nit*) gene products have been described as forming natural fusion proteins with the nucleotide-binding protein fragile histidine triad (Fhit). Fhit, an enzyme known to be crucial in apoptosis, is encoded as a fusion protein with Nit in invertebrates, and in mice *nit1* and *fhit* genes have almost identical expression profiles. The naturally occurring fusion protein, NitFhit (nitrilase-fragile histidine triad fusion protein) has recently had its crystallographic

structure solved by Pace *et al.* (Pace *et al.* 2000). The fusion of the two enzymes implies a functional, but as yet unknown, link of two separate pathways.

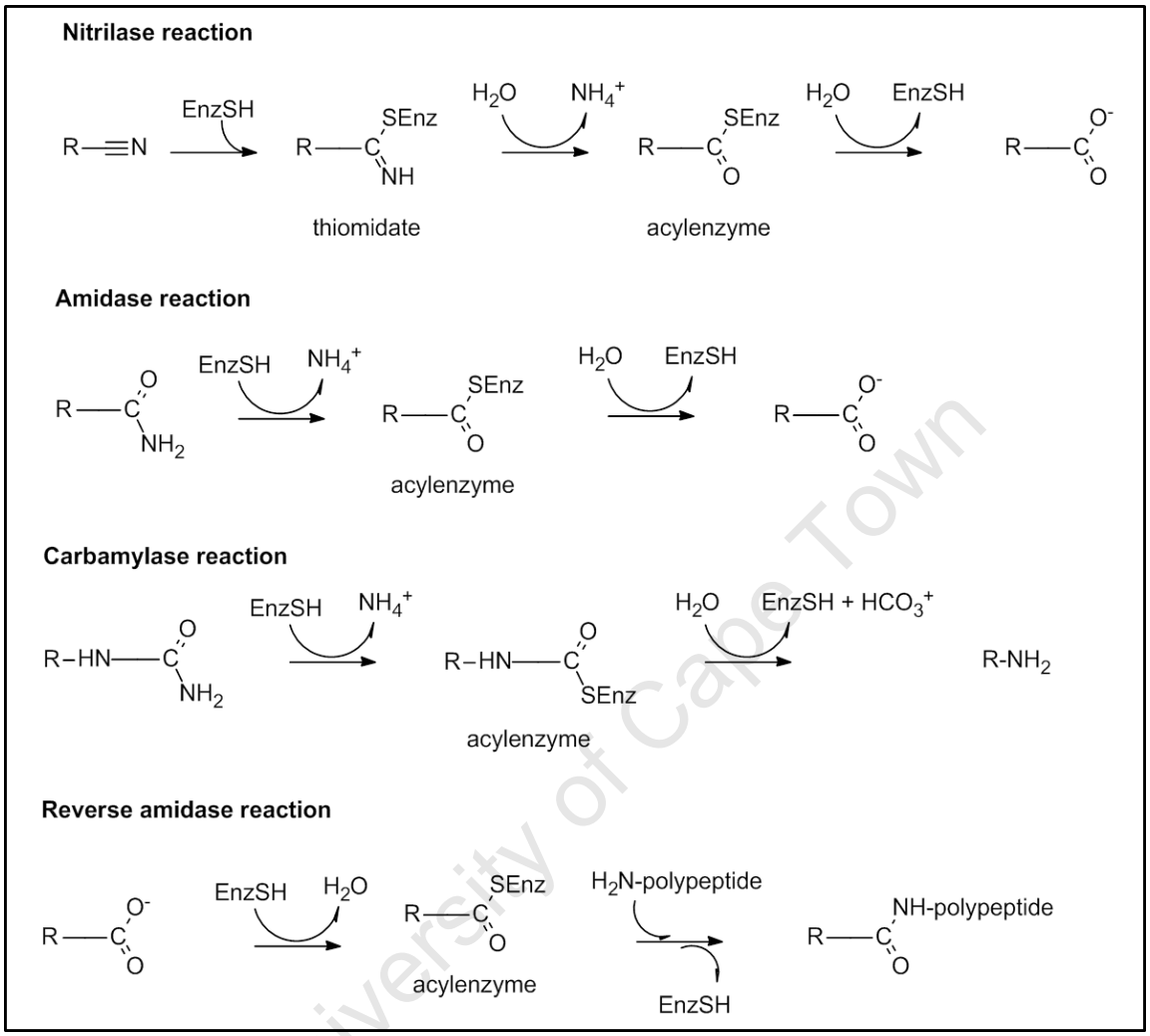


Figure 3.2: The four reaction types carried out by members of the nitrilase superfamily.

On the basis of sequence similarity and the presence of additional domains, the superfamily is classified into 13 branches, nine of which have known or deduced reaction specificity. The nitrilase reaction is carried out by branch 1 enzymes. In plants, this includes the production of indole-3-acetic acid from indole-3-acetonitrile. The amidase reaction is performed by branch 2-4 amidase enzymes and the nitrilase-related domains of branch 7 and 8 enzymes which are proposed to be glutamine-specific amidases. Amidase reactions are the most frequently observed activities in the nitrilase superfamily. The carbamylase reaction is carried out by branch 5 and 6 enzymes and is a special case of amidase reaction. The reverse amidase reaction, in which a phospholipid fatty acid is transferred to a polypeptide amino terminus, is performed by branch 9 N-acyltransferases. The amino-terminal domain of branch 9 enzymes is hydrophobic and has proposed functions in post-translational modifications. The reverse amidase reaction occurs in enzymes where the amidase is fused to a transmenbrane domain. All nitrilase-related reactions are thought to proceed through acylenzyme intermediates. The reaction scheme was adapted from Pace and Brenner (Pace and Brenner 2001).

3.1.2. Nitrilase mechanism

Nitrilase catalytic properties have been investigated by several groups. The cDNA for a number of nitrilases has been obtained, and a comparative amino acid sequence analysis reveals the presence of a conserved cysteine residue that is essential for catalytic activity (Bartling *et al.* 1994; Kobayashi *et al.* 1993a). The function of the conserved cysteine may be to act as a nucleophile in analogy to the bound water molecule used for catalysis in the metal-containing nitrile hydratases. The formation of an enzyme-bound intermediate upon the release of ammonia, and the subsequent hydrolysis of this intermediate to restore the cysteine residue, could account for the difference in reaction products from the two distinct types of nitrile-metabolizing enzymes.

The generally accepted nitrilase reaction mechanism (Figure 3.3) involves a nucleophilic attack on the nitrile carbon atom by a sulfhydryl (thiol) group of a cysteine residue of the nitrilase, with a concomitant protonation of nitrogen to form a tetrahedral thiomidate intermediate. The imine undergoes hydrolysis to the corresponding ketone while protonation of the nitrogen atom results in an ammonia by-product. The acyl-enzyme intermediate is hydrolysed by the addition of water and finally the carboxylic acid and regenerated enzyme are released (Gong *et al.* 2012; Kobayashi and Shimizu 1994; Stevenson *et al.* 1990).

A nitrilase isolated from *R. rhodochrous* ATCC 39484 is reportedly converted to an active form by subunit association when incubated with substrate (Stevenson *et al.* 1992) or when in the presence of higher concentrations of enzyme, salt or organic solvent (Nagasawa *et al.* 2000). These conditions may result in a hydrophobic effect that changes the conformation of the enzyme in such a way so as to expose hydrophobic sites which enable subunit assembly and hence enzyme activation.

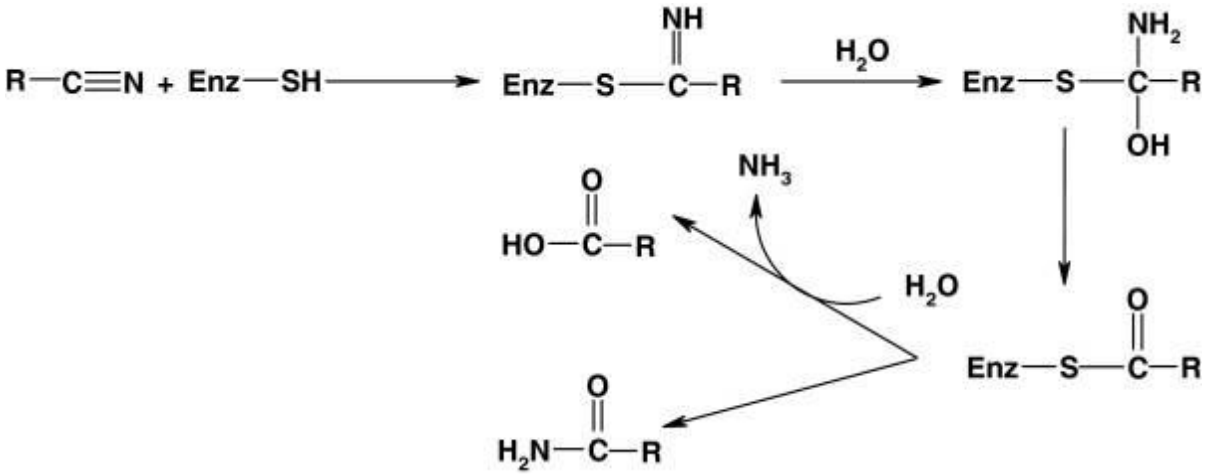


Figure 3.3: A reaction mechanism for a nitrilase reaction requiring a covalent enzyme intermediate.

Figure reproduced from Gong and colleagues (Gong *et al.* 2012).

3.1.3. Biotransformations with nitrilases

Nitrilases, the majority of which are enantioselective (Martínková and Kren 2002), can selectively hydrolyse dinitriles (Effenberger and Oβwald 2001; Kobayashi *et al.* 1998) as well as provide selective nitrile hydrolysis in the presence of labile functional groups (Singh *et al.* 2006). Applications of nitrilases have advanced in recent years due to ongoing studies detailing their structural (O'Reilly and Turner 2003; Thuku *et al.* 2009) and catalytic properties (Martínková *et al.* 2008; Martínková and Kren 2010). Industrial nitrilases are used in several biotransformations, including the manufacture of biologically active enantiomers such as (*R*)-mandelic acid, (*S*)-phenyllactic acid, and (*R*)-3-hydroxy-4-cyano-butyric acid, a key intermediate in the synthesis of Lipitor (the blockbuster drug by Pfizer Inc.) (Banerjee *et al.* 2002; O'Reilly and Turner 2003). Nitrilases are also used in the production of several other high-value compounds, including pyrazinoic acid (Kobayashi *et al.* 1990b), nicotinic acid (Almatawah and Cowan 1999), optically active amino acids (Bhalla *et al.* 1992) and ibuprofen (Yamamoto *et al.* 1990). Several nitrilases are used by DuPont, Lonza, BASF and Mitsubishi Rayon (Brady *et al.* 2004b).

3.1.4. Nitrilase structure

The true nitrilase superfamily branch may have as little as ~20% amino acid sequence identity, with huge variation in their substrate specificities (O'Reilly and Turner 2003). Indeed, cyanide degrading nitrilases/ cyanide dihydratases and aliphatic and aromatic nitrilases have been subclassified into branches of their own: aliphatic and aromatic nitrilases are broad range organic cyanide converting enzymes, whilst the cyanide dihydratases only hydrolyse inorganic cyanide. Cyanide hydratases, however, are a nitrilase subgroup branch of enzymes that can act as pure nitrile hydratases, since they can convert HCN to formamide only (Sosedov *et al.* 2010).

Nitrilases are typically homooligomers with a monomer size of ~40 kDa. The nitrilase monomers associate to form dimers across a conserved $\alpha\beta\beta\alpha$-fold. This association to form oligomeric complexes happens differently in the different nitrilase superfamily members, and nitrilases themselves often form homooligomeric spirals or helices (Thuku *et al.* 2009).

3.1.4.1. Nitrilase crystal structures

There is no atomic resolution structural data for nitrilases. The crystal structures of nine nitrilase-like enzymes have thus far been solved (Lundgren *et al.* 2008). These include, in addition to proteins with unknown function (Chin *et al.* 2007; Kumaran *et al.* 2003; Sakai *et al.* 2004), members of the carbamylase (Nakai *et al.* 2000; Wang *et al.* 2001), Nit and NitFhit (Pace *et al.* 2000) and aliphatic amidase subgroups (Andrade *et al.* 2007a; Hung *et al.* 2007; Kimani *et al.* 2007), and are summarized in Table 3.1.

A nitrilase from a moderately thermophilic bacterium, *Bacillus* sp. DAC521 is a ~41 kDa monomer with ~600 kDa relative molecular mass, suggestive of a ~15-mer complex. The size estimates were proposed to be somewhat obscured, however, due to the co-purification of a putative GroEL homologue (Almatawah *et al.* 1999). The nitrilase from *Geobacillus pallidus* RAPc8 (a gram-positive, moderately thermophilic bacterium) was recently described by Williamson *et al.* (Williamson *et al.* 2010). The nitrilase monomer is 35.8 KDa, with a

functional purified molecular weight of ~600 kDa by size-exclusion chromatography (without GroEL contamination), suggesting the nitrilase from *G. pallidus* RAPc8 is a functional 16-mer.

Table 3.1: Crystallised members of the nitrilase superfamily

Nitrilase member	PDB code	Source species	reference
Nit-fragile Histidine triad fusion protein	1EMS	*C. elegans*	(Pace *et al.* 2000)
Nitrilase 2	2W1V	*Mus musculus*	(Barglow *et al.* 2008)
N-carbamyl-D-amino acid amidohydrolase	1ERZ	*Agrobacterium sp.*	(Nakai *et al.* 2000)
	1FO6	*A. radiobacter*	(Wang *et al.* 2001)
	1UF4; 1UF5; 1UF7; 1UF8	*Agrobacterium sp.*	Hashimoto *et al.* unpublished
	2GGK	*A. tumefaciens*	(Chiu *et al.* 2006)
Amidase	2PLQ	*Geobacillus pallidus* RAPc8	(Agarkar *et al.* 2006) (Kimani *et al.* 2007)
	3HKX	*Nesterenkonia sp.*	(Nel *et al.* 2011)
Formamidase (AmiF)	2DYU; 2DYV; 2E2K; 2E2L	*Helicobacter pylori*	(Hung *et al.* 2007)
	2UXY	*Pseudomonas aeruginosa*	(Andrade *et al.* 2007b)
Putative CN hydrolase	1F89	*Saccharomyces cerevisiae*	(Kumaran *et al.* 2003)
hypothetical protein	1J31	*Pyrococcus horikoshii*	(Sakai *et al.* 2004)
Beta-alanine synthase	2VHH; 2VHI	*Drosophila melanogaster*	(Lundgren *et al.* 2008)
Supposed nitrilase	3IVZ; 3KI8; 3KLC; 3IW3	*Pyrococcus abyssi* ge5	(Mueller *et al.* 2006; Raczynska *et al.* 2010)
Glutamine-dependent NAD+ synthetase	3DLA	*M. tuberculosis*	(LaRonde-LeBlanc *et al.* 2009)
	3ILV	*C. hutchinsonii*	(Palani *et al.* 2009) unpublished

All crystallised superfamily members share a $\alpha\beta\beta\alpha$-fold (average root mean square deviation ~1.5 Å) (Thuku *et al.* 2009). The active site cys-glu_1-lys-glu_2 structural tetrad

residues are highly conserved (Kimani *et al.* 2007) but the N- and C-termini share little sequence or structural homology (Thuku *et al.* 2009). All nitrilase superfamily members interact across a two-fold symmetric dimer interface referred to as the "A-surface" (Sewell *et al.* 2003), forming an αββα-αββα fold with a characteristic extended shape. All members of the superfamily with significant nitrilase activity form helical oligomers (Kaplan *et al.* 2011b; Sewell *et al.* 2003; Thuku *et al.* 2007; Thuku *et al.* 2009; Vejvoda *et al.* 2008; Williamson *et al.* 2010; Woodward *et al.* 2008).

3.1.4.2. Nitrilase structural reconstructions

The structures of four nitrilase superfamily members have been determined at resolutions below 2 Å, but they not form the typical large homo-oligomeric complexes that microbial nitrilases do. The low resolution structures of two cyanide dihydratases were determined, *Pseudomonas stutzeri* AK61 (Sewell *et al.* 2003) and *Bacillus pumilus* C1 (Jandhyala *et al.* 2003), and found to have defined-length spiral structures with 14 and 18 subunits, respectively, under optimal activity (pH 7-8). The formation of the spiral structure is suggested to be essential for activity (Sewell *et al.* 2005). Sequence analysis showed that the nitrilases differ from the nonspiral-forming homologs by two insertions of between 12 and 14 amino acids, positioned at an intermolecular interface in the spiral, and a C-terminal extension of up to 35 amino acids. The insertions are likely to contribute to spiral formation and/or stabilization, and are remote from the active site. Mutations of the interacting residues were found to often lead to a loss of activity and could be as a result of impeded spiral formation (Sewell *et al.* 2005).

Three-dimensional maps of several oligomeric nitrilases (Figure 3.4) have been obtained at low resolution and combined with homology models in order to identify the amino acids potentially involved in interface formation (Thuku *et al.* 2009). In reconstructed low resolution structures of nitrilases, the dimer shape and connectivity can be discerned.

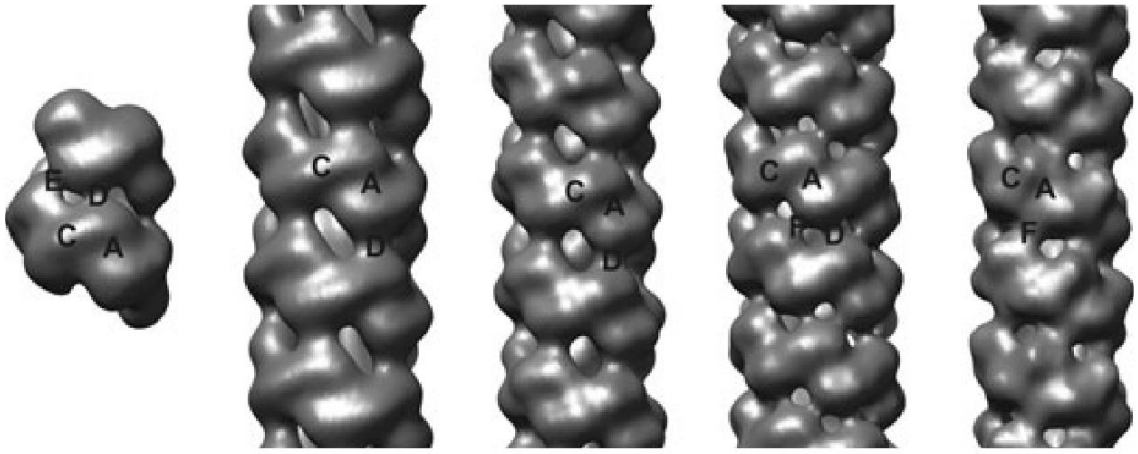


Figure 3.4: A comparative view of 3D EM reconstructed images of spiral forming oligomeric nitrilases studied at low resolution.

The figure is reproduced from a review of microbial spiral forming nitrilases (Thuku *et al.* 2009). The five depicted structures are, from left to right, the self-terminating 14-mer of *P. stutzeri* AK61 cyanide dihydratase (Sewell *et al.* 2003), cyanide dihydratase nitrilase helices from *B. pumilus* C1 (at pH 5.4) (Scheffer 2006) and *R. rhodochrous* J1 (Thuku *et al.* 2007) respectively, and cyanide hydratases from *N. crassa* (Dent *et al.* 2009) and *G. sorghi* (Woodward *et al.* 2008) respectively. The letters A, C, D, E and F refer to the various surface interactions described in the text.

The conserved nitrilase superfamily dimer forms the fundamental unit of oligomerization in spiral nitrilases (Sewell *et al.* 2003). The fibres are left-handed one-start spirals, and all structures have a two-fold axes at the dimer interface (the 'A' surface), which is found in all members of the nitrilase superfamily, in both crystalline and spiral structures. Dimers interact with each other across the so-called "C"-interface. A third interaction, the "D"-interface is formed across the helix after one complete turn of the spiral (Sewell *et al.* 2003). The third interaction may also be referred to as the "F"-interface depending on where it is situated (Woodward *et al.* 2008). An "E"-interface is asymmetric and described only in the cyanide dihydratase nitrilases that form a self-terminating spiral (Sewell *et al.* 2003).

3.1.5. Objectives

Nitrilases have drawn increasing attention for their application in nitrile degradations and the generation of products using green chemistry. Nitrilase biocatalysts can be derived from numerous sources (plants, yeasts, fungi and bacteria). Applications of nitrilase in industry have been successful in the past; however, in-depth structural knowledge and detailed catalytic mechanisms of many nitrilases are rare. Application of nitrilase in industry and further development of catalytic potential will include rational design and more than a superficial understanding of the protein structure.

A nitrilase was identified by sequence mining of the BAA-870 genome, and the further understanding of this nitrilase for use as a future biocatalyst was sought. The objectives of the study therefore included the molecular cloning and sequencing of the nitrilase gene identified in BAA-870, the construction of an expression plasmid for routine over-induction of the protein, and a purification protocol for isolation of nitrilase. Further elucidation of the structure of the BAA-870 nitrilase at the genetic and protein level is explored by crystallography, electron microscopy and modelling, and a comparison of the J1 nitrilase is presented herein.

3.2. Materials and Methods

3.2.1. Reagents

All restriction endonucleases were purchased from Inqaba Biotech, or Novagen. All chemicals were of molecular grade and were purchased from Sigma-Aldrich, unless otherwise stated. DNA polymerase, ligase and other molecular biology enzymes were obtained from Inqaba Biotech. Plasmids and expression strains were purchased from Novagen.

3.2.2. Expression, Purification and Identification of wild type nitrilase from *R. rhodochrous* ATCC BAA-870

See Appendix 6.3 for details of plasmids and vectors used for cloning.

3.2.2.1. Expression of wild-type nitrilase

R. rhodochrous ATCC BAA-870 was grown overnight in tryptone soya broth, and an aliquot (1:100 ml) used to inoculate minimal media containing 5 mM benzonitrile. Cells were grown at 25 – 30 °C overnight, and harvested by centrifugation at 15 000 x *g* for 30 min.

3.2.2.2. Cell disruption

The induced rhodococcal cells were freeze-thawed, and sonicated on ice for a total of 10 min (with 0.5 s pulsed sonication over 20 min). Cell-free extract was collected after centrifugation of cell debris at 15 000 x g for 30 min, and filtered through a 0.22 μm filter.

3.2.2.3. Purification and sequence identification

Cell free extract (50 ml) was applied to a Q-Sepharose anion exchange column (Amersham Biosciences). Equilibration was performed with 100 mM potassium phosphate buffer, pH 7.4, and protein was eluted using a 0.1 to 1 M KCl linear gradient. Fractions were tested for nitrilase activity using an ammonia release assay with benzonitrile as substrate. Pooled ion

exchange fractions showing nitrilase activity were purified by gel exclusion chromatography using potassium phosphate buffer, pH 7.4, containing 200 mM NaCl. Nitrilase-containing fractions were resolved using 12% sodium dodecyl sulphate gel electrophoresis (SDS-PAGE), and the appropriate band was cut out of the gel for further analysis by MALDI-TOF mass spectrometry. The daughter peptide fragments were compared against protein databases using the Basic Local Alignment Search Tool algorithm (BLAST); (Altschul *et al.* 1990), and similar sequences used to determine a strategy for cloning of the nitrilase gene from *R. rhodochrous* ATCC BAA-870.

3.2.3. Nitrilase Activity assays

Nitrile hydrolysis was measured by colorimetric detection of ammonia using the method developed by Fawcett and Scott (Fawcett and Scott 1960). Aliquots (5 μl) of fractions containing nitrilase were reacted at 37 °C for 30 min in 995 μl of 50 mM potassium phosphate buffer, pH 7.4, and 10 μl of a 125 mM nitrile substrate made in ethanol. The total reaction volume was 1.005 ml. The reaction was mixed with 40 μl each of a 1% phenol (made up in 96% ethanol) and 0.5 % sodium nitroprusside solution followed by addition of 100 μl freshly prepared oxidising solution (1 part sodium hypochlorite: 4 parts 20 % trisodium citrate, 1 % sodium hydroxide). The reaction was incubated at room temperature in the dark for 30 min, and the absorbance measured at 620 nm. The absorbance was compared to a standard curve for NH_4Cl prepared between 0 and 0.5 mM. One unit of enzyme activity is defined as the amount of enzyme needed to produce 1 μmol of NH_3 per minute.

3.2.4. Cloning, Expression and purification of the Nitrilase gene from *R. rhodochrous* for expression in *E. coli*

3.2.4.1. Polymerase chain reaction

Primers were designed according to the sequence of *R. rhodochrous* J1 nitrilase (Kobayashi *et al.* 1992c).

Forward primer: 5'-ATGGTCGAATACACAAACACATTCAAAGTTG-3'

Reverse primer: 5'-GATGGAGGCTGTCGCCCGG-3'

Amplification of the nitrilase gene was done with *Pfu* Polymerase (Fermentas, Lithuania) using a PeQLab primus 25 thermal cycler (from Biotechnologie GmbH, Germany). Cycling conditions were as follows: initial denaturation at 95 °C for 2 minutes; followed by 25 cycles of denaturation at 98 °C for 20 seconds, annealing at 60 °C for 15 seconds and extension at 72 °C for 1 min/kb. Control reactions where primers and the DNA template were omitted respectively, were performed.

3.2.4.2. Purification of amplified clone insert

DNA loading buffer (0.2 M Tris-Cl buffer pH 6.8, containing 10% w/v SDS, 20% v/v glycerol, 0.05% bromophenol blue, and 10 mM 2-mercaptoethanol) was added to the PCR reactions. Samples were loaded into wells of a 0.8 % agarose gel alongside PstΛ DNA marker, and run at 100 V. PCR reactions that showed an insert of the correct size were excised from the gel, and purified using a QiaQuick gel extraction kit (Qiagen, Germany). The insert was eluted using 20 mM Tris buffer, pH 8.0.

3.2.4.3. Blunt-end cloning of insert

Purified insert was ligated into pJet1.2 blunt clone vector (Fermentas, Lithuania) using T4 DNA ligase (Fermentas, Lithuania). Ligated insert/plasmid was transformed into *E. coli* EP300 and JM109 competent cells and plated on YT-ampicillin agar plates (10 g/L NaCl, 10 g/L yeast extract, 16 g/L tryptone and 31 g/L agar, pH 7.5, containing 100 µg/ml ampicillin). Purified plasmids (controls and reactions) were digested using *Bgl*II, which cuts on either side of a blunt-cloned insert in pJET1.2. The insert was restricted from the plasmid for directional cloning into an expression vector using the restriction enzymes *Hin*dIII and *Nde*I, purified, and concentrated for ligation. Cloning vectors pET20b(+) and pET28a(+) were purified, restricted with *Hin*dIII and *Nde*I and dephosphorylated. All restriction digestions were analysed by agarose gel electrophoresis.

3.2.4.4. Ligation and Transformation

The amount of vector and genomic DNA required for use in ligation reactions was determined, and ligations performed at a 3:1 insert: vector ratio. T4 DNA ligase (from Fermentas) was used for the ligation reaction at room temperature for 1 hour. Ligated restriction fragment insert-containing vector was transformed into E. cLoni® EXPRESS DUOs BL21(DE3) pLysS chemically competent cells from Lucigen® Corporation (USA). Transformed strain was plated onto nutrient agar plates containing ampicillin, X-gal and IPTG. Plates were incubated overnight at 37 °C. Blue/white screening was used to check for recombinants.

3.2.4.5. Ligation of insert and expression vector

The Nitrilase insert restricted with *Hin*dIII and *Nde*I was ligated with pET20b(+) and pET28a(+) using T4 DNA ligase (Fermentas). The ligated inserts were purified, transformed into *E. coli* EP300 competent cells and plated on LB-ampicillin or LB-kanamycin agar respectively.

3.2.4.6. Analysis of recombinant clones

Confirmation of insert ligation into vector was analysed by restriction digestion using the *Nde*I and *Hin*dIII digestion enzymes to restrict the insert from the plasmids, and by PCR using the original amplification primers. Selected plasmids containing insert were sequenced by Inqaba Biotech (Pretoria, South Africa).

3.2.4.7. Expression of recombinant nitrilase in *E. coli*

E. coli BL21(DE3) or *E. coli* Single Step KRX competent cells (Promega, USA), were cultured at 37 °C with shaking in LB broth supplemented with 37 μg/ml kanamycin. Cultures were grown to an OD at 600 nm of ~0.4 before adding IPTG to a final concentration of 1 mM, or in the case of Single Step cells, L-rhamnose to a final concentration of 0.1 %, and grown for 4 hours at 37 °C or overnight at 25 °C. Cells were harvested by centrifugation (15 min, 8 000 x g) and washed twice in 50 mM potassium phosphate buffer pH 7.4. Cell pellets were subjected to a freeze-thaw cycle before disruption.

3.2.4.8. Cell disruption

Cell paste was resuspended in B-PER bacterial protein extraction reagent (Sigma, Germany) containing lysozyme and DNase I (Fermentas, Lithuania) at 10 mg/ml and 20 μg/ml respectively, and incubated for 30 min at room temperature. The resuspended cells were sonicated using a Misonix sonicator (from QSonica, USA) for a total of 5 min (with 0.5 s pulsed sonication over 10 min). Cell-free extract was collected after centrifugation of cell debris at 15 000 x g for 30 min, and filtered through a 0.22 μm filter (Millipore™, USA).

3.2.4.9. Purification

Cell-free extract was applied either to a Ni-Sepharose 10/16 column (GE Healthcare) or Protino®Ni-TED batch resin (Macherey-Nagel GmbH, Germany) that was equilibrated with 50 mM potassium phosphate buffer pH 7.4 containing 500 mM NaCl and 40 mM imidazole. The column was washed until all unbound protein was eluted. Nitrilase was eluted by a linear salt and imidazole gradient, from 50 mM potassium phosphate buffer pH 7.4, to 50 mM potassium phosphate buffer pH 7.4 containing 500 mM NaCl and 500 mM imidazole. Batch purification of expressed nitrilase was done by applying clarified supernatant to Protino-Ni resin, and washed and eluted using the manufacturer's protocol. Nitrilase was eluted from the resin using 50 mM potassium phosphate buffer pH 7.4, containing 500 mM NaCl and 350 mM imidazole. Nitrilase-containing aliquots were concentrated using a 10 kDa cut-off filtration unit, and 10 mL applied to a Sephacryl S200 (16/100 HR) size exclusion column that was equilibrated with 50 mM sodium phosphate buffer pH 7.4 containing 100 mM NaCl. Fractions of eluted nitrilase were assessed for purity by SDS-PAGE, and protein concentrations were measured using the standard Bradford protein concentration assay (BioRad, USA).

3.2.5. Nitrilase Activity assays

Hydrolysis of nitrile was measured by colorimetric detection of ammonia using the method developed by Fawcett and Scott (Fawcett and Scott 1960) as described above, as well as by HPLC. The HPLC method was used to improve accuracy and involved the following procedure: Nitrilase-containing aliquots (0.1 ml) were added to 0.3 ml assay mix (50 mM potassium phosphate buffer, pH 7.4, containing 5 mM benzonitrile dissolved in a minimal

volume of methanol) and reacted for 5 min with shaking (1000 rpm) at 30 °C. Reactions were stopped by addition of 1.2 ml acidified stop solution consisting of 60:40 acetonitrile : 0.1 %TFA mix, centrifuged for 5 min at 13.2 x g to precipitate particulates, and 1 ml transferred to a 2 ml HPLC vial. Standard concentration curves of nitriles and corresponding acids were constructed between 0 – 5 mM. A Phenomenex Aqua C18 reverse-phased column (Phenomenex, USA), 250mm × 2mm, was used with a flow rate of 0.6 ml/min over 5 min using a 70:30 mixture of acetonitrile : 0.1 % TFA(aq) mobile phase.

3.2.6. Microscopy and Modelling

3.2.6.1. Attempted crystallization of nitrilase

Nitrilase was purified according to the protocol for recombinant nitrilase purification as described previously in Section 3.2.4.9. Nitrilase containing fractions were dialysed with 20 mM Tris-HCl buffer, pH 7.6, containing 50 mM NaCl and concentrated using a 6-8 kDa cut-off membrane (D-Tube™ Dialyzer Mega from Novagen, USA). Aliquots of purified nitrilase were dispensed into 96-well protein crystallization plates using the Mosquito® dispensing platform (TTP LabTech Ltd, U.K.) equipped with a Nanolitre pipettor. Nitrilase (1 - 2uL aliquots) was added to crystallization wells to a final concentration of 0.5, 2 and 3.2 mg/mL containing the relevant buffer screen (Table 3.2), and incubated at room temperature. Crystal formation was monitored using a light microscope.

Table 3.2: Crystallization screens and conditions used for nitrilase

Screen name	Description
JBScreen Basic HTS	No cacodylate, MES instead. Based on the classic sparse matrix crystallization screen, screening a wide range of pH and various salts and precipitants
JBScreen Classic HTS I	PEG based, cover the most prominent conditions for protein crystallization. Their compositions result from mining of literature data of several thousands of crystallized proteins. JBScreen represents the selected statistically most successful buffers that yielded protein crystals suitable for X-ray diffraction.
JBScreen Classic HTS II	Ammonium sulphate, MPD, alcohol and salt based
JBScreen PEG/Salt HTS	screening of PEG 3350 and PEG 5000 MME versus 48 different salts
JBScreen PACT++ HTS	pH, anion and cation testing in PEG, Systematic screen for **p**H, **a**nion, **c**ation **t**esting in the presence of polyethylene glycol
KBScreeb Membrane HTS	for hydrophobic and membrane proteins
JBScreen JCSG++ HTS	Optimized sparse matrix screen developed by the Joint Center for Structural Genomics (JCSG)
JBScreen Pentaerythritol HTS	Systematic crystallization screen based on pentaerythritol polymers as precipitants developed by Ulrike Demmer from the Max-Planck-Institute for Biophysics in Frankfurt.
Stura Footprint Screen	PACT screen, cation/PEG, anion/PEG, pH/PEG

3.2.7. Microscopy of nitrilase

3.2.7.1. Preparation of protein for electron microscopy

Samples were prepared for electron microscopy (EM) using conventional negative staining procedures. A 3 µl drop of purified nitrilase was adsorbed for 30 s to a carbon-coated 400 mesh copper grid that had been glow-discharged at 25 mA for 30 s. The protein sample was washed consecutively with 2 drops of deionised water and stained with 2 drops of 2 % uranyl acetate. Excess water and stain was removed by absorption onto Whatman 597 filter paper.

3.2.7.2. Imaging of nitrilase

Samples were imaged at room temperature using a LEO 912 or Philips Tecnai T20 electron microscope equipped with LaB_6 cathode filaments and operated at an accelerating voltage of 120 kV or 200kV, respectively. All images were taken at a magnification of 50,000x and a defocus value of 1.0 mm using low-dose procedures (<10 $e^-/Å^2$). Micrographs were recorded

using a Proscan Slow Scan 2048x2048 CCD camera and EsiVision Pro 3.2 software from Soft Imaging System (GmbH) or 4k×4k CCD camera (GATAN US4000 Ultrascan, California, USA) for collection on the LEO 912 and Philips Tecnai T20, respectively. All micrographs were visually inspected for astigmatism before collection and only drift-free images were selected for further processing.

Original stack = 13365 images, Scale: 2Å/pixel, Image dimensions: 100 x 100 pixels, Magnification:50K, HT: 200 kV.

3.2.7.3. Picking, filtering and masking

Images were selected and examined using Boxer (Ludtke *et al.* 1999). Ring-like particles were selected manually and windowed in 100 x 100 pixel boxes at a sampling rate of 2 Å/pixel. A total of ~16 000 images were captured. The sample set was analysed manually and filtered to a set of ~10 000 images. All selected particles were filtered and normalised to the same mean and standard deviation.

3.2.7.4. 2D Averaging and Image Processing

The EM software platforms used included SPIDER (Frank *et al.* 1996) and the single particle package SPARX version 1.0 with EMAN2 (Baldwin and Penczek 2007; Hohn *et al.* 2007; Ludtke *et al.* 1999; Tang *et al.* 2006; Tang *et al.* 2007)[3]. 2D Reference-free alignment and classification was performed using a high-pass filter applied to images (to reduce noise at high spatial frequency). Filtered images were aligned (by applying x-shifts only) using a reference-free procedure (AP SR) in Spider. The filtered images were then decimated (binned) by a factor of two and subjected to Kmeans classification using the 'AP C' routine in Spider. Number of groups (500 classes) was specified such that each class had 10-20

[3] **Spider**: http://www.wadsworth.org/spider_doc/spider/docs/techs/recon/mr.html

EMAN2:

http://blake.bcm.edu/emanwiki/Ws2011/Agenda?action=AttachFile&do=get&target=1_3_ludtke_EMAN2.pdf and

http://blake.bcm.edu/emanwiki/Ws2011/Agenda?action=AttachFile&do=get&target=2_4_ludtke_EMAN2.pdf

particles images. Using v2 or Jweb in Spider, class averages were assessed and bad classes (images) were discarded (>50%). The resultant image stack had 6954 images (not decimated). After application of SPIDER, further reference-free alignment was done in EMAN2 using same criteria as above. 348 classes were generated. Bad particles were excluded at this point to produce a final image stack with 5787 images. In total, 57% of image data was discarded (due to bad particles and heterogeneity in the data set). The centered images were then masked for subsequent image processing and 3D reconstruction both in Spider and EMAN2.

The image stack was assessed by manually assessing/editing using the program v2 in EMAN1. Bad particles were manually deleted and a new image stack with ~10000 images was generated using program proc2d in EMAN1. Further image assessment was then done in EMAN2.

Single particles were selected manually using sxboxer in the SPARX suite of programs (Baldwin and Penczek 2007; Hohn *et al.* 2007). Particles were windowed into 100 x 100 pixel boxes. Particles were subjected to rounds of alignment and classification and the Kmeans calculated for 12 – 84 classes. Unique averages were selected from the resulting class averages and used as the references for cycles of multi-reference alignment (Joyeux and Penczek 2002).

Sets of particle images were translationally and rotationally aligned. Alignment was done iteratively over many cycles until overall image shifts and rotations no longer decreased over further alignment cycles. Aligned images were then classified, and each class was averaged. Images in each class were visually compared to the corresponding class average, and inspected to verify that images were assigned correctly to the class. Unique classes were selected from the averages and used as the reference for multi-reference alignment. Images of a data set were cross-correlated with the references and assigned to the reference that yielded the highest correlation coefficient, and again averaged. Multi-reference alignment was considered complete when no further improvement in the resolution of the averages was seen upon further cycles of refinement.

Initial models of nitrilase oligomers containing 8 and 10-monomers were derived from co-ordinates of homology models fitted to the reconstruction of the *R. rhodochrous* J1 fibre (Thuku *et al.* 2007). Density was generated from these co-ordinates and filtered to a resolution of 20 Å. This density was projected at 84 equally spaced angles and the resulting projections were used to classify and align the images in the stack by multi-reference alignment in Spider (Frank *et al.* 1996). The resulting classes were back projected to create a new reconstruction. This procedure was iterated until a stable converged volume was achieved. The reconstructed 'c' structure of BAA-870 nitrilase was docked into the 8 and 10-mer model volumes of J1 nitrilase, to determine the likely multimer number for the BAA-870 nitrilase. Volumes and reconstructions were visualized using UCSF Chimera (Pettersen *et al.* 2004).

3D reconstruction using SPIDER: Reference volume was obtained using a 4-dimer model of the negative stain density of a nitrilase helix was generated and filtered to 20 Å using Spider routines. With the angular accuracy set to 15°, the model was projected in to 83 reference views.

3.3. Results

3.3.1. Wild-type nitrilase purification

Wild-type nitrilase was induced in the native organism by addition of benzonitrile to the culture medium. The purification of nitrilase was followed by ammonia assay using benzonitrile as substrate to track nitrilase containing fractions. A cell-free extract prepared from induced bacterium was applied to an anion exchange chromatography column (Figure 3.5).

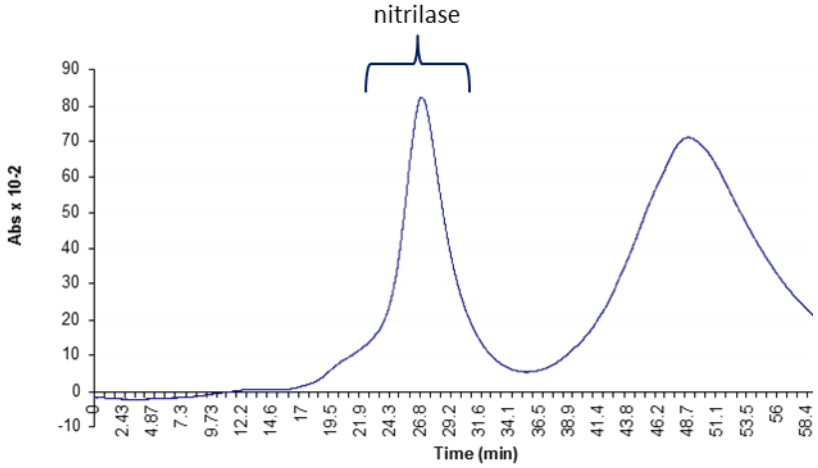


Figure 3.5: Ion exchange chromatography of cell free protein extract containing wild-type nitrilase.

The Q-Sepharose column was equilibrated with 50 mM potassium phosphate buffer containing 100 mM NaCl, pH 7.4, and protein was eluted (blue) using a 0.1 M to 1 M KCl gradient. Nitrilase eluted in fractions 20-30, corresponding to 20-30 minutes.

Fractions containing activity against benzonitrile were pooled, concentrated and applied to a S300 size exclusion column (Figure 3.6). The nitrilase containing fractions were immediately applied to EM grids for negative stain microscopy.

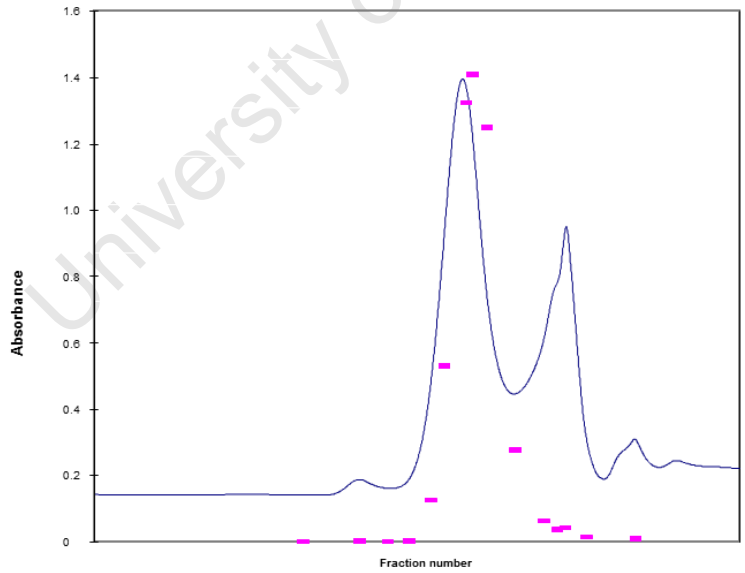


Figure 3.6: Gel filtration chromatography of pooled ion exchange fractions containing nitrilase.

A Sephacryl S-300 HR gel-filtration column (GE-Healthcare, USA) was equilibrated and run using 50 mM potassium phosphate buffer, pH 7.4 with 200 mM KCl. Activity against benzonitrile (pink) was consistent with protein elution (blue) at size ~608 KDa.

3.3.2. Microscopy of wild-type nitrilase

Benzonitrile activity in the eluted size exclusion fractions coincided with protein eluting at size ~608 KDa. The elution size (608 950 Da) corresponds to approximately 15 monomers of 41 kDa. The J1 nitrilase was shown to form fibres on assembly of subunits (Thuku *et al.* 2007), hence the behaviour of wild-type BAA-870 nitrilase was monitored by EM. Transmission electron microscopy (TEM) images showed that nitrilase from *R. rhodochrous* ATCC BAA-870 increasingly forms fibres over time, consistent with the nitrilase from *R. rhodochrous* J1 (Figures 3.7). Purified nitrilase was observed to be a mixed species of mostly small subunit complexes, with very few long assemblies. Nitrilase stored for 16 days showed assembly of subunits into fibres of variable length. When nitrilase was left at 4 °C for more than 30 days, fibre formation was mostly complete (Figure 3.8).

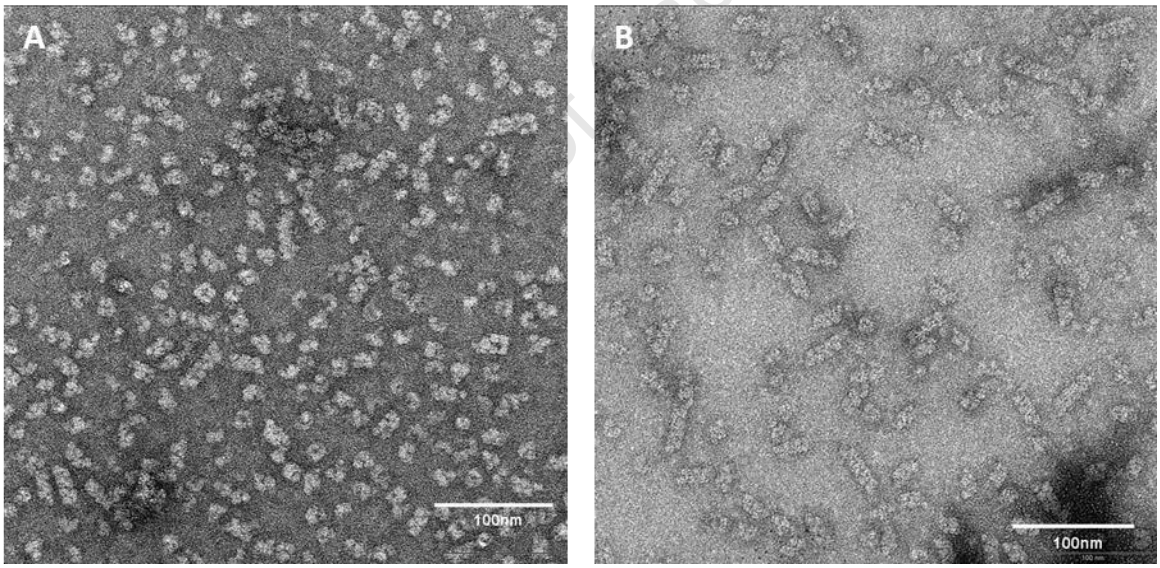


Figure 3.7: Wild-type nitrilase purified from *R. rhodochrous* ATCC BAA-870 forming fibres over time.

TEM grids of the purified gel filtration fractions containing nitrilase were prepared by negative staining with uranyl acetate. Purified nitrilase was stored at 4°C and electron micrographs taken over a period of 16 days to monitor fibre self-assembly. Left micrograph (A): recorded after one day. Right micrograph (B): recorded after 16 days.

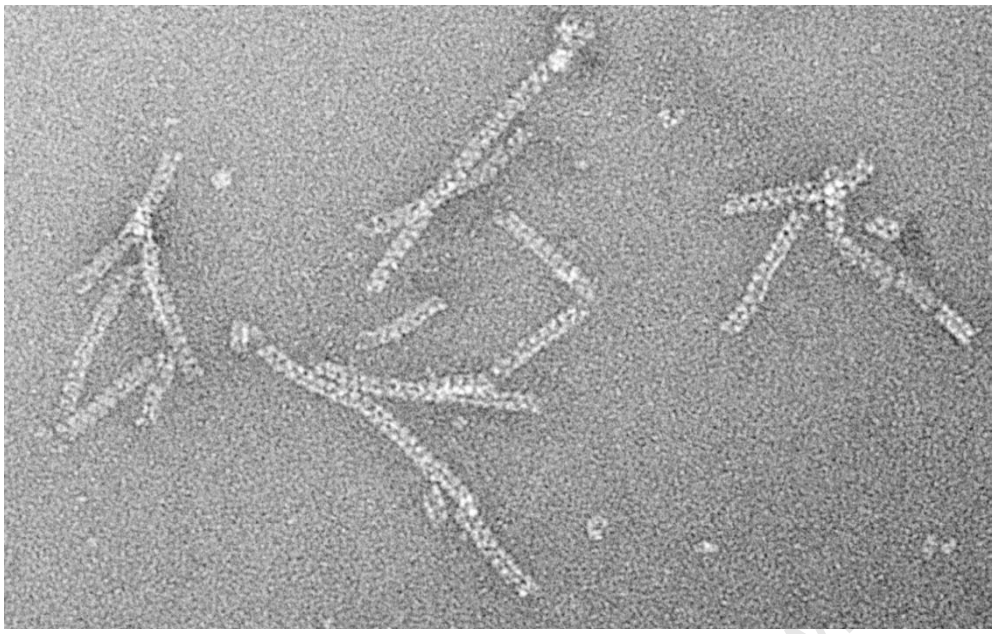


Figure 3.8: Nitrilase from *R. rhodochrous* ATCC BAA-870 forming fibres.

TEM grids of the purified nitrilase were prepared by negative staining with uranyl acetate. Purified nitrilase was stored at 4°C and electron micrographs taken after a period of 30 days to monitor fibre self-assembly.

3.3.3. Cloning and expression of recombinant nitrilase

3.3.3.1. Cloning of recombinant nitrilase

The nitrilase gene was amplified from the *R. rhodochrous* ATCC BAA-870 genome and blunt-cloned into the pJET1.2 vector. Clones containing the nitrilase gene were screened by restriction digestion (Figure 3.9).

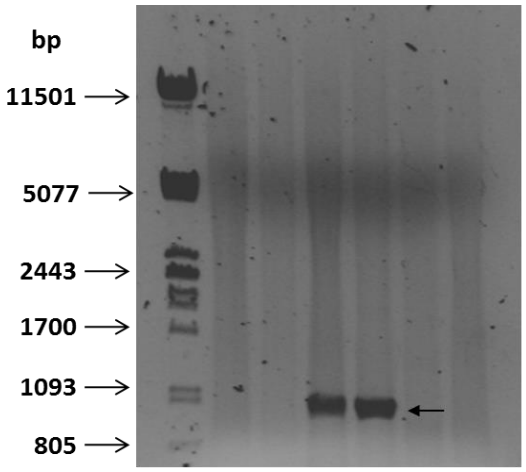


Figure 3.9: Agarose gel (0.8 %) showing amplification of the nitrilase gene from the template genome of *R. rhodochrous* ATCC BAA-870.

The nitrilase gene (arrow), relatively overloaded on the gel, is 1107 nucleotides. *Hin*dIII digested Lambda DNA was used as the marker.

Clones containing the nitrilase gene were sequenced and selected for directional expression cloning into pET28a(+). The nitrilase gene consists of 1107 nucleotides. The cloned nitrilase was BLASTed against a nucleotide database, and a 98% and 97% identity to the nitrilase gene from *R. rhodochrous* tg1-A6 and *R. rhodochrous* J1 respectively was found (Table 3.3).

Table 3.3: *R. rhodochrous* ATCC BAA-870 nitrilase BLAST results

Accession number	Strain and protein description	Nucleotide Identities	Authors
EF465367.1	*R. rhodochrous* strain tg1-A6 nitrilase gene, complete cds	1087/1098 (98%)	(Luo *et al.* 2008)
D11425.1	*R. rhodochrous* gene for nitrilase, complete cds	1070/1098 (97%)	(Kobayashi *et al.* 1992c)

The translated sequence of the nitrilase gene cloned from *R. rhodochrous* ATCC BAA-870 was aligned with the two closest *R. rhodochrous* nitrilase BLAST matches and compared (Figure 3.10).

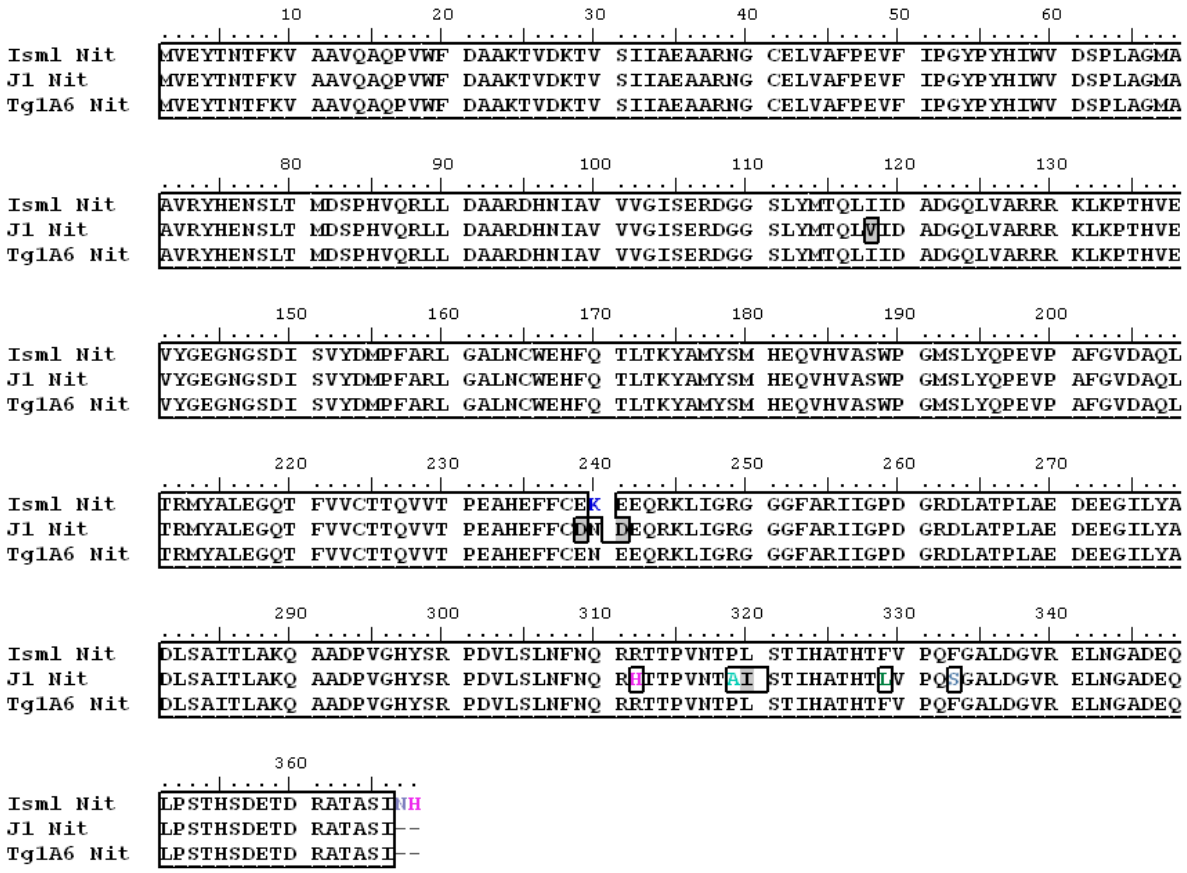


Figure 3.10: Alignment of nitrilases from *R. rhodochrous* ATCC BAA-870 (Isml), J1 and tg1-A6.

The protein identity score is 98.65 % in DNAMAN version 5.2.9 by Lynnon BioSoft©. ClustalW alignment was performed using *MEGA* version 3.1 (Kumar *et al.* 2004) using a Gonnet protein weight matrix. Redrawn in BioEdit Sequence Alignment Editor, version 7.0.5.2 (Hall 1999). Isml Nit, J1 Nit and Tg1A6 Nit refer to nitrilases from *R. rhodochrous* ATCC BAA-870, *R. rhodochrous* J1 and *R. rhodochrous* Tg1-A6, respectively. Boxed area indicates conservation of sequence. Grey and coloured highlights show conservative and non-conservative changes, respectively.

The nitrilase proteins from *R. rhodochrous* BAA-870, tg1-A6 and J1 share a 98.6% identity. Few major differences are observed, and amino acid changes are mostly conserved. For example, at position 117, J1 has a valine residue while BAA-870 and tg1-A6 have isoleucine (both are aliphatic residues). A major difference is observed between positions 238 and 240, a block of 3 consecutive amino acids: the 'DND' (Asp-Asn-Asp) of J1 nitrilase corresponds to 'ENE' (Glu-Asn-Glu) in the case of Tg1-A6, all conservative changes, whereas for BAA-870, it corresponds to the less conservative "EKE" (Glu-Lys-Glu). The positive Lys mutation at position 239 could be significant for structural and/or catalytic properties. While nitrilase

from BAA-870 and tg1-A6 are identical at the C-terminus (Figure 3.11), J1 nitrilase has a His at 311 as opposed to Arg, Ala at 318 instead of Pro, and Leu and Ser at positions 328 and 332 respectively instead of Phe. BAA-870 has an extra 2 amino acids (Asn - His) at the C-terminal.

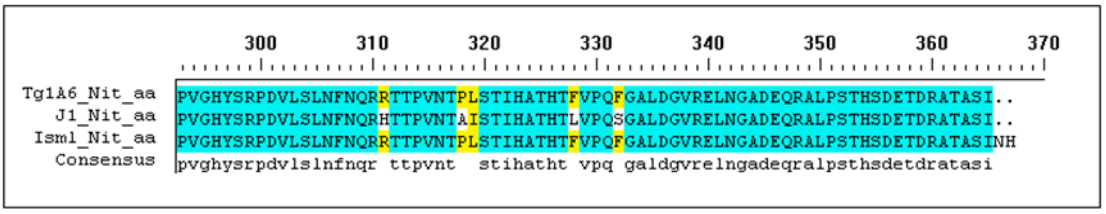


Figure 3.11: Alignment of the nitrilases from *R. rhodochrous* tg1-A6, J1 and BAA-870, highlighting differences at the C-terminus.

The alignment was performed using ClustalW (Thompson *et al.* 1994). Isml_Nit, J1_Nit and Tg1A6_Nit refer to nitrilases from *R. rhodochrous* ATCC BAA-870, *R. rhodochrous* J1 and *R. rhodochrous* Tg1-A6, respectively. Yellow and white highlights indicate conserved and non-conserved amino acid changes, respectively.

3.3.3.2. Expression of recombinant nitrilase in pET28a(+)

Expression of nitrilase was tested using L-rhamnose induction and IPTG induction. Expression of 41 kDa nitrilase was induced using L-rhamnose in single step cells (Figure 3.12).

Standard protein expression using BL21(DE3) and the inducer IPTG was also tested, since protein expression routinely is carried out using this protocol (Figure 3.13). L-Rhamnose is an expensive chemical, and a cheaper induction protocol was sought after with future high volume expression in mind for industrial applications.

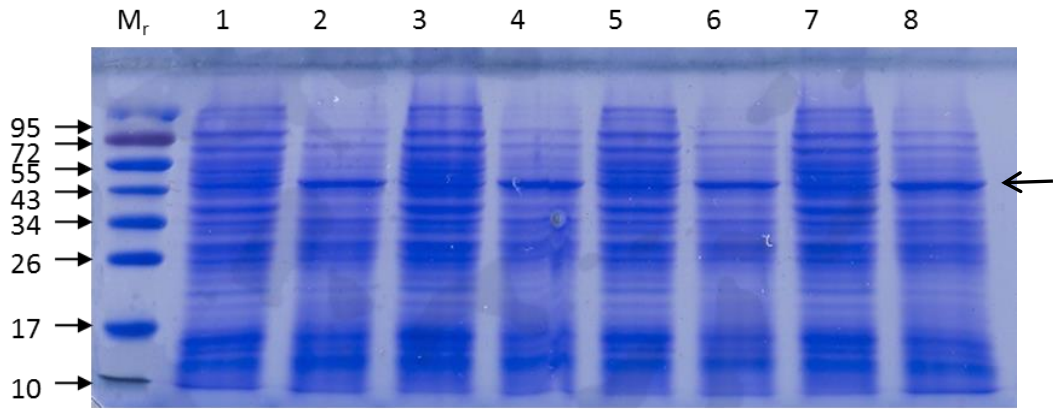


Figure 3.12: Analysis of nitrilase induction by 12% SDS-PAGE showing uninduced vs induced nitrilase from multiple clones.

Nitrilase (indicated by arrow) was expressed from pET28(+) plasmid in E. coli Single Step cells, rhamnose-induced cells for~3 hrs at 25°C. HPLC of uninduced and induced cell culture shows induced clones are active against benzonitrile. Odd lanes: uninduced, even lanes: induced nitrilase. Aliquots (1 mL) of uninduced and induced cells were standardised by addition of (OD_{600} X 100) SDS-PAGE sample buffer.

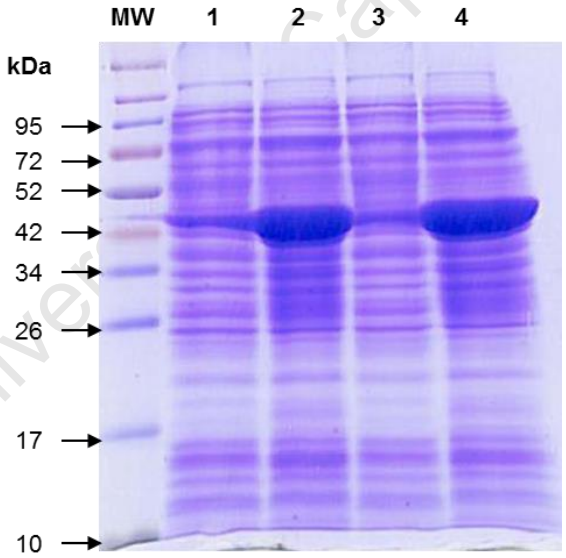


Figure 3.13: IPTG Induction of nitrilase from pET28a(+) using *E. coli* BL21(DE3).

Cells were induced at an OD (600 nm)of ~0.3 by addition of 0.1 M IPTG, and grown overnight at 23 °C. Aliquots (1 mL) of uninduced and induced cells were standardised by addition of (OD_{600} X 100) SDS-PAGE sample buffer, and analysed by 12 % SDS-PAGE.

Nitrilase from BAA-870 was successfully over-induced using a standard IPTG induction in *E. coli* BL21 cells.

3.3.4. Purification of recombinant nitrilase

Nitrilase expression in the pET28a(+) construct incorporates an N-terminal His-tag, and was purified using nickel column chromatography. A quick purification protocol was sought after for future industrial purification of nitrilase. Nitrilase could be purified using a two-step chromatographic protocol, by nickel column and size exclusion. First, a batch purification of nitrilase using the nickel affinity resin, Protino, was tested (Figure 3.14).

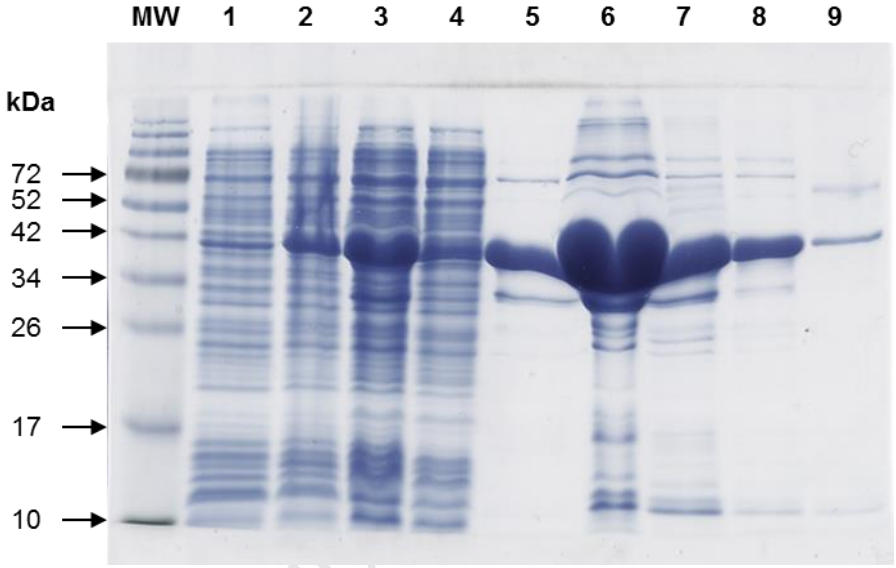


Figure 3.14: Purification of His-tagged nitrilase using nickel resin batch purification.

Lane MW, marker; Lane 1, uninduced; Lane 2, induced; Lane 3, cell free extract; Lane 4, flow through; Lane 5 – 9, Elution fractions from His-tag purification using Protino nickel resin.

Batch elutions (consecutive washes with elution buffer containing 350 mM imidazole) containing nitrilase were pooled, concentrated and eluted by size exclusion chromatography (Figure 3.15). The purification was assessed by SDS-PAGE (Figure 3.16).

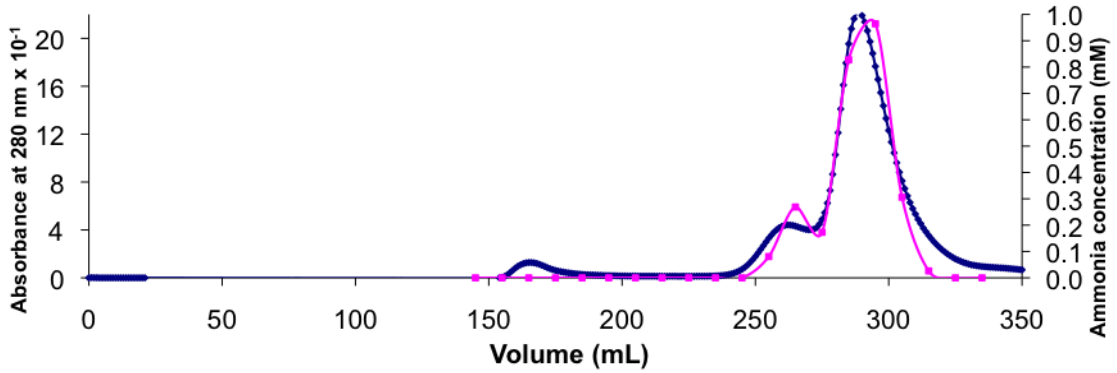


Figure 3.15: Elution of nitrilase from Sephacryl S400 size exclusion chromatography, following purification by batch His-tag purification using Protino Ni resin.

Activity against benzonitrile (pink line) was present in both peaks eluting after 250 mL, eluting at sizes between 600 – 150 kDa.

The size exclusion run was repeated using an S200 Sephacryl size exclusion column. The elution profile was the same as that from S400, with two peaks containing nitrilase activity.

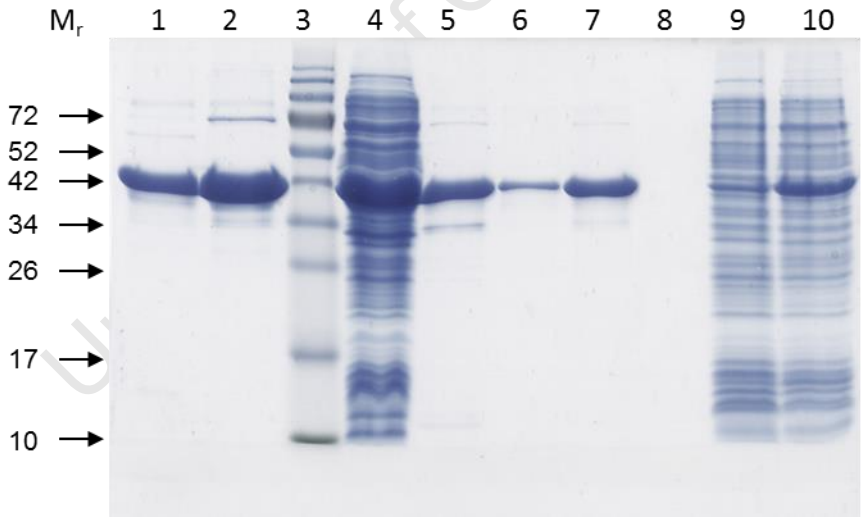


Figure 3.16: Analysis by 12% SDS-PAGE of purification of recombinant nitrilase.

Nitrilase was purified by batch nickel resin and size exclusion. Lane 1, first Ni-column elution; Lane 2, second nickel column elution; Lane 3, marker; Lane 4, cell free extract; Lane 5, pre-S400 protein sample; Lane 6 and 7, purified nitrilase eluted from S400 size exclusion fractions; Lane 9 and 10, uninduced and induced nitrilase from BL21(DE3) cells respectively.

The fractions containing nitrilase at lower apparent molecular mass (and hence the pre-fibre form) was concentrated and subjected to a further round of size exclusion chromatography

using an FPLC PWXL4000 size exclusion column (Figure 3.17). The nitrilase-containing fraction eluted by FPLC was shown to be homogeneous by SDS-PAGE (Lane 9, Figure 3.18), and further used for imaging by EM.

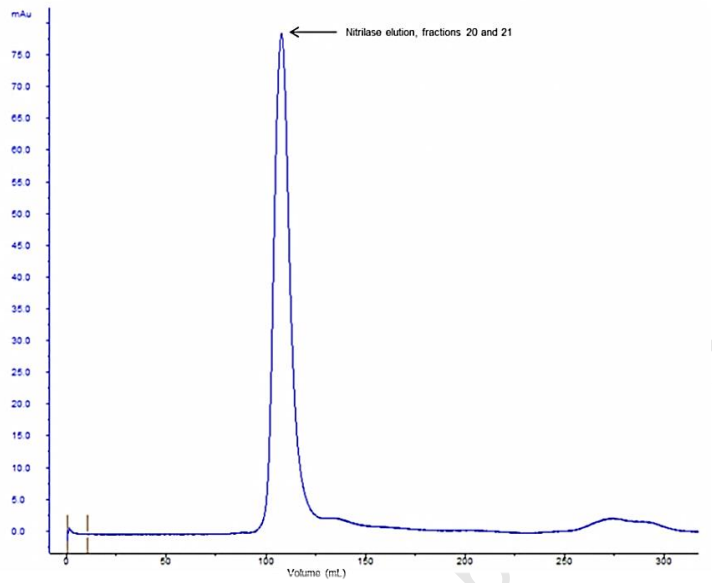


Figure 3.17: Nitrilase elution by FPLC size exclusion chromatography.

Nitrilase was induced in *E. coli* BL21, and purified by His-tag purification prior to purification by S200 size exclusion column chromatography. Nitrilase containing fraction were pooled, concentrated and subjected to a further round of size exclusion using an HPLC PXWL 4000 column.The column was equilibrated and eluted with 50 mM sodium phosphate buffer, pH 7.4, containing 100 mM NaCl. The nitrilase elution is indicated.

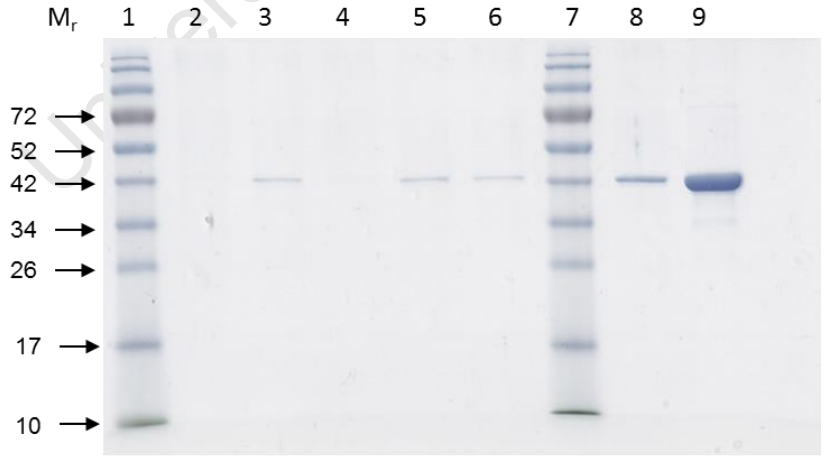


Figure 3.18: Analysis by 12% SDS-PAGE of purification of nitrilase using size exclusion chromatography.

Lanes 1 and 7, marker; Lane 8 and 9, dilutions of pure nitrilase eluted from S400 size exclusion chromatography. Lanes 2-6, dilutions of nitrilase from pure fractions of S200 size exclusion elution.

In order to deduce whether the nitrilase from BAA-870 also undergoes a C-terminal truncation via autolysis, a fraction of purified nitrilase that was stored at 4 °C for more than one week (and noted to be precipitated and collected on the bottom of a glass fraction tube) was analysed by SDS-PAGE and compared to a freshly eluted nitrilase sample (Figure 3.19).

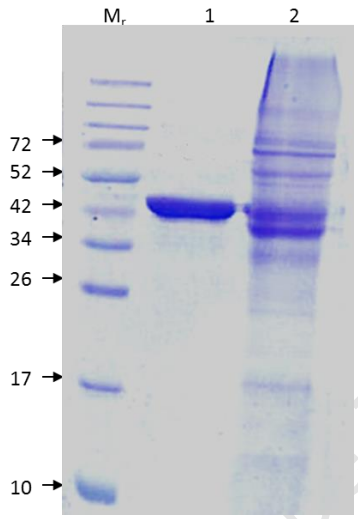


Figure 3.19: Analysis by 12% SDS-PAGE of recently purified nitrilase and precipitated nitrilase incubated for several days.

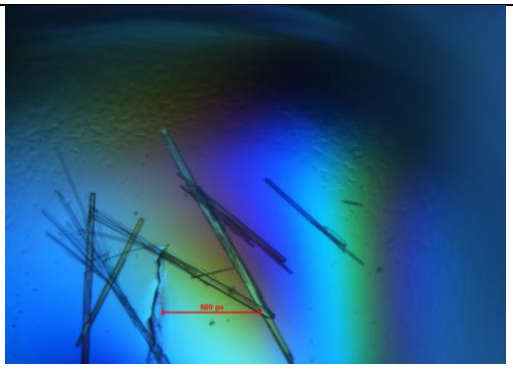
Nitrilase was purified using affinity and size exclusion chromatography. Lane 1: Freshly purified nitrilase. Lane 2, nitrilase stored for 7 days at 4 °C (resuspended precipitate).

It is likely, as observed from a shift in size of band as analysed by SDS-PAGE, that the BAA-870 nitrilase also undergoes autolysis before polymerization into fibres. Purified nitrilase is active against both benzonitrile, used as a control substrate measurement throughout purification steps, as well as against 3-cyanopyridine (data not shown).

3.3.5. Nitrilase Crystallization attempt

Several crystallisation screens yielded crystals and also precipitates, of which the most promising is shown (Table 3.4) since this condition yielded large crystals at all concentrations of nitrilase tested. Unfortunately, however, these crystals did not diffract, and hence no structural data was obtained.

Table 3.4: Most promising result of nitrilase crystallization

Condition*	Image
10% w/v PEG 6000 100 mM bicine pH 9	

*Using the JBScreen JCSG++ HTS from Jena Bioscience.

3.3.6. Electron microscopy and reconstruction of recombinant nitrilase

Microscopy of recombinant expressed nitrilase from *R. rhodochrous* ATCC BAA-870 revealed that the protein forms 'c'-shaped oligomers in the pre-fibre form (Figure 3.20).

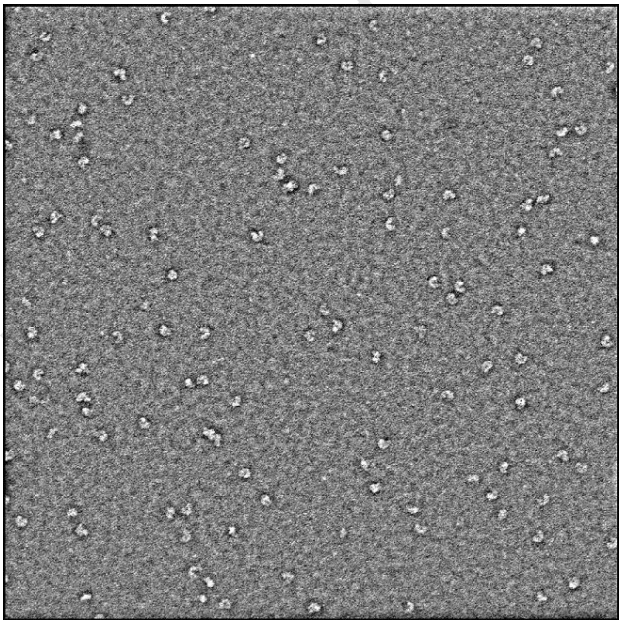


Figure 3.20: Low-dose electron micrograph of purified recombinant nitrilase from *R. rhodochrous* ATCC BAA-870.

The negatively stained ‘c’-shaped oligomers of BAA-870 were picked into 100 x 100 pixel boxes and manually ‘cleaned’ by viewing the shapes and representative class averages selected (Figure 3.21 and Figure 3.22).

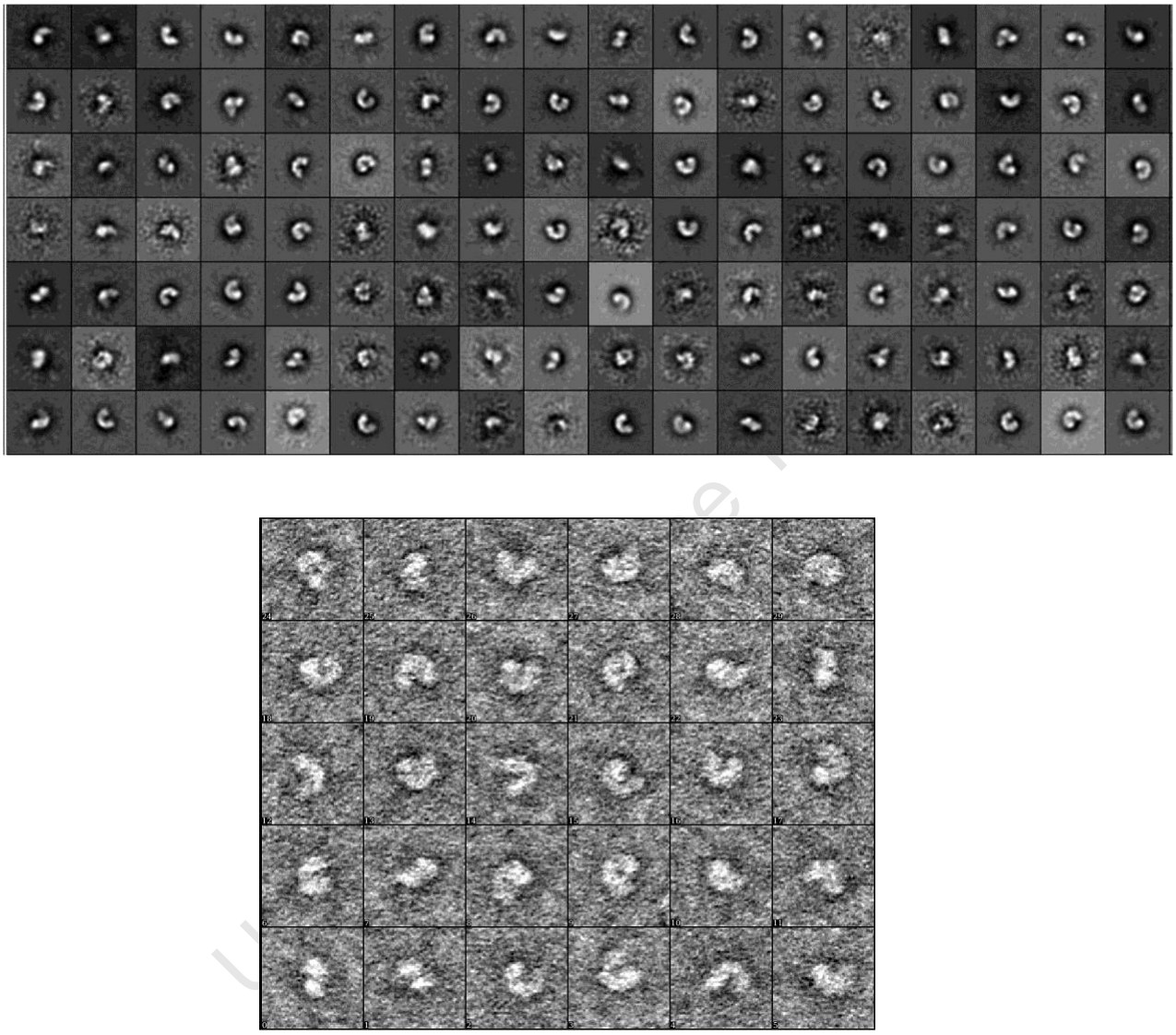


Figure 3.21: Representative class averages after K-means classification using SPIDER (top) and EMAN (below).

Single particles were boxed using sxboxer in the SPARX suite of programs (Baldwin and Penczek 2007; Hohn *et al.* 2007).

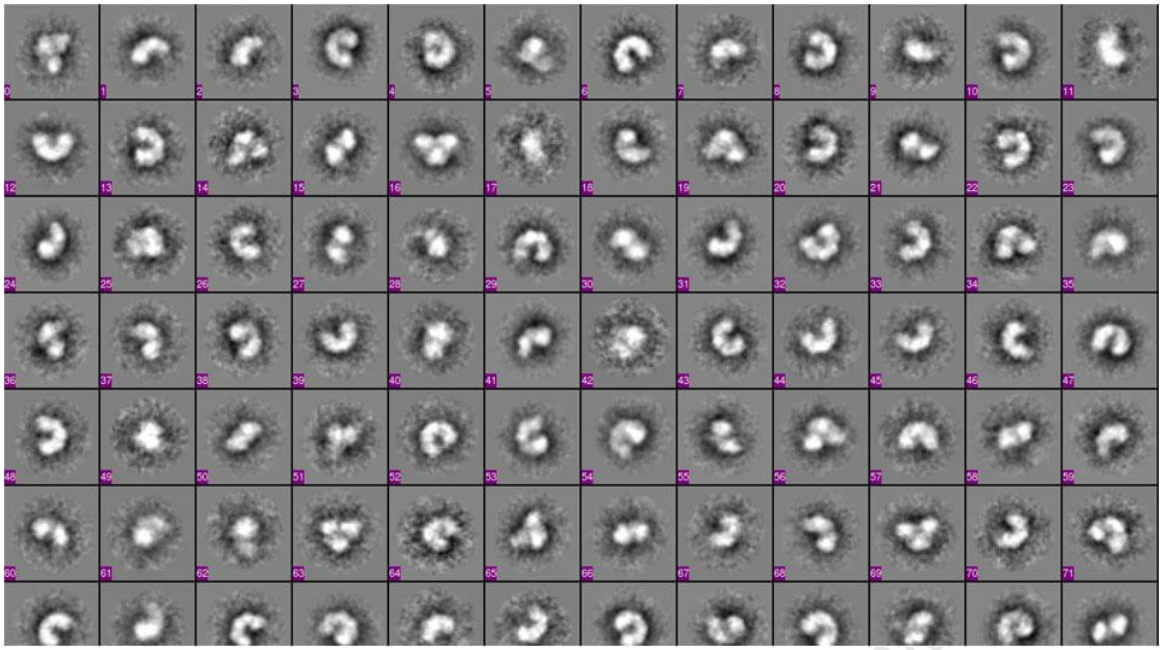


Figure 3.22: Views of various boxed (100 x 100 pixel) single particles of homogeneous purified recombinant oligomeric nitrilase from *R. rhodochrous* ATCC BAA-870 oligomers.

Single particles were boxed using sxboxer in the SPARX suite of programs (Baldwin and Penczek 2007; Hohn *et al.* 2007).

The reference volume was projected into 83 views and a reconstructed volume model made (Figure 3.23).

The reconstructed volume of the BAA-870 nitrilase was scaled and viewed using UCSF Chimera (Pettersen *et al.* 2004), and the resulting structure docked into volumes of the J1 fibre of varying length. The 8-mer docking was the closest structural volume fit. The 'c'-shaped oligomer, therefore, likely consists of 8 nitrilase subunits and is arranged in a left-handed helix, as deduced by docking of the model into an 8-mer volume from the reconstructed nitrilase structure from J1 (Figure 3.24).

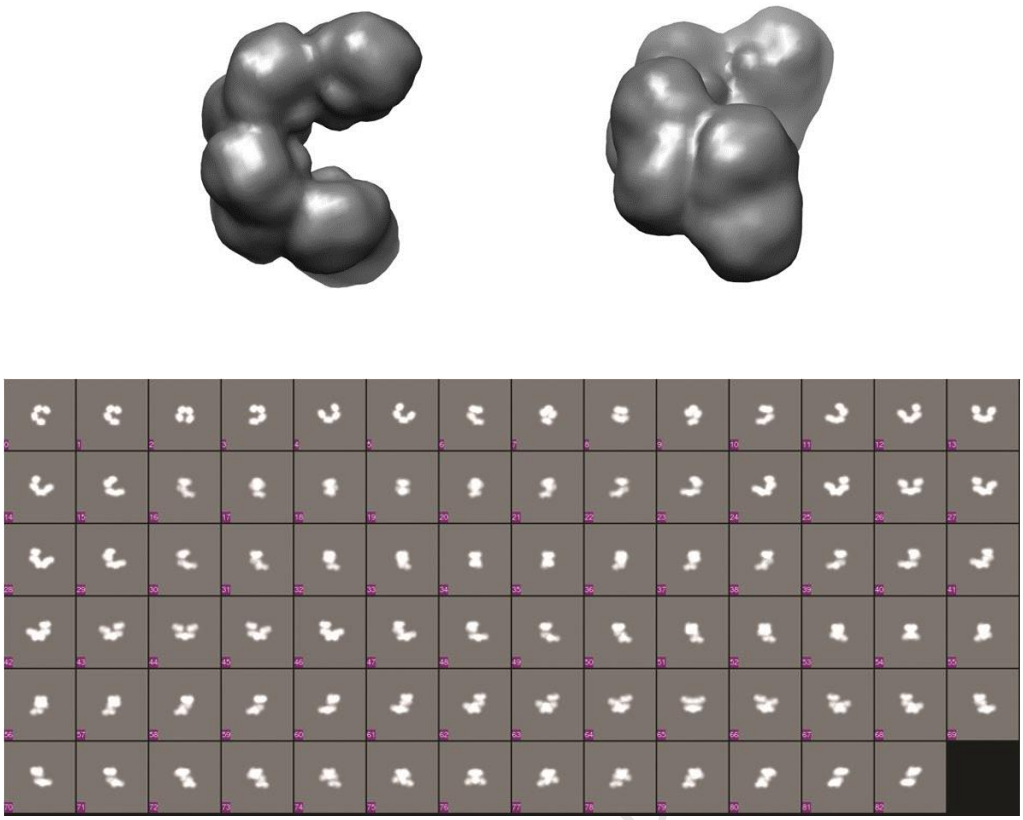


Figure 3.23: Views (top and side) of the reference volume that was projected into 83 views using Spider routines.

Top – top and side views. Bottom – 83 views from Spider. Volumes of reconstructed models of recombinant nitrilase oligomeric 'c' structures depicted using UCSF Chimera (Pettersen *et al.* 2004).

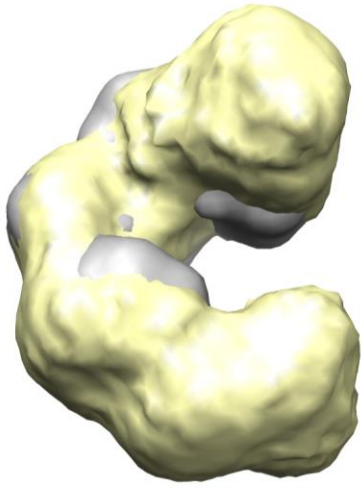


Figure 3.24: Volume of recombinant nitrilase oligomer docked into a modelled 8-mer of nitrilase from J1.

Volumes were viewed and created using UCSF Chimera (Pettersen *et al.* 2004).

3.3.6.1. 2D multireference alignment (MRA)

The image stack (5787) was aligned to the reference projections using the 'AP SH' routine in Spider. This procedure generated a parameter doc file and an aligned image stack (Figure 3.25). The images were then assigned to corresponding projections of the reference volume (Figure 3.26).

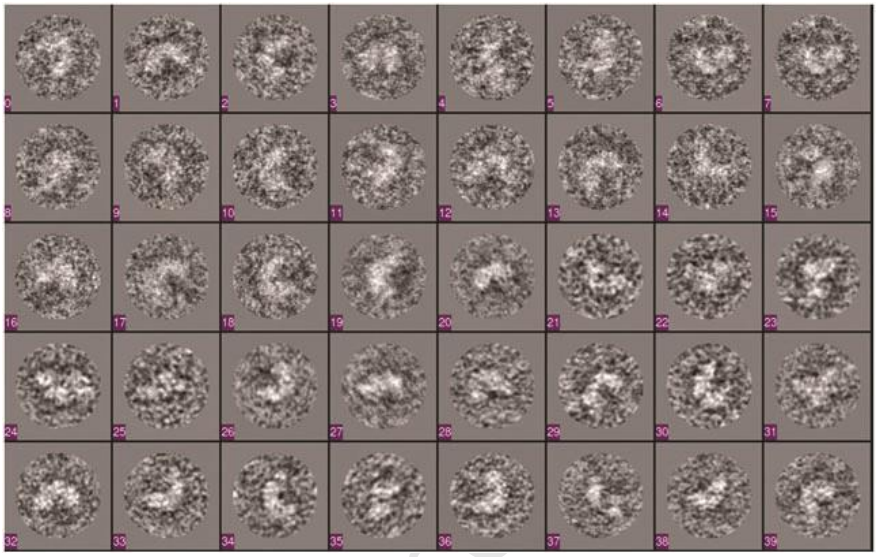


Figure 3.25: Gallery of aligned and masked images of c-shaped rhodococcal nitrilase.

For all projections, all aligned particles of a given reference view were averaged together (Figure 3.26). In cases where there were no images assigned to a single view of the reference, a blank image was created.

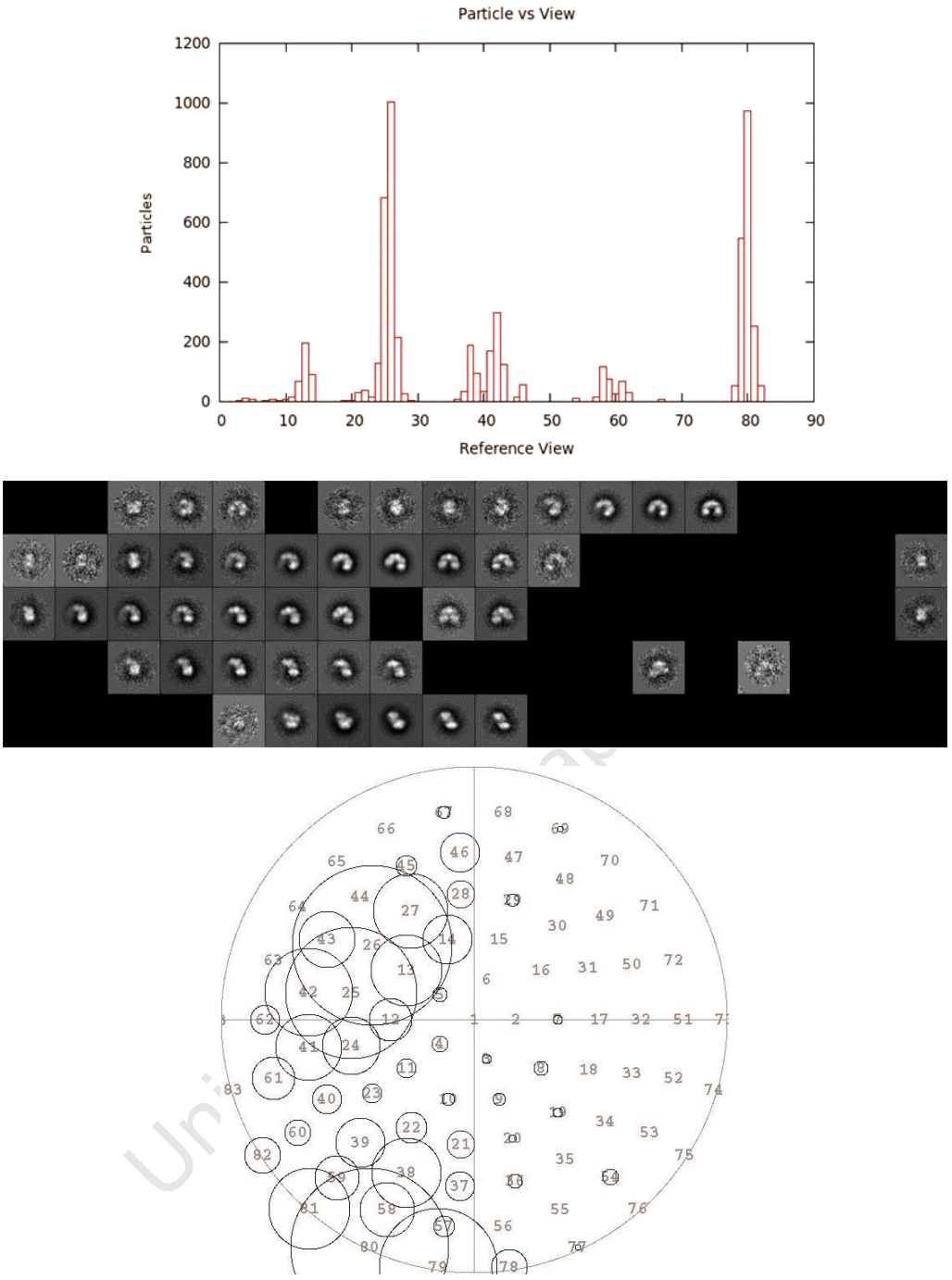


Figure 3.26: Assignment of images to corresponding projections (47) of the reference volume after MRA.

Top – reference volume. Middle - average of particles matching the assigned views. Views without images assigned to them were replaced with a blank. Bottom – Map of angular coverage. Numbers in the circle denote the Eulerian angles. The areas of the circles are proportional to the numbers of particles in assigned to that view/direction.

Further particle selection was made by first identifying an appropriate cross-correlation value and then selecting a correlation cut-off threshold to reject some particles (Figure 3.27). About 5% of the particles with low cc values were discarded before 3D reconstruction.

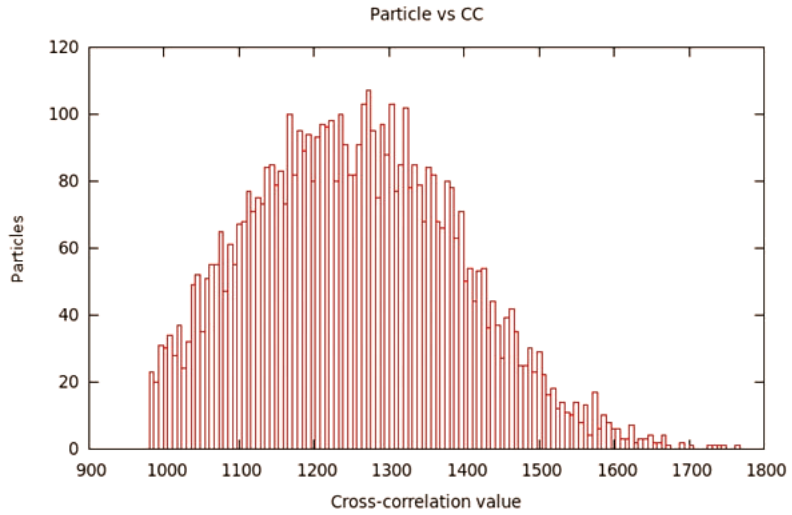


Figure 3.27: Cross-correlation of the particles with projected views of the reference volume.

A cc threshold of 1000 was applied to exclude ~5% of poor images.

3.3.6.2. 3D reconstruction and refinement

The selected aligned particles (5718) were used to create an initial 3D volume of the c-shaped rhodococcal nitrilase via projection matching in Spider (Shaikh *et al.* 2008). The initial volume was refined repeatedly by decreasing the angular steps for reference projections during each cycle (Cong and Ludtke 2010). This gives the data particles a chance to find a better approximating reference match for each iteration ('settle in') and find better fitting angles than the initial choices. A total of 7 iterations of refinement were executed (Figure 3.28). The resolution of the final reconstruction was estimated by splitting the particle images into two equal sets and comparing the two resulting volumes. The two volumes were then merged into the final reconstruction which was then filtered to the calculated resolution (Figure 3.29).

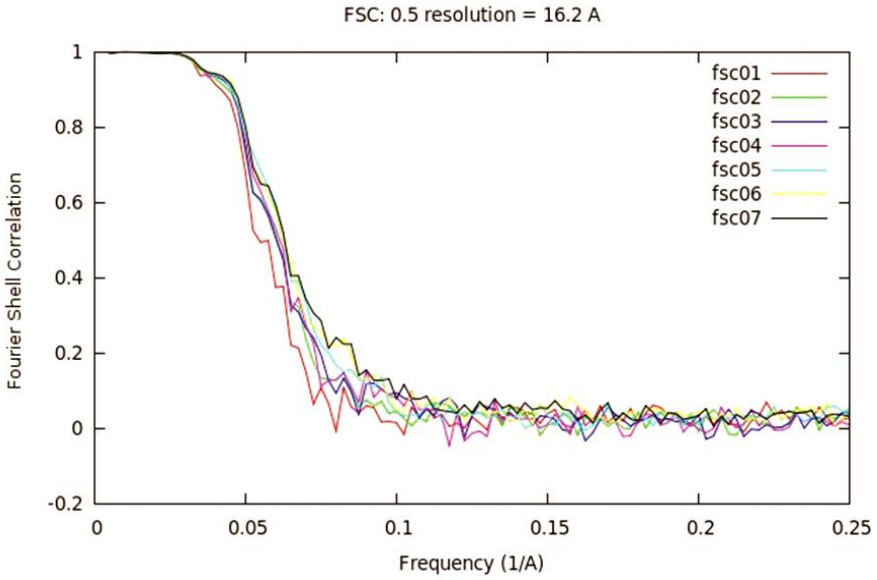


Figure 3.28: FSC curves for all seven refinement cycles.

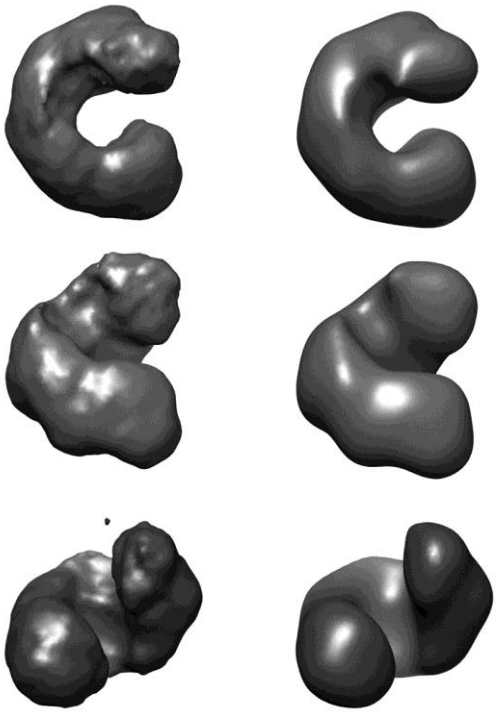


Figure 3.29: Reconstructed left-handed, c-shaped *R. rhodochrous* ATCC BAA-870 nitrilase.

Left – unfiltered volumes. Right – filtered volumes. Resolution ($FSC_{0.5}$) was determined to be ~16 Å.

3.3.6.3. 3D reconstruction using EMAN2

The masked images (5787) and reference model (Figure 3.23) were imported via the e2workflow.py interface in EMAN v2.05 for 3D reconstruction and refinement. MRA was performed by projecting the reference volume at an angular distribution of 15°, cross-correlating each image with these projection views and back projection into an initial 3D volume for subsequent refinement (Figure 3.30).

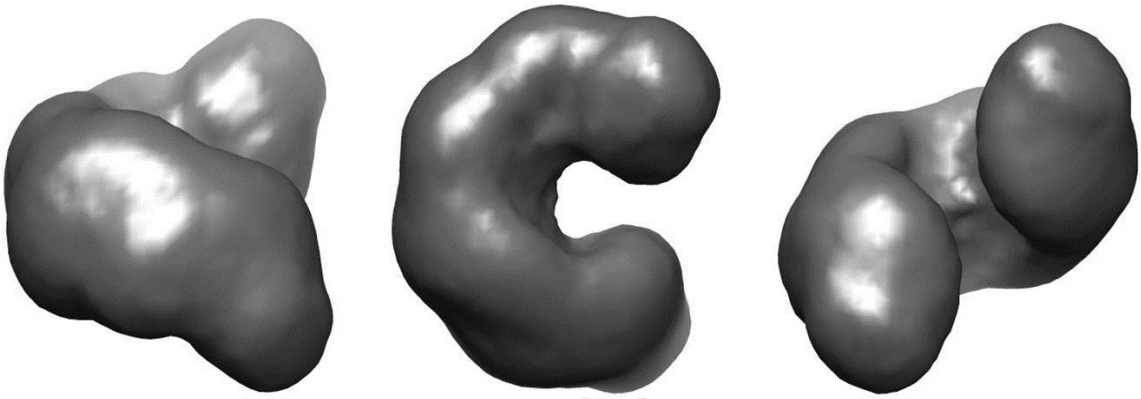
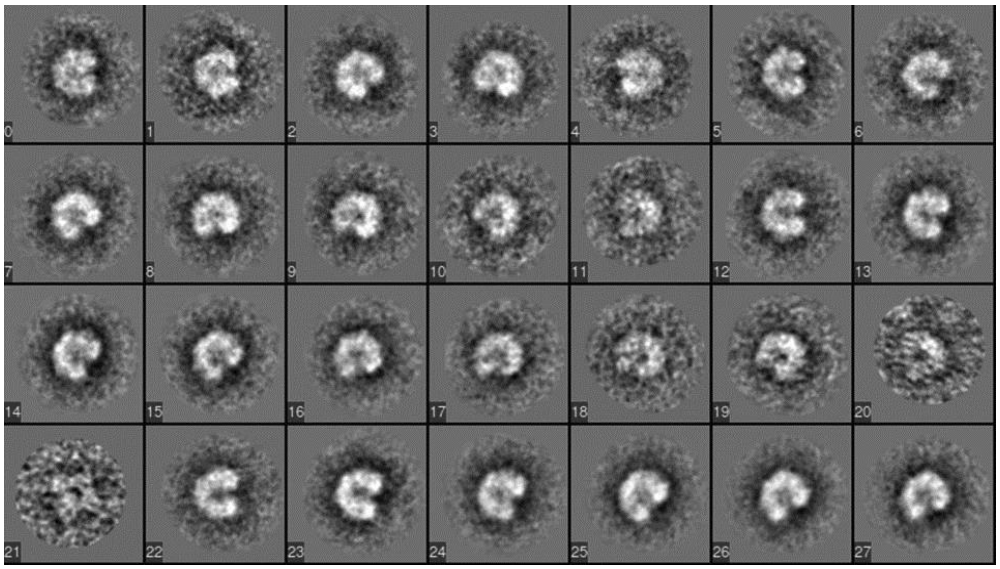


Figure 3.30: Views of the initial 3D volume which was used for further 3D refinement using e2refine.py in EMAN2.

Further 3D refinement was performed by generating more views (angular spacing = 2.5) and cross-correlating these with the image data. During each refinement cycle, only the 'good' particles (determined using a sigma cut-off of 0.8 in each class) were back projected on the basis of the cross-correlation score. This process was computationally intensive and takes ~8 days to complete on a Dell Precision T7400 workstation with 4 processors and 20 GB of memory.



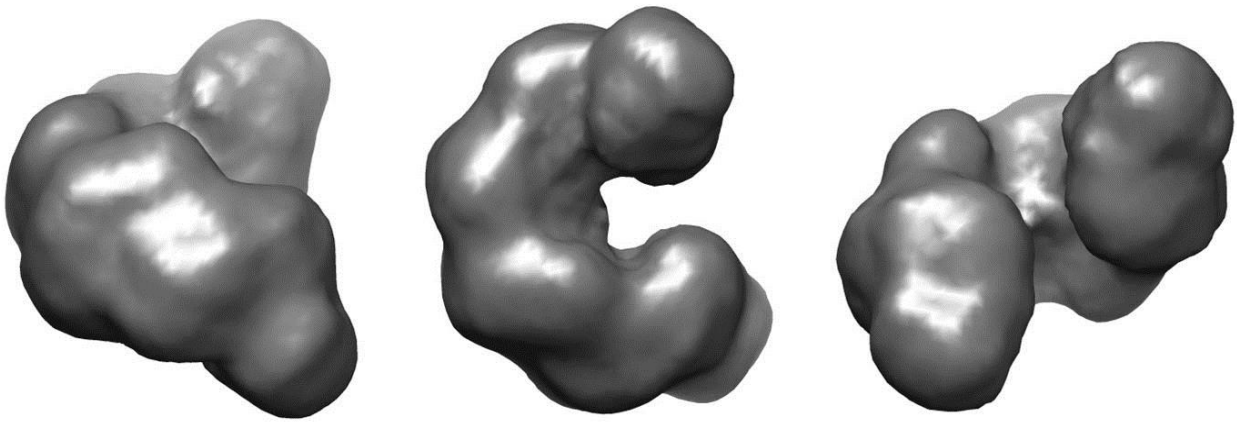


Figure 3.31: Representative gallery of final class averages after iterative cycles of refinement and views of the final 3D reconstruction.

A representative gallery of final class averages after four iterative cycles of refinement (top).Views of the final 3D reconstruction (without imposing two-fold symmetry) following refinement of the initial volume with EMAN2. There are 9-10 subunits (~5 dimers due to possible partial occupancy in the terminal dimers of the molecule) forming the 1-start left-handed c-shaped oligomer.

3.3.6.4. Generation of an initial model using 2D reference class averages and subsequent 3D refinement

Using e2initialmodel.py, an initial model was generated from class averages (a subset of good class averages (~30) was selected prior to initial model reconstruction) obtained after

reference free alignment. The symmetry for the particle was defined as dihedral (D_1). This process starts with random pattern of blobs and does a full single particle refinement using the class averages. The algorithm generated ten random models. Each model was evaluated for how well its 2D projections agree with the class averages obtained from the reference free 2D analysis and then ranked.

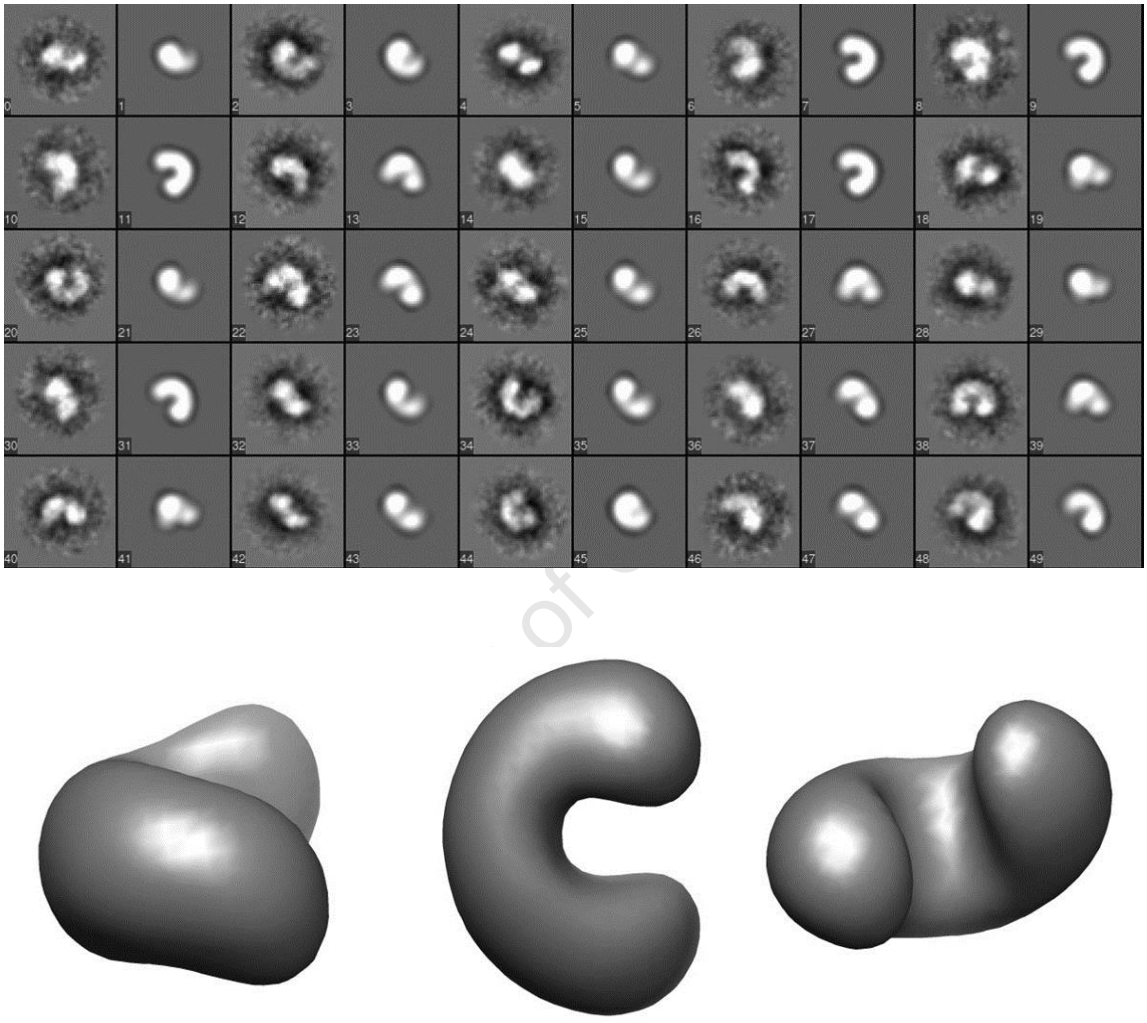


Figure 3.32: Representation of pairs class average and model projection images, and views of the best initial model generated from reference-free 2D class averages.

Top - a representation of pairs of images: class average on the left and model projection on the right. These show good agreement and suggest that the starting model (bottom) resembles the actual structure of the rhodococcal nitrilase. Bottom – side, top and staggered view of the best initial model generated by EMAN2 from reference-free 2D class averages. Out of the 10 models generated, this was considered best because it was left-handed and resembles the crescent shape seen in the 2D class averages.

The initial model was refined against the data using e2refine.py. The model was projected every 2.5° and each projection was cross-correlated with the raw data. Classification was performed during each cycle of refinement and poor class images were excluded (on basis of a sigma cutoff = 0.8) prior to back projection.

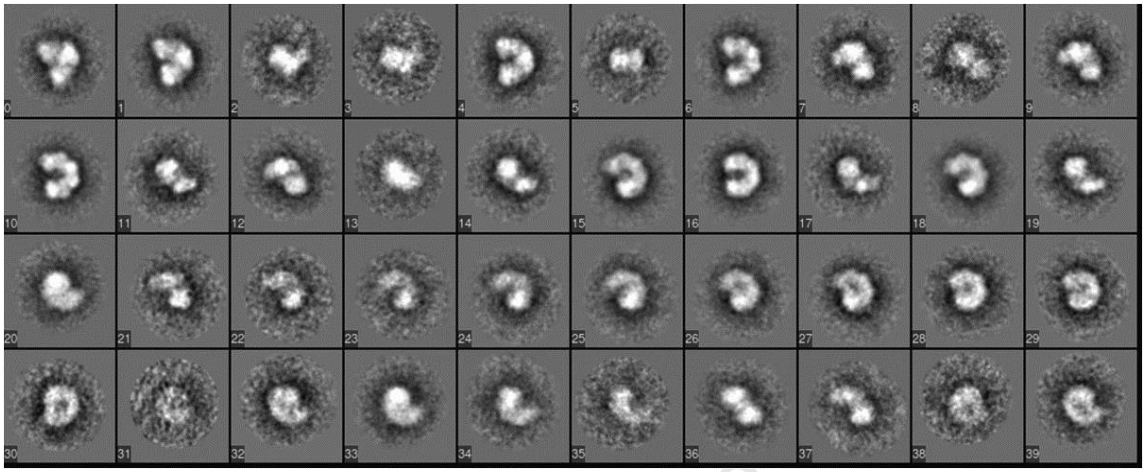


Figure 3.33: Representative set of final class averages after 3D refinement using EMAN2.

The particle outer diameter was measured to be ~11.4 nm. There are ~5 dimers suggesting that the c-shaped nitrilase could form either an octamer or a decamer. The closure of the spiral oligomer does not occur probably because there is not enough space between the terminal subunits to accommodate addition of another subunit. Furthermore, the presence of an extended carboxy-terminus would prevent closure due to steric hindrance in the centre of the molecule.

3D refinement converged to the same structure obtained using the Spider model (Figure 3.23) as the initial reference. This suggests that the choice of initial model (Figure 3.32) was correct.

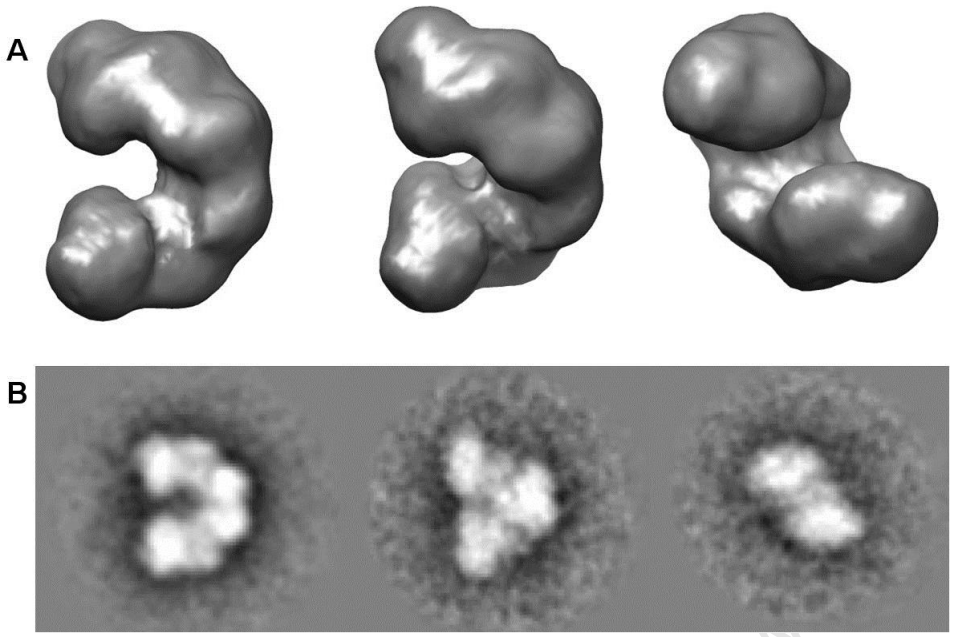


Figure 3.34: The final structure of *Rhodococcus rhodochrous* ATCC BAA-870 nitrilase. Representative views of the negative stain density of the c-shaped oligomer (A) correspond to 2D class averages (B). The density of each dimer and a two-fold axes located at the 'A' surface (perpendicular to the spiral axis) can be identified. There are ~5 dimers in the c-shaped left-handed spiral.

3.3.6.5. Resolution of 3D reconstruction

The quality of the reconstruction was assessed by splitting the raw data set into two halves and running full refinement for each half data set with starting models whose resolution was randomized to 25 Å. This procedure was executed in the EMAN2 software package using e2refine_evenodd.py algorithm which performs a 'gold standard' resolution test. The resolution of the final negative stain density was determined at $FSC_{0.5}$ to be 16 Å (Figure 3.35).

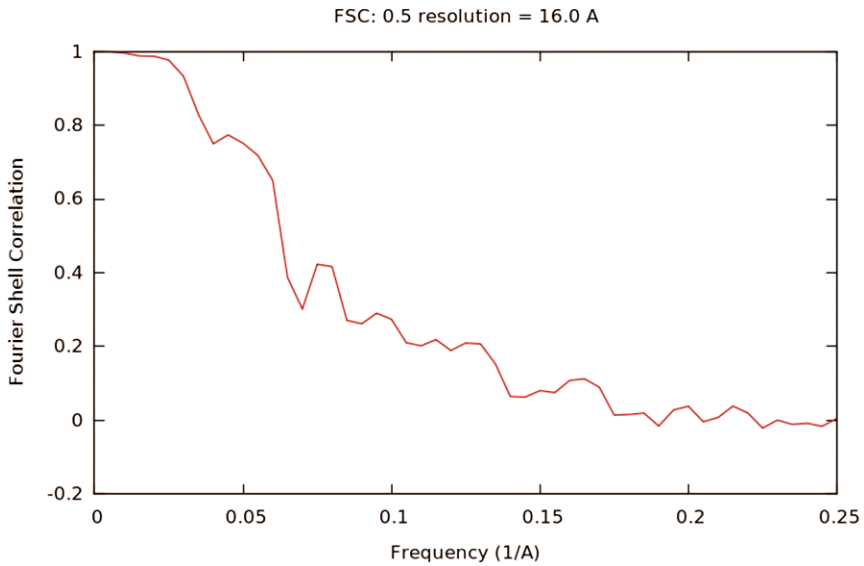


Figure 3.35: The resolution (FSC = 0.5) of the 3D refined map of *Rhodococcus rhodochrous* ATCC BAA-870 was determined to be 16 Å.

3.4. Discussion and conclusions

In summary, the wildtype BAA-870 nitrilase was expressed by nitrile induction in minimal media and the formation of fibres from pure nitrilase was observed over time. This behavior was similar to that reported in J1 nitrilase (Thuku *et al.* 2007). The nitrilase was then amplified from the genome and cloned for heterologous expression. Both the wildtype and cloned nitrilase were purified. The gene sequence confirmed a high percentage similarity to J1 nitrilase at the sequence level. Electron microscopy of the pre-fibre form of BAA-870 nitrilase showed a c-shaped particle, which was further clarified by reconstruction. The BAA-870 nitrilase was found to be similar, but not identical to, the J1 nitrilase. Nitrilase expression in the pET28a(+) construct incorporates an N-terminal His-tag, however, and like the J1 nitrilase, it is likely that BAA-870 nitrilase also undergoes a C-terminal truncation via autolysis (which should be confirmed via mass spectrometry). The oligomerization into c-shaped particles, and further oligomerization into fibres, does not appear to be impacted by the presence of the His-tag, although a slight impact on the 3D structure cannot be ruled out completely.

The first nitrilase superfamily crystal structures were N-carbamyl-D-amino acid hydrolase from *Agrobacterium* (PDB code 1ERZ) (Nakai *et al.* 2000), and the *C. elegans* NitFhit protein (PDB code 1EMS) (Pace *et al.* 2000). Both were found to exist as tetramers. Subsequently members of the superfamily which formed dimers and hexamers having 3_2 symmetry were crystallized. The most interesting crystal structure for the present work is that of beta alanine synthetase which forms a helical ramp comprising eight monomers (Lundgren *et al.* 2008). Crystallization attempts of the pre-fibre form of nitrilase from BAA-870 using a variety of buffers failed to produce viable crystals for x-ray diffraction studies. Promising needle-like crystalline objects were obtained from 10% w/v PEG 6000, 100 mM bicine, pH 9. Although no diffraction was obtained from these, they represent a starting point for further optimization.

EM of negatively stained molecules offers several advantages however, when no crystal structure is available. Negatively stained images offer increased signal-to-noise ratio allowing the molecule's finer detail to be visualised and random orientations of the molecule can be captured and computationally combined. The purification of nitrilase to homogeneity was crucial so that only identical molecules are combined to form the 3D reconstruction.

The nitrilase from *R. rhodochrous* ATCC BAA-870 was verified to have 98 % identity with that from *R. rhodochrous* J1 by mass spectroscopic analysis of wild-type purified nitrilase and by sequence analysis of the cloned nitrilase. Nitrilase was expressed in the native organism as a 600 kDa complex which matured over time to form long fibres. The nitrilase gene was cloned in an N-terminal His-tagged expression plasmid pET28a(+). The recombinantly expressed protein could be purified in a pre-fibre homogeneous form, shown by negative stain EM to be 'c'-shaped particles. Images of these particles were recorded by negative stain electron microscopy and they were reconstructed in three dimensions which verified that they were octamers. When left for an extended time, the octamers assemble into long fibres, as observed with the nitrilase from *Rhodococcus rhodochrous* J1.

The *G. pallidus* RAPc8 nitrilase is similar to some bacterial and several plant nitrilases, with a greater sequence similarity to plant, than to other bacterial nitrilases. It has been suggested that the phylogenetic positions of nitrilases are not reflected as expected in the host

organism's taxonomy due to ecological associations which resulted in horizontal gene transfers amongst eukaryotic and prokaryotic kingdoms (Pace and Brenner 2001; Podar *et al.* 2005). Negative stain EM of *G. pallidus* RAPc8 nitrilase showed heterogeneous isoforms of nitrilase: protein complexes that were "crescent-formed", "c-shaped", "circular" and "figures-of-8" with models suggesting the various complexes were composed of 6, 8, 10 and 20 subunits, respectively. Class averages showed 3 distinct open and closed ring forms, suggesting there was a progressive elongation of the oligomers along a helical ramp. These assembly stages culminated in the closed, circular structure, with two such circles associating at staggered ends to form the "figure-of-8".

The structural oligomerisation of BAA-870 nitrilase appears to display only the 'c'-shaped complex, and the larger fibre helices. The 'c'-shaped octomer appears to closely resemble a nitrilase homologue, β-alanine synthase from *Drosophila melanogaster*, which has also been shown to form eight-monomer helices (Lundgren *et al.* 2008). When the structural coordinates of β-alanine synthase from *D. melanogaster* deposited in the PDB (Lundgren *et al.* 2008) are used to generate projections using SPIDER (Frank *et al.* 1996), the overall shape of the "c" is remarkably similar to the nitrilase from *R. rhodochrous* ATCC BAA-870 and the nitrilase from *R. rhodochrous* J1. β-Alanine synthase from *D. melanogaster* is a homooctomeric assembly in the shape of a left-handed helical turn in which tightly packed dimers are related by a two-fold symmetry. β-Alanine synthase shares the nitrilase αββα-fold, but it has an extended N-terminus due a number of additional secondary structural elements. It also contains the conserved Cys, Glu, Lys catalytic triad residues. The octomeric spiral is left-handed.

Nitrilase from *R. rhodochrous* J1 also forms "c-shaped" oligomers of 480 kDa. Autolysis of a 39-amino acid C-terminal fragment causes the oligomers to rearrange into helices (Thuku *et al.* 2007). The "c-shaped" helical ramps of *R. rhodochrous* J1 formed by the truncated monomers, have an axial rise per dimer (Δz) of 15.8 Å and a rotation per dimeric subunit (Δϕ) of -73.5° (Thuku *et al.* 2007). Nitrilase from *R. rhodochrous* ATCC BAA-870 is proposed to be structurally highly similar to the structure of J1, based on the sequence identity and observed fibre formation over time. The outer diameter of the *R. rhodochrous* ATCC BAA-870 fibre was found to be 120 Å on average. The *R. rhodochrous* J1 nitrilase forms helical

fibres with molecular masses of over 1.5 MDa (Thuku *et al.* 2007). The BAA-870 fibres have diameters of 95 – 130 Å.

The modelled β-alanine synthase from *D. melanogaster* oligomer has a comparable diameter of ~120 Å and a smaller pitch of 65 – 70 Å (Lundgren *et al.* 2008). The 39 C-terminal residues that are truncated in *R. rhodochrous* nitrilases, causing fibre formation, are not present in the *D. melanogaster* β-alanine synthase sequence which has its last native residue in the subunit terminating at the concave, inner surface if the homooctamer.

For the *Rhodococcus rhodochrous* J1 nitrilase to form long helices, it is necessary for the residues beyond 327 in the C-terminus to be truncated (Thuku *et al.* 2007). If the C-terminal truncation does not happen, the J1 nitrilase also forms 'c'-shaped oligomers similar to those of the *D. melanogaster* β-alanine synthase (Lundgren *et al.* 2008). It is likely that a truncation of the C-terminus also happens in the nitrilase of BAA-870, as suggested by SDS PAGE observation after purified nitrilase is left for some time, although this should be confirmed by mass spectrometry to rule out the actions of contaminating proteases. A sequence alignment of spiral forming nitrilases showed that the C-terminal residues are not conserved (Thuku *et al.* 2009). The C-terminal region may therefore have two roles in spiral formation; to facilitate spiral formation by playing a positioning residues and strengthening interactions at the interfaces, or by interaction with structurally adjacent residues located at or near the active site, influencing the activity and stability of the spiral. Removal of residues from the C-terminus could remove effects of additional steric hindrance which may otherwise prevent the 'c'-shaped oligomers from forming complete turns to form fibres.

Purified recombinant nitrilase is active against the aromatic nitriles benzonitrile and 3-cyanopyridine (nicotinonitrile). Nitrilase-mediated bioconversion of 3-cyanopyridine is an attractive method for production of nicotinic acid (Vitamin B_3). Recently, a recombinant 35.8 kDa nitrilase from *Alcaligenes* sp. ECU0401 was found to enantioselectively hydrolyze racemic mandelonitrile to (R)-(-)-mandelic acid, with applications for future production of (R)-(-)-mandelic acid by this highly thermostable enzyme (Zhang *et al.* 2010b). Further characterization of the substrate specificity can now be carried out, and the enantioselective potential of BAA-870 nitrilase tested.

Over the last two decades, the *Rhodococcus rhodochrous* J1 nitrilase has been the most reported and well-known of the industrially relevant nitrilases (Kobayashi *et al.* 1989; Kobayashi *et al.* 1992c; Komeda *et al.* 1996c; Nagasawa *et al.* 2000; Thuku *et al.* 2007), but despite this success, it is surprising that there have been no reports of industrial applications of pure isolated J1 nitrilase. Only whole cell applications of nitrilase activity in *R. rhodochrous* J1 resting cells have been used industrially, for example the production of nicotinic acid and pyrazinoic acid (Kobayashi *et al.* 1990a; Mathew *et al.* 1988). Given the aromatic substrate preference of *Rhodococcus rhodochrous* BAA-870 nitrilase, the similarities of *Rhodococcus rhodochrous* J1 and *Rhodococcus rhodochrous* BAA-870 nitrilases, and the availability of an expressed and easily purified *Rhodococcus rhodochrous* BAA-870 nitrilase, perhaps the success of a large-scale industrial process such as those employing J1 nitrile hydratase can be repeated (large-scale nicotinamide production by Lonza, China, and (R)-(-)-mandelic acid production by Mitsubishi Rayon in Japan and BASF in Germany.

Spiral and fibre forming features in a protein can make the structural characterization relatively difficult and therefore perhaps the rhodococcal nitrilase structural characterization has lacked. For example, the 4.9-fold screw axis of the *Rhodococcus rhodochrous* J1 nitrilase fibre precludes crystallization in 3D. It is still possible however, that the c-shaped particle could be crystallized, and perhaps an atomic resolution crystal structure obtained. This study presents a full 3D reconstruction of the *Rhodococcus rhodochrous* BAA-870 nitrilase c-shaped oligomer, whereas in the case of the *Rhodococcus rhodochrous* J1 nitrilase structure only 2-D averages of two different projections have been reported (Thuku *et al.* 2007). There is still active and ongoing research supporting the case for finding new nitrilases for industrial processes and green chemistry alternatives (Gong *et al.* 2012; Kaplan *et al.* 2011a; Martínková and Kren 2010; Thuku *et al.* 2009). In future, a better structural understanding of the mechanisms governing nitrilase fibre formation could be used to control industrial reactions advantageously; for example, by engineering the enzyme behavior for controlled oligomerization of the fibre for industrial reactions, or by inducing immobilization of stable fibres onto substrates. In future, protein engineering of the nitrilase active site may also be used to control the ratio of amide/acid ratio products produced from nitriles.

Chapter 4

4. Nitrile hydratase

4.1. Nitrile hydratase

4.1.1. Introduction

Nitrile hydratases, E.C. number 4.2.1.84, are metalloenzymes found in actinomycetes. The catalytic centre contains noncorrin cobalt (Kobayashi *et al.* 1991; Payne *et al.* 1997; Yamaki *et al.* 1997) or a nonheme iron ion (Ikehata *et al.* 1989; Nishiyama *et al.* 1991), referred to as Co-type and Fe-type, respectively. In Co-type NHase there is one Co ion in each α-subunit. The postulated role for Co is in water activation, CN-triple-bond hydration and protein folding. Co-type NHase enzymes have a maximum absorption at 410 nm with no corrin absorbance in visible spectrum. There is a high sequence similarity between Co and Fe NHases, and both contain the characteristic 'CXXCSC' amino acid sequence motif in each alpha subunit. The side chains of 3 Cys residues and main chain amide nitrogens of both serine and the third cysteine residue in the motif are ligands in the iron centre of *Rhodococcus* sp. R312 (studied by X-ray crystallography) (Huang *et al.* 1997).

4.1.1.1. Low versus high molecular mass nitrile hydratase

Two forms of cobalt-containing nitrile hydratase from *R. rhodochrous* J1 have been described (Kobayashi *et al.* 1991), a low molecular mass (L-NHase) and high molecular mass (H-NHase) nitrile hydratase. L-NHase is a ~130 kDa protein containing 2 alpha and 2 beta subunits, with one Co ion per alpha/beta pair of subunit. L-NHase can be selectively induced using cyclohexanecarboxamide in the presence of Co ions. H-NHase is a 520 kDa multimer containing 9/10 of each subunit, and can be induced using urea in presence of Co. The nitrile hydratase cloned from *R. rhodochrous* ATCC BAA-870 shares identity with the low molecular mass nitrile hydratase of *R. rhodochrous* J1. L-NHase has been used in the industrial production of the vitamin nicotinamide from 3-cyanopyridine.

H-NHase and L-NHase are both composed of alpha and beta subunits, but the alpha and beta subunits of L- and H-NHase differ. H- and L-NHase can be selectively induced using urea and crotonamide, respectively (Kobayashi *et al.* 1991). Full alignment of H-NHase and L-NHase from *R. rhodochrous* J1 shows a 42.89% identity between the full length proteins (Figure 4.1).

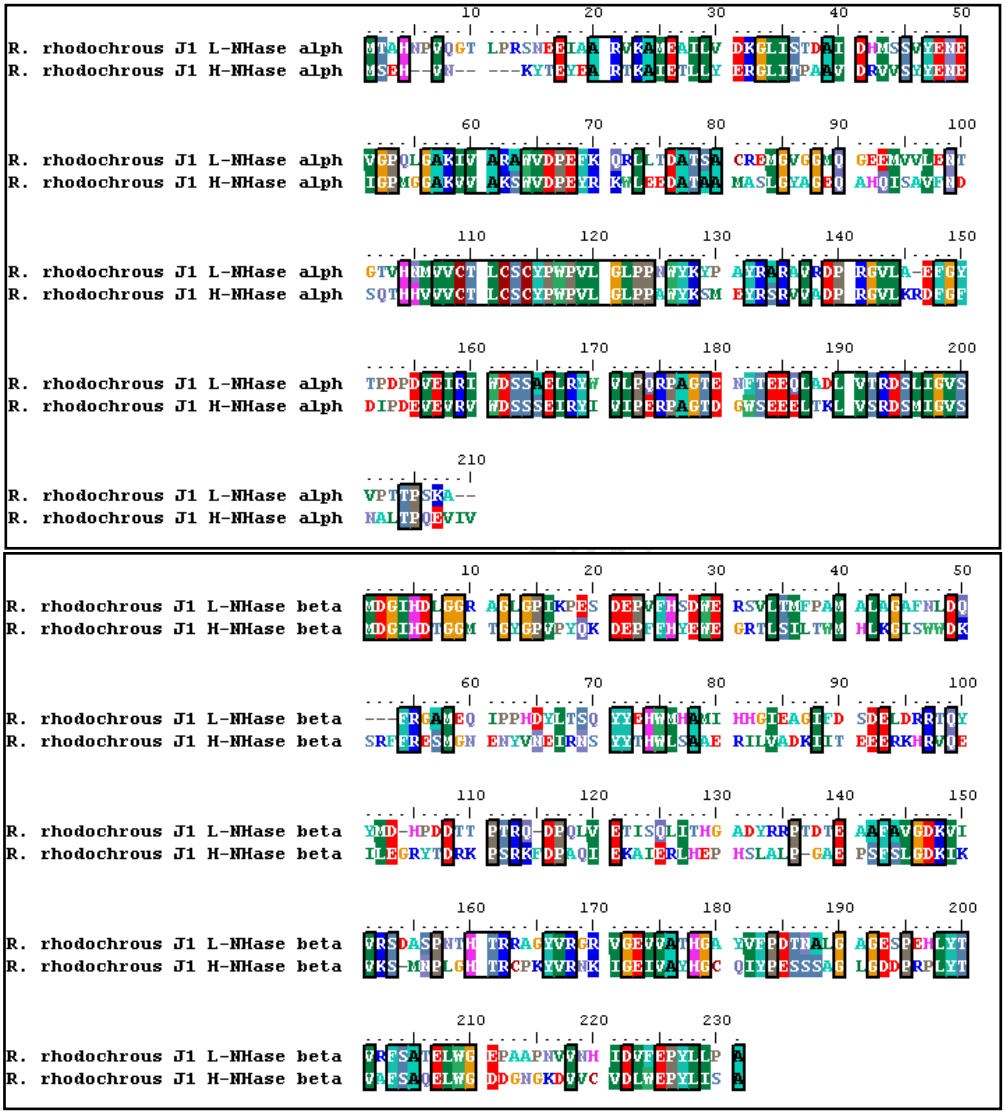


Figure 4.1: Alignment of the alpha and beta subunits of L- and H-NHase from *R. rhodochrous* J1.

Compared to the L-NHase gene, H-NHase has an extra 40 nucleotides on the N-terminal end, and an extra 146 nucleotides on the C-terminal end. The active site amino acid residues "CTLCSC" in the alpha subunit are conserved in both high and low molecular mass forms. The

cobalt ion involved in enzyme activity is contained in the alpha subunit. Alignments of the alpha subunits of L- and H-NHase show both conservative and non-conservative amino acid changes, as well as sequence insertions (Figure 4.2). It is possible that horizontal gene transfer, and gene duplication, has taken place during the evolution of the organism.

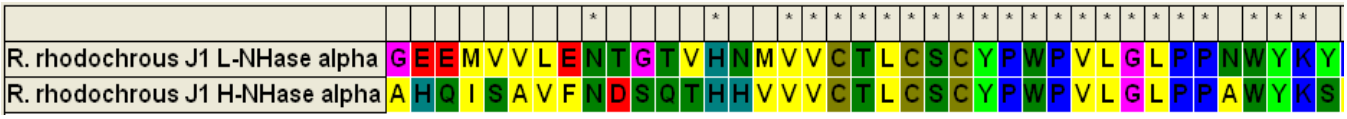


Figure 4.2: Partial ClustalW alignment of the alpha subunits of low and high molecular mass nitrile hydratase from *R. rhodochrous* J1 showing the conserved active site residues "CTLCSC".

The alignment was generated using ClustalW (Thompson *et al.* 1994) in the DNAMAN sequence analysis programme (Lynnon Corporation).

4.1.1.2. Cobalt versus Iron containing nitrile hydratase

The properties of metal-containing NHases have been reviewed in detail (Kobayashi and Shimizu 1998) and the catalytic mechanism is likely similar for all. An iron-containing NHase isolated from *Rhodococcus* sp. N-774 has been characterised by biochemical and genetic methods, and shown to be an $(\alpha\beta)_2$ dimer. The α and β subunits, as calculated from their amino acid sequences, are roughly 22 870 and 23 850 Da, respectively.

The alpha and beta subunits of nitrile hydratase have no amino acid similarity to each other, but each subunit has highly similar amino acid sequences to other alpha and beta subunits. Fe- and Co-type NHases have similar structures, but bind only their specific metals (Nojiri *et al.* 2000). It is unknown why this is so and there is only one report of experimental metal substitution in NHase (Miyanaga *et al.* 2004). Fe-type NHase can be cultured without addition of iron to the medium or coexpression of the Fe-NHase activator, and cobalt will be incorporated into the catalytic site, but cobalt substituted Fe-type NHase shows very little activity (Nojiri *et al.* 2000). There are no reports on Fe substitution in a cobalt enzyme.

Most Co-NHases show a high affinity for aromatic nitriles whereas Fe-type NHases mostly hydrate aliphatic nitriles, with some exceptions. The substrate binding pocket of Co-type NHase seems to be bigger, and may account for the apparent substrate difference (Figure 4.3). The ion coordination cluster region in the alpha subunit consists of 3 cysteine residues and one serine that are particularly highly conserved, and two arginine residues in the beta subunits. Co-type NHase does not exhibit the unique photoreactivity that Fe-type NHase exhibits, which binds nitric oxide (NO) molecules (Bonnet *et al.* 1997; Endo *et al.* 1999b; Odaka *et al.* 1997). In Fe-type NHase post-translational modification of 2 cysteine residues to cysteine-sulfinic acid and cysteine-sulfenic acid, forming a claw-like setting, is essential for association with NO and subsequent photoactivation (Endo *et al.* 2001; Murakami *et al.* 2000; Nagashima *et al.* 1998b; Piersma *et al.* 2000).

Co-type NHase has also been seen to have the claw-like setting with two modified cysteines coordinating to the cobalt ion (Co-type NHase at 1.8 Å resolution: (Miyanaga *et al.* 2001)). The proposed mechanism of action is based on the formation of an imidate as a reaction intermediate before being converted to an amide product (Kobayashi and Shimizu 1998). It was found that 2 Arg residues (βArg52 and βArg157, conserved in the beta subunit of NHase) formed 4 H-bonds with the modified oxygen atoms, thought to stabilize the claw setting. In Fe-NHase, mutation of the codons for these 2 Arg to those encoding other amino acids reduced activity and stability markedly (Endo *et al.* 2001; Piersma *et al.* 2000).

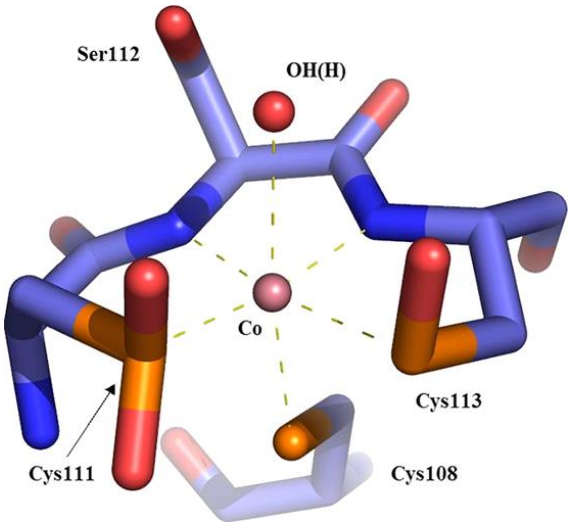


Figure 4.3: The active site of cobalt-containing nitrile hydratase from P. thermophila (Protein Data Bank code IRE).

The active site contains a six-coordinate trivalent metal ion, with three cysteine sulfurs, two amide nitrogens, and one water molecule. Cys111 and Cys113 is post-translationally modified to cysteine sulfinic acid and cysteine-sulfenic acid respectively. The image was generated from PDB code 1IRE (Miyanaga *et al.* 2001), by Mitra and Holz (Mitra and Holz 2007) and is reprinted as it appears in Mitra, S. and Holz, R. C. (2007).

4.1.1.3. Fe- versus Co-type nitrile hydratase active site conserved sequence

The conserved active site sequence of NHase is "C(**T/S**)LCSC(**Y/T**)". Fe-type NHase displays the active site sequence '**CS**LCSC**T'** while Co-type NHase is '**CT**LCSC**Y**'. The serine/threonine side chain differences are proposed to play an important role (Payne *et al.* 1997), however no mutational studies on these residues have confirmed this. In Co-type NHase, T and Y correspond to αThr109 and αTyr114, while in Fe-type NHase, S and T correspond to αSer109 and αThr114. In Co-type NHase, αThr109 has a hydrophobic interaction with the side chain of αVal36 (the distance between the Cγ2 atom of Thr and Cγ1 atom of Val is 3.7 Å), but in Fe-type NHase, the side-chain of the corresponding Ser does not interact with Val. αTyr114 is also located near the Cys cluster region. In Co-type NHase, the OH of αTyr114 forms an H-bond with main-chain O atoms of αLeu119 and αLeu121 *via* a water molecule. The αTyr114 residue also undergoes hydrophobic interactions with its environment. In Fe-type NHase, the

corresponding Thr, αThr115, forms an H-bond with the main-chain O atom of Ser (αSer113) in the cysteine cluster. The conformation of the cysteine cluster in Fe-type NHase differs compared to Co-type NHase, since the H-bond seems to pull αSer113 closer to αTyr115, making the Cys region conformation slightly more open.

4.1.1.4. Fe- versus Co-type nitrile hydratase activators

Fe and Co-type NHase activator proteins (also referred to as 'beta subunit homologs' and NHase 'chaperones') are not homologous, and their functions are believed to be different (Lu *et al.* 2003; Nojiri *et al.* 1999; Wu *et al.* 1997). Fe-type NHase activators show significant similarity to ATP-dependent ion transporters. Co-type activators have a slight similarity to the β subunit of NHase, and would likely assist in formation of active site centre during protein folding/maturation steps. The chaperone of Co-NHase may incorporate the metal ion at the active site to prevent formation of a disulphide bond (which forms readily under oxidative conditions) and facilitate the correct oxidative modification of the active site Cys residues.

4.1.1.5. Nitrile hydratase crystal structures

Excluding two eukaryotic isopenicillin *N*-synthase structures (Ge *et al.* 2008), there are 39 bacterial NHase X-ray crystal structures deposited in the PDB. Various NHase crystal structures have been obtained, including the apoenzyme of cobalt-containing nitrile hydratase (Miyanaga *et al.* 2004) shown in Figure 4.4, a thermophilic *Bacillus smithii* nitrile hydratase (Hourai *et al.* 2003), and nitrile hydratase complexed with nitric oxide (Nagashima *et al.* 1998a) or cyclohexyl isocyanide (Nojiri *et al.* To be published) (Table 4.1). All protein ligands to the iron in ferric NHase are provided by the α subunit, and there is one iron atom per αβ unit. There is no crystal structure of a wild-type NHase with a nitrile or amide substrate at the active site, although Hashimoto and Nojiri have successfully crystallised NHases with the isonitriles *tert*-butyl isocyanide (Hashimoto *et al.* 2008) and cyclohexyl isocyanide (Nojiri *et al.* To be published) respectively. Known NHase ligands are nitric oxide which binds reversibly to Fe-type NHase in a photosensitive manner (Huang *et al.* 1997; Nagashima *et al.* 1998b), and the inhibitors 1,4-dioxane, *n*-butyric acid (Miyanaga *et al.* 2001; Miyanaga *et al.* 2004) and cyclohexyl isocyanide (Hashimoto *et al.* 2008). Interestingly,

despite the apparent abundance of rhodococcal NHase enzymes in general, there is no Co-type NHase structure reported that has been isolated from a *Rhodococcus*. This would suggest that groups researching NHase have not attempted to crystallize the enzyme (unlikely), or more likely, that the enzyme may be a difficult crystallization target.

The α subunit of NHase has the primary amino acid sequence motif Cys-X-X-Cys-X-Cys. The observed amino acid motif and iron stoichiometry is comparable to the motif that provides two of the four cysteine ligands in the rubredoxin iron site, suggesting that the active site resides in the α subunit. It has been shown, through separation of the two subunits, that iron is present in the α subunit but not the β subunit (Okada *et al.* 1996).

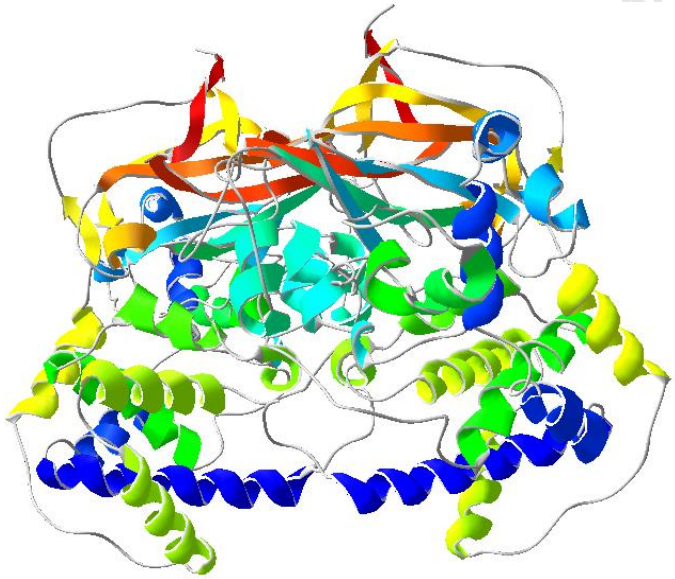


Figure 4.4: Ribbon representation of cobalt-containing Nitrile Hydratase from *Nocardia sp*.

The enzyme is shown as a homotetramer made up of the αβ asymmetric unit. The image was modelled on the protein data bank entry 1IRE by Miyanaga *et al.* 2001 using SwissPDB Viewer version 3.7 by Berman *et al.* (Berman *et al.* 2000).

Table 4.1: Summary of reported Co-type nitrile hydratase crystal structures

NHase type	Organism source	ligand	Resolution (PDB ID)	Reference
Mutant Co(III)-type NHase with enhanced thermostability	*Geobacillus pallidus*	none	1.16 Å (3HHT)	Van Wyk *et al.* 2009 Not published
Native Co-NHase	*Bacillus sp*. RAPc8	none	2.52 Å (2DPP)	(Tsekoa *et al.* 2004)
Native Co-NHase	*Bacillus smithii*	none	2.6 Å (1V29)	(Hourai *et al.* 2003)
Native Co(II)-NHase	*Pseudonocardia thermophila*	none	1.8 Å (1IRE)	(Miyanaga *et al.* 2001)
Co(II)-NHase	*Pseudonocardia thermophila* JCM 3095	complexed with inhibitor, n-butyric acid	1.63 Å (1UGP)	(Miyanaga *et al.* 2004)
Co(II)-NHase	*Pseudonocardia thermophila* JCM 3095	apoenzyme	2.0 Å (1UGQ)	(Miyanaga *et al.* 2004)
Co(II)-NHase with mutation α-T109S	*Pseudonocardia thermophila* JCM 3095	none	1.8 Å (1UGR)	(Miyanaga *et al.* 2004)
Co(II)-NHase with mutation α-Y114T	*Pseudonocardia thermophila* JCM 3095	none	2.0 Å (1UGS)	(Miyanaga *et al.* 2004)
Co(III)-NHase with beta subunit mutation β-H71L	*Pseudomonas putida*	none	2 Å (3QYH)	(Brodkin *et al.* 2011)
Co(III)-NHase	*Pseudomonas putida*	glycerol	2.1 Å (3QXE)	(Brodkin *et al.* 2011)
Co(III)-NHase with beta subunit mutation β-E56Q	*Pseudomonas putida*	glycerol	2.3 Å (3QYG)	(Brodkin *et al.* 2011)
Co(III)-NHase with alpha subunit mutation β-E168Q	*Pseudomonas putida*	glycerol	2.5 Å (3QZ5)	(Brodkin *et al.* 2011)
Co(III)-NHase with alpha subunit mutation β-Y215F	*Pseudomonas putida*	glycerol	2.4 Å (3QZ9)	(Brodkin *et al.* 2011)
Thiocyanate hydrolase	*Thiobacillus thioparus*	sulfate	2 Å (2DD5)	(Arakawa *et al.* 2007)
Thiocyanate hydrolase	*Thiobacillus thioparus*	fructose and D(-)-tartaric acid	2.06 Å (2DD4)	(Arakawa *et al.* 2007)
Thiocyanate hydrolase with modified Co centre	*Thiobacillus thioparus*	phosphate	2.25 Å (2DXB)	(Arakawa *et al.* 2009)
Thiocyanate hydrolase	*Thiobacillus thioparus*	L(+)-tartaric acid	1.9 Å (2DXC)	(Arakawa *et al.* 2009)
Thiocyanate hydrolase, oxidised	*Thiobacillus thioparus*	L(+)-tartaric acid, fructose and β-D-glucose	1.78 Å (2ZZD)	(Arakawa *et al.* 2009)

4.1.1.6. NHase gene organization

Although they are involved in hydrolysis of the nitrile bond, the nitrile hydratases are not members of the nitrilase superfamily, and hence have no structural or mechanistic similarity to the nitrilases. All characterised bacterial nitrile hydratases are composed of two kinds of subunits assembled into heterooligomers. Mostly they form dimers or tetramers, with a few larger molecules containing up to 20 subunits (Komeda *et al.* 1996b; Zhou *et al.* 2010). Nitrile hydratases typically form $\alpha\beta$ or $\alpha_2\beta_2$ functional proteins with the catalytic centre at the α-β subunit interface. Nitrile hydratases contain a metal ion at the active site, and bacterial nitrile hydratases usually contain Fe^{3+} or Co^{3+} (Yamada and Kobayashi 1996), whilst the fungus *Myrrothecium verrucaria* has been shown to contain Zn^{2+} (Maier-Greiner *et al.* 1991).

The microbial strain that has been the most exploited for its activity by industry, *R. rhodochrous* J1, produces both high molecular mass (H-NHase) (Komeda *et al.* 1996b) and lower molecular mass (L-NHase) nitrile hydratase (Komeda *et al.* 1996a; Wieser *et al.* 1998). The enzymes are each composed of two subunits, α and β, with the α subunit differing in size from the β, and both subunits of H-NHase differing from those of L-NHase. H-NHase is overproduced in this strain (as more than 50% of the soluble protein fraction) when urea is used as an inducer in the culturing medium.

It is usual for the expression of nitrile hydratase and amidase enzymes to be associated in the same organism. Indeed, the structural genes are adjacent on the same operon and are under the control of the same activator protein. The nitrile hydratase (NHase) operon consists of six genes, *nhr2*, *nhr1*, *ami*, *nha1*, *nha2* and *nha3*, that encode NHase regulator 2, NHase regulator 1, amidase, NHase α subunit, NHase β subunit and NHase activator, respectively (Lu *et al.* 2003). Expression of nitrilase, however, is separate from NHase and amidase expression in the same organism, and can be selectively modulated without affecting the expression of nitrile hydratase and amidase genes (and *vice versa*). The amount of each enzyme produced by an organism can be increased by addition of an inducer such as propionitrile to the fermentation medium (Gradley *et al.* 1994). Interestingly, it has been repeatedly reported that nitrile hydratase is generally induced by amides (the reaction product) rather than the nitrile (substrate) in microorganisms that metabolise nitriles

(Kobayashi *et al.* 1992a; Komeda *et al.* 1996a). This may mean that a low level constitutive expression of NHase may occur, allowing for formation of amide from nitrile, thereby causing induction.

4.1.1.7. Nitrile Hydratase Reaction Mechanism: insights into substrate binding and the role of the metal ion at the active site

Resonance Raman spectroscopy (Brennan *et al.* 1996; Nelson *et al.* 1991), EPR (electron paramagnetic resonance) (Sugiura *et al.* 1987), ENDOR (electron nuclear double resonance) (Doan *et al.* 1996; Jin *et al.* 1993) and EXAFS (extended X-ray absorption fine structure) (Nelson *et al.* 1991; Scarrow *et al.* 1996) studies have been used to propose a structural model for the ferric NHase active site shown in Figure 4.5. It was proposed to have a biologically relatively rare octahedral coordination geometry consisting of two cysteine thiol ligands in *cis* configuration, three histidine nitrogen ligands in *mer* configuration, and one solvent-exchangeable hydroxide ligand (Okada *et al.* 1996).

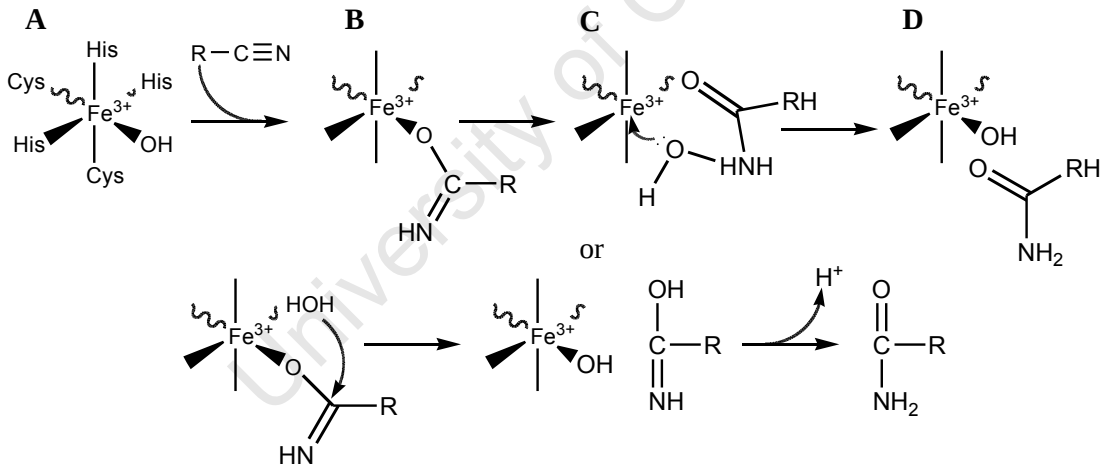


Figure 4.5: Iron coordination at the active site proposed for nitrile hydratase and a possible reaction mechanism involving a coordinated hydroxyl group.

The mechanism for nitrile hydration was proposed on the basis of EXAFS and EPR studies and two X-ray crystal structures, and redrawn from Shearer *et al.* (2002) and Nagashima *et al.* (1998) (Nagashima *et al.* 1998b; Shearer *et al.* 2002).

The first proposed reaction mechanism for ferric NHase shown in Figure 4.5 involves attack of the metal-bound hydroxide on the nitrile group, which gives rise to an unstable amide

tautomer coordinated to the metal active site (Figure 4.5B). Rearrangement of the amide and subsequent dissociation of the product implies there is a transient formation of a five-coordinate iron site upon transfer of the bound hydroxide (Figure 4.5C). Re-coordination of a solvent-derived hydroxide completes the catalytic cycle and accounts for the products observed from NHases. A possible alternate mechanism in which the bound intermediate is activated for attack by a solvent molecule is also shown. This mechanism, however, is still not widely accepted by many nitrile hydratase researchers.

4.1.1.8. Photoreactivity of Fe-NHase

Purified Fe-type NHase may be obtained as a mixture of 'low pH' and 'high pH' forms. Low pH forms are catalytically active while the high pH form needs to be activated. This activation can be accomplished through photolysis (Honda *et al.* 1994). Resonance Raman spectroscopy studies have shown that low- and high-pH forms have extensive differences in hydrogen bonding (Brennan *et al.* 1996). NHase, therefore, possesses unique intrinsic activity traits in the presence of light in that it can undergo photo-reactivation when irradiated with visible light after aerobic incubation in the dark. Endo and colleagues (Endo *et al.* 1999a) have proposed a biochemical mechanism for this photoreactivity. Light irradiation of an iron complex (chromophore) in the β-subunit of the enzyme induces a conformational change of the subunit. Inactive NHase has a non-heme iron (III) centre to which an endogenous nitric oxide (NO) molecule is bound. Upon photo-reactivation, the NO is released and the activity of the enzyme recovered. It has been suggested through the use of Fourier-transform infrared difference spectroscopy (Endo *et al.* 1999a) that irradiation causes a conformational change in the β subunit, and a subsequent breakage of the Fe-N bond. This results in photo-dissociation of the endogenous NO molecule from the NHase catalytic centre. Further studies of photo-activation of NHase is required to define the structure of the enzyme fully, since the mechanism of photoactivation is still poorly understood and reports are generally conflicting (Brennan *et al.* 1996; Endo and Odaka 2000; Honda *et al.* 1994; Noguchi *et al.* 1995; Popescu *et al.* 2001; Scarrow *et al.* 1996). Photoregulation and NO-addition both allow points of control of NHase activity, and afford potential in the use of the enzyme in biocatalysis.

In dark conditions the Fe-N (NO) bond is extremely stable (Odaka *et al.* 1997), however light irradiation both breaks the Fe-N covalent bond and also induces local structural changes around the iron center (Nowak *et al.* 2002). A study of photoreactive nitrile hydratase involved analysis of a molecular dynamics simulation of the photodissociated NO ligand in the enzyme channel (Kubiak and Nowak 2008b). This, together with their previous study, also through simulated modeling, suggested that NO induced substantial structural changes upon binding to the iron center by causing an 'inverted doming' effect, indicating that a degree of mechanical change is likely involved in photodissociation (Nowak *et al.* 2002). Modeling results by Rose *et al*. 2010 suggest that NO photolability is strongly related to Cys-S oxygenation since S-oxygenation weakens Fe−S bonding which allows electron transfer at strong transitions near 470 nm (Rose *et al.* 2010).

The biological role of the NO photo response of nitrile hydratases is not known. Bacterial NO metabolism is generally related to nitrogen assimilation through the denitrification process which is a mode of respiration of many bacteria during times of reduced oxygen tension or under anaerobic conditions; by transforming oxyanions of nitrogen to N_2. The reaction reverses nitrogen fixation in the biogeochemical nitrogen cycle, which is sustained by Prokaryotes and is a central part of the overall global nitrogen cycle. Denitrification is controlled by many metalloenzymes, including nitrate and nitrite reductases, and nitric and nitrous oxide reductases (Zumft 1993). Nitric oxide is an obligatory intermediate in denitrification, as well as a signaling and defense molecule of major importance. Bacteria are known also to be sensitive to high NO concentrations due to its reactivity and membrane permeability (Poole 2005), but the basis of resistance to NO and other reactive nitrogen species is poorly understood in many microbes. The cellular targets of NO and interacting proteins include reactive thiols, heme groups, iron-sulfur clusters, phenolic or aromatic amino acid residues, tyrosyl radicals, and amines, in metalloproteins, thiol proteins, glutathione and homocysteine. These protein targets may themselves in turn act as signal transducers, sensing NO and RNS, and resulting in altered gene expression and synthesis of protective enzymes (Rodionov *et al.* 2005). NO may modify the activity of regulatory proteins via direct reaction with the heme moiety, or indirectly, via S-nitrosylation of thiol groups or nitration of tyrosine residues. NO can interact with iron contained in heme proteins such as guanylyl cyclase, which accounts for many of its roles in physiological signal

transduction (Murad 1994), but although guanylyl cyclase is activated by NO, NO–heme interactions can result in the inactivation of other heme proteins, such as catalase and cytochrome systems (Kim *et al.* 1995). The biological significance of a light activated NO bound metalloenzyme remains an interesting point of further study.

4.1.1.9. Catalytic mechanism of NHase

In the last few years, several simple catalytic mechanisms have been proposed for NHase on the basis on theoretical modelling studies, synthetic models, kinetic and spectroscopic studies (Desai and Zimmer 2004; Harrop and Mascharak 2004; Kovacs 2004; Sugiura *et al.* 1987), theoretical density functional calculations (Hopmann *et al.* 2007; Hopmann and Himo 2008), molecular dynamics simulations (Kubiak and Nowak 2008a; Peplowski *et al.* 2008) and substrate docking studies (Peplowski *et al.* 2007). NHase X-ray crystal structures have further added to the mechanism theories (Huang *et al.* 1997; Kobayashi and Shimizu 1998).

One mechanism, proposed based on observations of changed electronic absorption and EPR spectra on nitrile addition (Sugiura *et al.* 1987), was that the substrate nitrile moiety binds directly to metal center, displacing active site water molecule/hydroxide group. This allows the metal ion to act as Lewis acid, activating coordinated nitrile towards nucleophilic attack (Desai and Zimmer 2004). The nucleophilic water/hydroxide is likely provided or activated by an active-site base. In all cases, the metal ion was suspected to function as a Lewis acid. The general reaction included binding of a nitrile directly to the metal, facilitating nucleophilic attack of a water molecule on the nitrile carbon. The metal activates the water molecule which directly, or indirectly, attacks nitriles trapped near the metal.

The direct interaction of the nitrile with the active-site metal was supported by structural studies and synthetic model studies that showed nitriles can readily exchange with low spin Co(III) and Fe(III) centres (Shearer *et al.* 2001; Shearer *et al.* 2002). An X-ray crystal structure of wild-type NO-bound Fe-type NHase from *Rhodococcus erythropolis* showed that NO was bound in place of the metal-coordinated water molecule (Nagashima *et al.* 1998b). The X-ray crystal structure of Co-type NHase from *P. thermophila* bound by the weak inhibitor, *n*-butyric acid, revealed that the metal-coordinated water molecule was displaced by a

carboxylate oxygen atom that binds to the metal ion (Miyanaga *et al.* 2004). Moreover, theoretical modelling studies of the NHase from *Rhodococcus erythropolis* have suggested that nitriles can occupy the active site and be bound by the active site metal (Desai and Zimmer 2004; Greene and Richards 2006). Reaction mechanisms of nitrile hydratase proposed on the basis of butyric acid bound at the active site should, however, be taken with caution since ferric nitrile hydratase was found to interact with butyric acid with a high affinity and is a competitive inhibitor of some substrates (Kopf *et al.* 1996). Most of the early nitrile hydratase mechanisms proposed referred to Fe-type NHase enzymes purified and stored in high concentrations of *n*-butyric acid (Nagasawa *et al.* 1986; Nagasawa *et al.* 1987). The interaction of the NHase active site with *n*-butyric acid was shown to affect the spectral properties and activity of the enzyme (Kopf *et al.* 1996), and should therefore be taken into consideration when assessing proposed mechanisms.

Ligand exchange reaction of low-spin Co-type NHase model complexes suggested that the *trans*-thiolate sulfur played an important role in promoting ligand exchange at the sixth site (Shearer *et al.* 2001). It was also shown that protonation/deprotonation states of the sulfenate oxygen were modulated by the unmodified Cys thiolate ligand (Lugo-Mas *et al.* 2001). Supported by several studies, oxidized cysteine ligands, especially the cysteine sulfenic acid ligand, play an important role in catalysis (Heinrich *et al.* 2001; Noverson *et al.* 1999; Tyler *et al.* 2003). Post-translational modification of cysteine ligands are essential for catalytic activity (Murakami *et al.* 2000), and oxidation by 2-cyano-2-propyl hydroperoxide of specifically the cysteine-sulfenic acid ligand to cysteine-sulfinic acid (αCys-SOH to αCys-SO_2H) in Fe-type NHase results in an irreversible inactivation of the enzyme (Tsujimura *et al.* 2003).

4.1.1.10. Residues participating in recognition of substrate

Miyanaga *et al.* reported the mutational and structural analysis of substrate binding and metal specificity of cobalt NHase (wild-type, apoenzyme and various mutants) from *Pseudonocardia thermophila* JCM 3095 (Miyanaga *et al.* 2004). Mutants involved in substrate binding and catalysis, and formation of the active site centre were constructed, characterized and the structures investigated. Three residues in particular, βLeu48, βPhe51

and βTrp72, form a hydrophobic pocket that likely accommodates the aromatic ring or alkyl chain of the nitrile substrate (Miyanaga *et al.* 2001). Residues βLeu48, βPhe51 and βPhe37 are involved in formation of a hydrophobic environment around the *n*-butyric acid alkyl group. It appears that the hydrophobic pocket of Co-type NHase is bigger than that of Fe-type NHase, and may explain the difference in preference of Co-type NHase for aromatic substrates as opposed to aliphatic.

The three conserved residues that putatively determine the substrate specificity of Co-LNHase from *P. thermophila* (βLeu48, βPhe51 and βTrp72) are also conserved in Co-LNHase from *R. rhodochrous* J1. The kinetic parameters for similar sized substrates are also similar for *P. thermophila* and Co-type LNHase from J1 (Wieser *et al.* 1998). In H-NHase J1 (Nagasawa *et al.* 1991), kinetic parameters for small nitriles are similar to those for *P. thermophila* NHase, but the K_m values for aromatic nitriles are 1000x higher. In H-NHase J1, Trp and Ser at positions corresponding to βLeu48 and βLeu51 respectively of *P. thermophila* NHase, presumably play a role in higher aromatic substrate specificity. *P. chlororaphis* NHase minimally hydrates aromatic nitriles (Nagasawa *et al.* 1987), and contains Val, Val and Tyr at residues corresponding to βLeu48, βPhe51 and βTrp72 respectively (these 3 residues are conserved in all Fe-type NHases), and form a narrower substrate pocket (Miyanaga *et al.* 2001).

Kinetic parameters of *P. thermophila* Co-type NHase studied by Miyanaga *et al.* showed Co-NHase had significantly smaller K_m values for aromatic than small aliphatic nitriles and was inhibited competitively by *n*-butyric acid (K_i of 1.3 mM), similar to *Rhodococcus* sp. N-771 NHase at 1.6 mM (Miyanaga *et al.* 2004; Piersma *et al.* 2000). Generally, a stronger inhibition by aromatic organic acids than aliphatic was observed. In Fe-type NHase, *n*-butyric acid is competitive inhibitor but also a stabilizer (Kopf *et al.* 1996; Nagasawa *et al.* 1987), and activity gradually decreases when there is a lack of *n*-butyric acid. In Co-type NHase, native enzyme is more stable, for example, *P. thermophila* NHase retained complete activity for a year stored at 4 °C with an unchanged catalytic centre (Miyanaga *et al.* 2004).

The proposed reaction model for NHase involves the formation of an imidate intermediate (Kobayashi and Shimizu 1998); n-butyric acid may inhibit as an imidate intermediate

analogue. *n*-Butyric acid at the active site of *P. thermophila* Co-NHase (Miyanaga *et al.* 2004) showed two conformations present in mixed state, designated type I and II. In the type I state, one carboxylic group O atom H-bonds with βTyr68 (2.9 Å), presumed to be the key residue in substrate binding and/or catalysis. This Tyr residue is fully conserved in all NHases.

The βTyr68 also forms another H-bond (2.55 Å) with the O atom at the side chain of αSer112; this H-bond is present in native (Miyanaga *et al.* 2001) and Fe-type NHase (Nagashima *et al.* 1998b). βTyr68 was clearly shown to be important for substrate binding and catalysis, and mutation of the residue (βTyr68 to βY68F) significantly decreased activity compared to wild-type enzyme (Miyanaga *et al.* 2004). In the type II state, one carboxylic group O is positioned a short distance (~1.4 Å) from the O atom of αCys-SOH113, possibly indicating formation of a covalent bond between the 2 O atoms, yet to be confirmed. In both types however, other carboxylic O atoms directly coordinate to the Co ion (2.6 Å in type I and 2.3 Å in type II) instead of a water molecule as in the native structure. This carboxylic O atom is trapped by 3 O atoms of the claw setting (αCys-SO_2H111, αSer112 and αCys-SOH113).

4.1.1.11. Docking studies

Several docking studies proposed to offer insights into the mechanism of NHase. Docked model substrates in the Co-NHase from *P. thermophila* JCM 3095 (Peplowski *et al.* 2007) correlated well with measured activities (Miyanaga *et al.* 2001), leading to the proposal that it is the conformational freedom allowed of aliphatic compounds at the active site that lead to higher activity, and that spatial restrictions at the active site cavity impede aromatic substrates. This is in agreement with the results of conformational searches for Fe-type NHases done by others (Desai and Zimmer 2004). Aromatic nitriles and amides have smaller number of orientations closer to the active site then aliphatic. Both Miyanaga and Peplowski used the Co-NHase from *P. thermophila* JCM 3095, with a 100x higher efficiency for aliphatic nitriles than aromatic, to propose their mechanisms (Miyanaga *et al.* 2004; Peplowski *et al.* 2007). In studied docking scenarios of the amides, energetically favoured positions of amides (aliphatic and aromatic) were oriented towards αSer112 by the amide group. However, only

small aliphatic amides were interacting with sulfinic and sulfenic oxygens of the claw setting of NHase proposed by Nagashima *et al.* (Nagashima *et al.* 1998b).

4.1.1.12. The Hashimoto Mechanism

The so-called 'Mitra and Holz mechanism' was based on theoretical kinetic parameter studies of Co-type NHase from *Pseudonocardia thermophila* JCM 3095 (Mitra and Holz 2007) and previously reported X-ray structural data, spectroscopic, theoretical modelling and synthetic model complex studies (Desai and Zimmer 2004; Harrop and Mascharak 2004; Kovacs 2004; Sugiura *et al.* 1987). Presuming that nitriles bind directly to the trivalent metal ion active site, the Mitra-Holz mechanism could be supported by some kinetic, EPR, UV-visible and theoretical modelling studies (Desai and Zimmer 2004; Harrop and Mascharak 2004; Kovacs 2004; Sugiura *et al.* 1987). However, a detailed and robust catalytic mechanism was still lacking. The NHase mechanisms of action proposed by previous researchers has now been superceded by a sophisticated time-resolved X-ray crystallography experiment performed by Hashimoto and colleagues (Hashimoto *et al.* 2008).

The NHase catalytic mechanism (Figure 4.6) proposed by Hashimoto (Hashimoto *et al.* 2008) presents structural evidence for the mechanism of Fe-type NHase from *R. erythropolis* N771, which converts isonitriles to corresponding amines (Taniguchi *et al.* 2008). Fe-type NHase from *R. erythropolis* N771 possesses a unique photoreactivity. The enzyme is activated by photo-induced denitrosylation, and inactivated by nitrosylation in the dark (Bonnet *et al.* 1997; Noguchi *et al.* 1996; Odaka *et al.* 1997). The NHase converts *tert*-butylisonitrile (*t*BuNC) to *tert*-butylamine with slow reactivity, and together with the photoreactivity of nitrosylated NHase, allowed a mechanistic study of the enzyme by time-resolved X-ray crystallography. The results showed that the isonitrile carbon was coordinated to the iron, and then attacked by a solvent molecule which is activated by αCys^{114}-SOH (sulfenic cysteine/*s*-hydroxycysteine/CSO). It also showed that the sulfenate group of αCys114-SO^- plays a key role in catalysis.

Protein structure is highly conserved amongst known NHases (Hourai *et al.* 2003; Huang *et al.* 1997; Miyanaga *et al.* 2001; Nagashima *et al.* 1998b), and the related thiocyanate

hydrolase enzyme (Arakawa *et al.* 2007). The distorted octahedral geometry of the metal-binding site is also conserved. The active site motif Cys^1-Xaa-Leu-Cys^2-Ser-Cys^3, is strictly conserved. There are six ligand sites: the metal ion coordinates to two amide nitrogens of Ser and Cys^3, and three Cys sulphurs (Huang *et al.* 1997), while the sixth ligand site is occupied by a water/solvent molecule (Miyanaga *et al.* 2001), or NO in the case of nitrosylated Fe-type NHase (Nagashima *et al.* 1998b). Cys^2 and Cys^3 are post-translationally modified to cysteine-sulfinic acid and cysteine-sulfenic acid respectively (Nagashima *et al.* 1998b), and are likely in the deprotonated form at the metal active site (Noguchi *et al.* 2003).

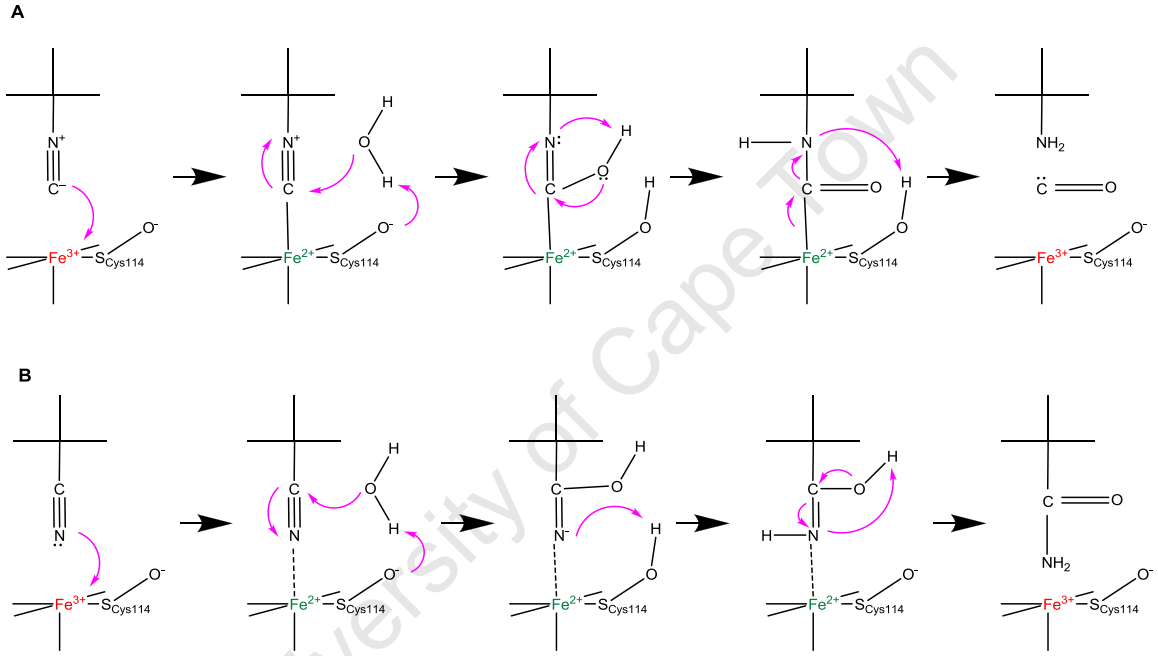


Figure 4.6: Nitrile hydratase mechanism proposed by Hashimoto *et al.* 2008.

A. Isonitrile hydrolysis. B. Nitrile hydrolysis. Re-drawn and adapted slightly from (Hashimoto *et al.* 2008).

4.1.2. Objectives

Given the extensive possible applications of NHase activities in the fine chemical and pharmaceutical industries, there are drives to find new sources of NHase, and NHases with different activities. The biotransformation capabilities of NHase from BAA-870 have already been explored (Brady *et al.* 2004b; Chhiba *et al.* 2012; Kinfe *et al.* 2009), and are highlighted in Chapter 1. The sequencing of BAA-870 (Chapter 2) has shown there is one NHase present

in the genome. The isolation of the NHase gene by cloning, and expression of the NHase protein in *E. coli* will allow easier routes of purification, and access to the enzyme without possible interfering activities from the whole cell transformations used previously. We aim to have an easily expressible NHase, and a route for purification of the enzyme. Insights into the structure of the NHase from BAA-870 will also give possibilities for site-directed mutagenesis for improving or changing the current activity of the enzyme.

4.2. Materials and Methods

4.2.1. Reagents

All restriction endonucleases were purchased from Inqaba Biotech, or Novagen. All chemicals were of molecular grade and were purchased from Sigma-Aldrich, unless otherwise stated. DNA polymerase, ligase and other molecular biology enzymes were obtained from Inqaba Biotech. Plasmids and expression strains were purchased from Novagen.

4.2.2. Methods Outline

Figure 4.7 shows the chronology of the research towards isolation of the NHase gene and protein.

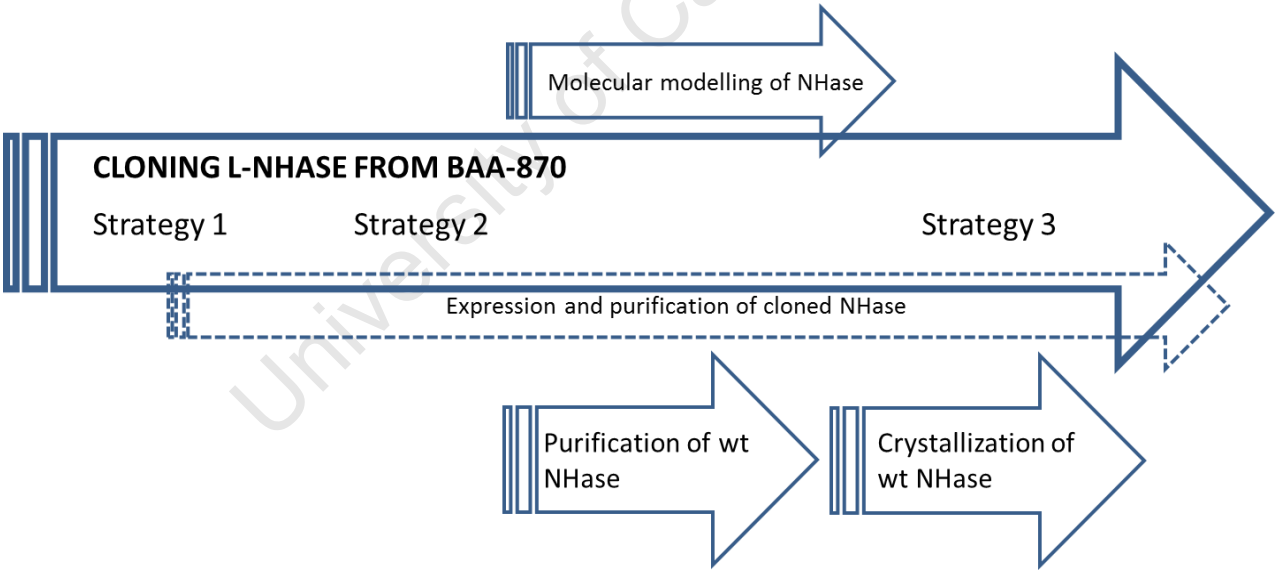


Figure 4.7. Outline of cloning and purification of L-NHase showing the relative time line of strategy used to isolate the gene and enzyme.

The cloning of NHase involved three strategies in chronological order, each strategy an improvement on the last in order to solve expression and optimise the isolation of NHase. Molecular modelling of NHase gave insights into cloning strategy three. A purification of wildtype NHase was done simultaneously during cloning procedures in order to find a suitable purification strategy and crystallization conditions.

4.2.3. Cloning and expression of the wild-type nitrile hydratase gene cassette from *R. rhodochrous* ATCC BAA-870

4.2.3.1. Cloning

Polymerase chain reaction

Primers were designed according to the sequences of *R. rhodochrous* J1 low and high molecular weight cobalt-containing nitrile hydratases, assuming that BAA-870 also contained both low and high molecular weight NHases. Restriction sites were added to the primers for directional cloning into vectors (*Nde*I 5'-CATATG-3' in the forward, and *Hin*dIII 5'-AAGCCT-3' in the reverse).

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'

Amplification of the low molecular mass nitrile hydratase gene cassette (including the beta subunit, alpha subunit, and beta-homolog protein/chaperone) was done with KapaHiFi Polymerase (Kapa Biosystems, South Africa) using a PeQLab primus 25 thermal cycler (Biotechnologie GmbH, Germany). Cycling conditions were as follows: initial denaturation at 95 °C for 2 minutes; followed by 25 cycles of denaturation at 98 °C for 20 seconds, annealing at (Tm - 5 °C) for 15 seconds and extension at 68 °C for 1 min/kb. Control reactions where primers and DNA template were omitted respectively, were performed.

4.2.3.2. Purification and blunt-end cloning of amplified clone insert

DNA loading buffer was added to the PCR reactions. Samples were loaded into wells of a 0.8 % agarose gel alongside PstI DNA marker, and run at 100 V. PCR reactions that showed an insert of the correct size were excised from the gel, and purified using a QiaQuick gel extraction kit (Qiagen, Germany). Insert was eluted using 20 mM Tris buffer, pH 8.0. Purified

insert was ligated into pJET1 (Fermentas, Lithuania) using T4 DNA ligase. A control experiment was performed using a control insert provided with the CloneJET PCR cloning kit (Fermentas, Lithuania). Ligated insert/plasmid was transformed into '*E. cloni* 10GF' DUO competent cells (from Lucigen, USA) and plated on LB-ampicillin agar plates.

4.2.3.3. Confirmation of blunt-cloned insert

LB-ampicillin media was inoculated with randomly selected transformation colonies and grown overnight at 37 °C with shaking (200 rpm). Plasmids were purified using a Zyppy™ plasmid miniprep kit (from Zymo Research, USA). PCR was used to amplify the nitrile hydratase insert using the original amplification primers to confirm presence of the insert of the expected size.

4.2.3.4. Restriction digestion

Purified plasmids (controls and reactions) were digested using *Bgl*II, which cuts on either side of a blunt-cloned insert in pJET1. The insert was cut from the plasmid for directional cloning into an expression vector using the restriction enzymes *Hin*dIII and *Nde*I, purified, and concentrated for ligation. Cloning vectors pET20b(+) and pET28a(+) were purified, restricted with *Hin*dIII and *Nde*I and phosphorylated. All restriction digestions were analysed by agarose gel electrophoresis.

4.2.3.5. Ligation of insert and expression vector

Nitrile hydratase insert restricted with *Hin*dIII and *Nde*I was ligated with pET28a(+) using T4 DNA ligase. Ligations were purified, transformed into *E. coli* BL21(DE3) competent cells and plated on LB-kanamycin agar.

4.2.3.6. Analysis of recombinant clones

Confirmation of insert ligation into vector was analysed by restriction digestion and PCR using the original amplification primers. Selected clones containing insert were sequenced by Inqaba Biotech (Pretoria, South Africa).

4.2.3.7. Expression

See Appendix 7.3 for plasmids and vector details.

4.2.3.8. Expression of nitrile hydratase in pET20b(+)

Clones were transformed into *E. coli* BL21(DE3) expression cells, and grown in the presence of ampicillin. Overnight inoculums grown at 37 °C with shaking, were used to inoculate LB-ampicillin. Cultures were grown to an OD_{600} of 0.4 – 0.6, and various concentrations of $CoCl_2$ (0.05, 0.1, 0.2 mM) were added half an hour before protein induction. Induction of protein was initiated by addition of different concentrations of IPTG (with a range of 0.5 – 4 mM), and cultures were grown at 25 °C – 37 °C, with shaking, to test the effects of different concentrations of inducer.

4.2.3.9. Expression of nitrile hydratase in pET28a(+)

Clones were transformed into *E. coli* BL21(DE3) expression cells, and grown in the presence of kanamycin. Overnight inocula grown at 37 °C with shaking, were used to inoculate LB-kanamycin. Cultures were grown to an OD_{600} of ~0.3, and 0.1 mM $CoCl_2$ was added. Cultures were grown for half an hour at 37 °C with shaking before induction of protein was initiated by addition 0.1 mM IPTG. Cultures were induced at 23 °C, with shaking at 180 rpm, for at least 12 hours.

4.2.4. Improving NHase expression by ribosome binding site (RBS) mutagenesis

4.2.4.1. Mutagenesis of the alpha subunit RBS

L-NHase-containing clones prepared as described above were mutated at the ribosome binding site for the alpha subunit using the Phusion® Site-Directed Mutagenesis Kit (Invitrogen, USA). The desired point mutation was verified by DNA sequencing of clones. The mutagenesis reaction, however, incorporated undesired mutations at the start codon of the gene insert. The mutated insert was amplified using the original cloning primers containing the *Nde*I and *Hin*dIII cloning sites, and re-ligated into pJET1.2 for cloning and screening of the

correct clones, before the correct insert was re-cloned into the expression vector pET28a(+). Expression of the mutated clones was done using the protocols described above.

4.2.5. Cloning and co-expression of NHase in *E. coli* using a two-construct system

4.2.5.1. Cloning of nitrile hydratase in pRSFDuet-1 and pET21a(+)

Polymerase chain reaction

Primers were designed to amplify the sequence of *R. rhodochrous* ATCC BAA-870 low molecular weight cobalt-containing nitrile hydratases, with restriction sites added for directional cloning into vectors (Table 4.2). The pRSFDuet-1 cloning restriction sites included *Nde*I at the forward position and either *Pac*I or *Eco*RV in a reverse primer, for cloning of the alpha-chaperone sequence. Two reverse primers were designed due to difficult sequence design options in that region. The pET21a(+) cloning restriction sites included *Nde*I at the forward position and *Xho*I at the reverse primer, for cloning of the beta subunit sequence.

Table 4.2: List of primers used for coexpression amplification and directional cloning

Primer name	Target for PCR amplification	Primer sequence (5' – 3')	Restriction sites added
pRSF Forward NdeI	Alpha-chaperone	TAGCTACATATGACCGCCCACAATCCC	5' – *Nde*I-alpha
pRSF Reverse PacI	Alpha-chaperone	GCTATTAATTAACTACCCGTCGG	Chaperone-*Pac*I – 3'
pRSF Reverse EcoRV	Alpha-chaperone	GCATGATATCCTACCCGTCGG	Chaperone-*Eco*RV – 3'
pET21 Forward NdeI	Beta	GCTAGCTACATATGGATGGAATC	5' – *Nde*I-beta
pET21 Reverse XhoI	beta	TATACTCGAGGGCCGGTAGCAGATACG	Beta-*Xho*I – 3'

Amplification of the beta subunit, and the alpha subunit and chaperone of the nitrile hydratase gene cassette was achieved with KapaHiFi Polymerase (Kapa Biosystems, South Africa) using a PeQLab primus 25 thermal cycler. Cycling conditions were as follows: initial denaturation at 95 °C for 2 minutes; followed by 25 cycles of denaturation at 98 °C for 20 seconds, annealing at (Tm - 5 °C) for 15 seconds and extension at 68 °C for 1 min/kb. Control

reactions were performed where primers and DNA template were omitted respectively. Various genomic DNA templates were used to optimise the amplification of products.

Blunt-end cloning of purified PCR products

DNA loading buffer was added to the PCR reactions. Samples were loaded into wells of a 0.8 % agarose gel alongside *Pst*λ DNA marker, and run at 100 V. PCR reactions that showed an insert of the correct size were excised from the gel, and purified using a QiaQuick gel extraction kit (Qiagen, Germany). Insert was eluted using Tris buffer, pH 8.0. Purified insert was ligated into pJET1.2 (Fermentas, Lithuania) using T4 DNA ligase. A control experiment was performed using a control insert provided with the CloneJET PCR cloning kit (Fermentas). Ligated insert/plasmid was transformed into *E. coli* DH5α competent cells (Lucigen, USA) and plated on LB-ampicillin or LB-kanamycin agar plates for selection of beta and alpha-chaperone-containing clones respectively.

Confirmation of blunt-cloned insert

LB-ampicillin or -kanamycin media was inoculated with randomly selected transformation colonies and grown overnight at 37 °C with shaking (200 rpm). Plasmids were purified using a Zyppy™ plasmid miniprep kit (from Zymo Research). Plasmids (pJet1.2 containing insert) were restricted with *Bgl*II to release the insert, and reactions run on 0.8 % agarose gel with 0.5 µg/ml ethidium bromide (Sigma-Aldrich, Germany) to check for the presence of correctly sized inserts.

Restriction digestion

Plasmids containing the expected inserts were cut from the blunt cloning plasmid for directional cloning into the relevant expression vectors. Restriction enzymes *Nde*I, *Xho*I (for the beta insert) and *Pac*I or *Eco*RV (for the alpha-chaperone insert) were reacted with pJet1.2 plasmids for 1 hour at 37 °C to release restricted inserts. Reactions were analysed and purified from 0.8 % agarose gel. Cloning vectors pET21a(+) and pRSFDuet-1 were purified, restricted with the relevant restriction enzymes and dephosphorylated using a standard shrimp alkaline phosphatase treatment (SAP enzyme, Fermentas, Lithuania).

Ligation of inserts and expression vectors

The gene encoding the nitrile hydratase beta subunit was restricted with *Nde*I and *Xho*I and then was ligated into pET21a(+). The gene encoding the alpha-chaperone was ligated into pRSFDuet-1 restricted with *Nde*I and *Pac*I, using T4 DNA ligase (Fermentas, Lithuania). Ligations were purified, transformed into *E. coli* DH5α cells for plasmid maintenance, and BL21(DE3)cells for expression, and plated on LB-ampicillin or -kanamycin agar, respectively.

Sequence confirmation of recombinant clones

Selected insert-containing clones were sequenced by Inqaba Biotech (Pretoria, South Africa).

4.2.5.2. Coexpression of nitrile hydratase in pRSFDuet-1 and pET21a(+)

Plasmids were co-transformed into *E. coli* BL21(DE3) expression cells, and grown in the presence of 50 µg/ml kanamycin. Overnight inocula grown at 37 °C with shaking, were used to inoculate LB-kanamycin. Cultures were grown to an OD_{600} of ~0.3, and 0.1 mM $CoCl_2$ was added. Cultures were grown for half an hour at 37 °C with shaking before induction of protein was initiated by addition 0.1 mM IPTG. Cultures were induced at 23 °C, with shaking at 180 rpm, for at least 16 hours.

4.2.6. Purification of nitrile hydratase

All purification steps are common to purification of wildtype and recombinant nitrile hydratase, unless otherwise described. All chromatography elution fractions were assayed for protein concentration using the Bradford assay from BioRad (Bradford 1976) and activity was measured by HPLC or continuous spectrophotometric assay. Protein was visibly inspected by 12 % SDS-PAGE using the Laemmli discontinous buffer system (Laemmli 1970). All ultrafiltration procedures were performed using an Amicon ultrafiltration stirred cell

system on ice, with a Millipore PM10, 10 kDa, size exclusion membrane under nitrogen gas pressure.

4.2.6.1. Cell-free extract preparation

After a freeze-thaw cycle, cell paste was resuspended in B-PER bacterial protein extraction reagent (Sigma, Germany) containing lysozyme and DNase I (Fermentas, Lithuania) at 10 mg/ml and 20 μg/ml, respectively, and incubated for 30 min at room temperature. The resuspended cells were sonicated for a total of 5 min (with 0.5 s pulsed sonication over 10 min). Cell-free extract was collected after centrifugation of cell debris at 15 000 x g for 30 min, and filtered through a 0.22 μm filter.

4.2.6.2. Column chromatography

Ion exchange chromatography

A Q-Sepharose column (Amersham Biosciences) was equilibrated with two column volumes of 100 mM potassium phosphate buffer pH 7.4, containing 100 mM KCl at a flow rate of 2.5 ml/min. Cell free extract was applied to the column and unbound protein was washed off. Nitrile hydratase was eluted using potassium phosphate buffer pH 7.4, containing 100 mM – 1M KCl, at a flow rate of 2 ml/min. The protein concentration was measured at an absorbance of 280 nm and recorded by an in line UV/Vis detector (Gilson Inc, USA). Fractions (5 ml/tube) were collected using a Gilson FC203B fraction collector.

Size exclusion chromatography

Pooled ion exchange fractions showing nitrile hydratase activity were purified by size exclusion chromatography, and eluted using a Sephacryl S-200 or S-400 HR column (Amersham Biosciences) equilibrated with potassium phosphate buffer, pH 7.4, containing 200 mM NaCl. Flow rates were 0.8 or 0.5 ml/min respectively.

4.2.6.3. Additional steps for purification of wildtype nitrile hydratase

Batch fermentation of *Rhodococcus rhodochrous* ATCC BAA-870

Fermentation media consisted of an autoclaved bulk medium of 14.6 g/L K_2HPO_4, 2g/L $(NH_4)_2SO_4$, 3.6 g/L Na_2HPO_4, 2.5 g/L citric acid, 1.2 g/L $MgSO_4$, 5 g/L NH_4NO_3 and 20 g/L yeast extract, to which a solution of separately autoclaved 36 g/L glucose monohydrate was added, as well as filter sterilized 5 ml/L trace element solution (0.2 g/L $ZnSO_4.5H_2O$, 0.4 g/L $CuSO_4.5H_2O$, 0.004 g/L $CoSO_4.7H_2O$, 50 g/L trisodium citric acid and 0.4 g/L $MnSO_4.H_2O$), 30 µg/mL kanamycin and 20 mM benzonitrile (or other nitrile) as inducer. All nitriles were prepared in 100% methanol. Log phase growth in minimal media was at least 28 - 30 hours at 30 °C.

Ammonium sulphate precipitation

Ammonium sulfate was ground to a fine powder using a mortar and pestle. A 20-70% ammonium sulfate saturation cut was applied to cell free extract using standard ammonium sulfate precipitation techniques. Between each ammonium sulfate addition step, the sample was left on ice for 20 min, followed by centrifugation at 10 000 x g for 10 min at 4 °C to precipitate the saturated protein fraction. The pellet was resuspended in a 5 ml volume of 50 mM potassium phosphate buffer pH 7.4. A 5X salt-free buffer dilution of each fraction (total volume 20 µl) was used to analyse by SDS-PAGE.

Batch ion exchange chromatography

Volumes of clarified supernatant from a 30 % ammonium sulfate fractionation were batch eluted from a 50 ml slurry of Q-Sepharose in 50 mM steps of salt concentration, from 200 mM – 500 mM NaCl. The Q-Sepharose batch exchange media was equilibrated with 50 mM potassium phosphate buffer, pH 7.4.

4.2.6.4. Additional steps for purification of cloned nitrile hydratase from pRSFDuet-1 and pET21a(+)

Nickel affinity column chromatography

Clarified supernatant was applied to Ni-Protino batch resin, which was equilibrated with 50 mM sodium phosphate buffer pH 7.4 containing 500 mM NaCl and 40 mM imidazole. The column was washed until all unbound protein was eluted. Nitrile hydratase was eluted by consecutive washes of the resin with 50 mM sodium phosphate buffer pH 7.4 containing 500 mM NaCl and 500 mM imidazole until all bound protein was removed. Nitrile hydratase-containing aliquots were concentrated using a 10 kD cut-off filtration unit, and applied to an Unosphere Q ion exchange column (BioRad, USA) that was equilibrated with 50 mM sodium phosphate buffer pH 7.4 containing 100 mM NaCl. Nitrile hydratase was eluted using a linear salt gradient that was increased from 100 mM NaCl to 1 M NaCl. Fractions of eluted nitrile hydratase were assessed for purity by SDS-PAGE, and protein concentrations were measured using the standard Bradford protein concentration assay (BioRad, USA).

4.2.7. Continuous Spectrophotometric Assay

The production of corresponding amide from nitrile was measured using a spectrophotometric assay developed specifically for measuring benzamide formation during the purification of NHase. A 1.5 mL reaction total containing 50 mM sodium phosphate buffer, pH 7.4, with 300 µl of a 10 mM nitrile substrate made up in a minimal volume of methanol, was initiated by addition of an appropriate amount of enzyme, and the corresponding amide product measured (Table 4.3). Wavelength scans from 200 – 300 nm were done of nitriles and the corresponding amide, and the wavelength at their maximum absorbance differences were chosen as the wavelength of assay measurement (see Figure 6.9, Appendix). Reactions were scaled accordingly to a 150 µl total volume, and measured at the desired wavelength using UV-compatible 96-well plates (Corning Life Sciences, USA and Canada).

Table 4.3: Parameters for spectrophotometric assay

Substrate	Product	Wavelength at which amide measured (nm)	Molar extinction coefficient of amide (mM^{-1}/cm)
Benzonitrile	Benzamide	252	0.3818
3-Cyanopyridine	Nicotinamide	240	0.8609

4.2.8. HPLC

Nitrile hydratase-containing aliquots (0.1 ml) were added to 0.3 ml assay mix (50 mM potassium phosphate buffer, pH 7.4, containing 5 mM benzonitrile dissolved in a minimal volume of methanol) and reacted for 5 min with shaking (1000 r.p.m.) at 30 °C. Reactions were stopped by addition of 1.2 ml acidified stop solution consisting of 60:40 Acetonitrile : 0.1 % TFA mix, centrifuged for 5 min at 13 200 x g to precipitate particulates, and 1 ml transferred to a 2 ml HPLC vial. Standard curves of nitriles and corresponding amides and acids were constructed from 0 – 5 mM. A Phenomenex Aqua C18 reverse-phased column, 250 mm × 2 mm, was used to analyse retention times and peaks of nitriles, amides and acids with a flow rate of 0.6 ml/min over 5 min using a 70:30 acetonitrile : 0.1 % TFA(aq) mobile phase. In fractions where benzoic acid was detected, the reaction was repeated with an appropriate dilution and reacted for 2 min to detect benzamide before conversion to benzoic acid by contaminating amidase.

4.2.9. Structure and Modeling of NHase

4.2.9.1. Molecular modeling of NHase from *R. rhodochrous* ATCC BAA-870

The BAA-870 NHase alpha and beta subunit sequences were submitted to the mGenTHREADER Fold recognition tool web server for structural sequence alignment to identify protein folds and distant homologues (Jones 1999; McGuffin and Jones 2003). Low complexity regions were masked as a filtering option. Protein secondary structure prediction was done using the PSIPRED (protein secondary structure prediction) server (Jones *et al.* 2000). Similar protein folds were identified in other proteins and homology models were

built using MODELLER (Eswar *et al.* 2001; Martí-Renom *et al.* 2000; Šali and Blundell 1993). Side chain conformations were adjusted using the SCWRL program for prediction of protein side-chain conformations (Altschul *et al.* 1990; Canutescu *et al.* 2003). Resulting models were aligned with the template structure using UCSF Chimera (Pettersen *et al.* 2004). A BLAST of the BAA-870 L-NHase full length protein sequence showed a 100% identity to the sequence of L-NHase from *R. rhodochrous* J1.

4.2.9.2. Attempted crystallization of nitrile hydratase

The wild-type nitrile hydratase was purified as described in Section 4.2.6. Purified nitrile hydratase was noted to be easily precipitated on standing, with increasing instability observed in relation to protein concentration. Hence, limited concentrations were available for crystallography trials. The sequence of BAA-870 nitrile hydratase was submitted to the XtalPred sequence analysis tool (Slabinski *et al.* 2007).

4.3. Results

4.3.1. Nitrile hydratase cloning and expression

4.3.1.1. Cloning and expression of the nitrile hydratase gene cassette from *R. rhodochrous* ATCC BAA-870

The gene encoding the L-NHase cassette (beta and alpha subunits and chaperone) in *R. rhodochrous* ATCC BAA-870 were successfully amplified from the genome (Figure 4.8), and cloned into various cloning and expression vectors. The total size of the NHase gene cassette was 1.8 kb. Failure to amplify the H-NHase was supported by the genome sequence information obtained during this study, since no H-NHase was identified in the assembled and annotated sequence.

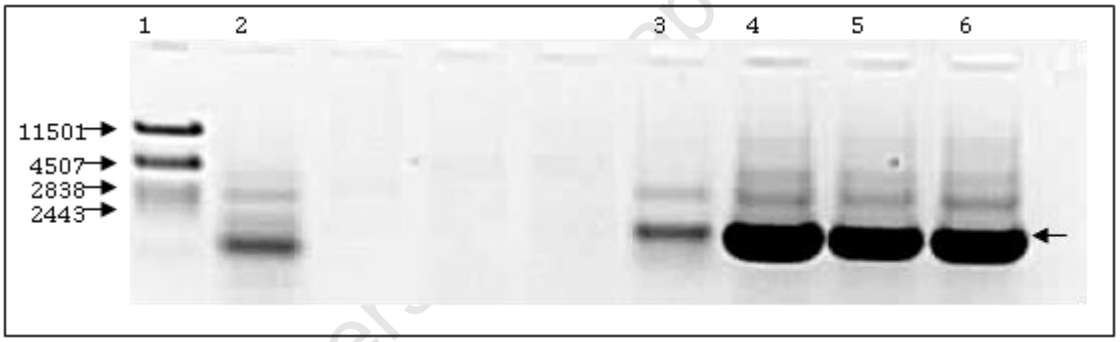


Figure 4.8: Agarose gel (0.8 %) of amplified inserts from pET vectors.

Lane 1, PstΛ DNA marker; Lane 2, insert amplified from pET28a(+); Lanes 3, 4, 5 and 6, nitrile hydratase beta and alpha subunits and chaperone amplified from pET20b(+). The arrow to the right of the gel indicates the amplified nitrile hydratase gene (~1.8 kb).

The NHase gene cassette was expressed using clones contained in both the pET20b(+) and pET28a(+) vectors. However, only expression of the beta subunit and beta subunit homolog (chaperone) was noted (Figure 4.9). Increasing the induction time did not improve expression of the alpha subunit. The expression of the chaperone and beta subunits was not in equal ratio, as noted by the intense protein bands at 16 kDa.

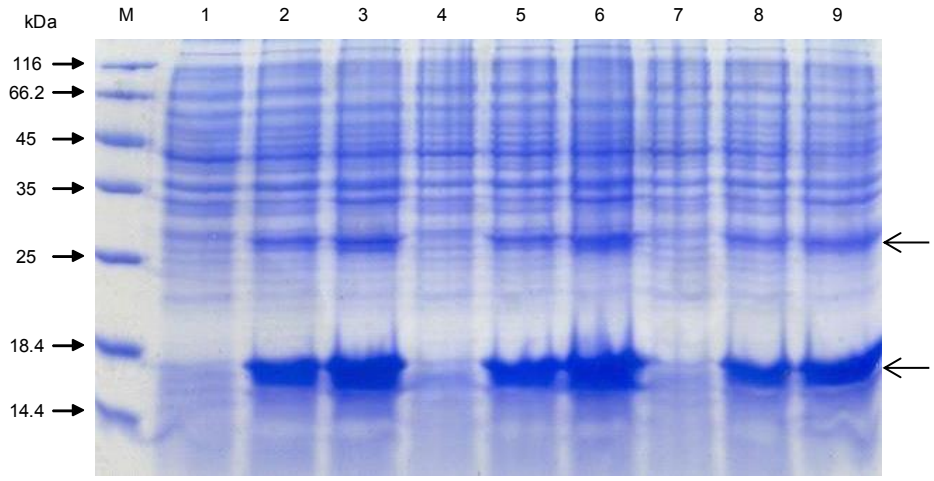


Figure 4.9: Expression of nitrile hydratase beta subunit and the beta homolog using pET-28a(+) showing increased expression of the beta subunit and chaperone protein over time.

Cultures (LB containing kanamycin) were inoculated 1:100 with O/N growth and grown to an OD at 600 nm of ~0.6. IPTG was added to 0.1 mM and induced at 37 °C for various times. Lanes 1, 4 and 7: Uninduced protein sample. Lanes 2, 5 and 8: Induced proteins after 4 hours of induction. Lanes 3, 6 and 9: Induced protein sample after 8 hours of induction. The 26 kDa beta subunit and 16 kDa chaperone subunit are induced (indicated by arrows).

The pET28a(+) and pET20b(+) plasmid constructs containing the genes for the beta and alpha subunit, and chaperone of L-NHase showed only overexpression of the beta subunit (26 kDa) and the chaperone protein (17 kDa). Since the chaperone protein was overexpressed relative to the beta subunit, it was clear that the ribosome binding site (RBS) for the chaperone was recognised by *E. coli* and the ribosome binding sites for each gene in the L-NHase cassette were investigated. The chaperone gene lies in a different frame to the beta and alpha genes, a naturally occurring frameshift. The RBS for the chaperone gene, however, lies in the C-terminus of the alpha gene. As the chaperone was expressing well in the *E.* coli host, the RBS for the alpha subunit was mutated to a sequence identical to that for the chaperone in order to improve expression of the alpha subunit.

Unfortunately the RBS-mutation did not appear to significantly enhance the expression of the alpha subunit (Figure 4.10), although the expression of the chaperone was reduced compared to the beta subunit However, measurement of cell free extract (CFE) by HPLC showed conversion of benzonitrile to benzamide, indicating there was some NHase activity, despite the very low expression of the alpha subunit.

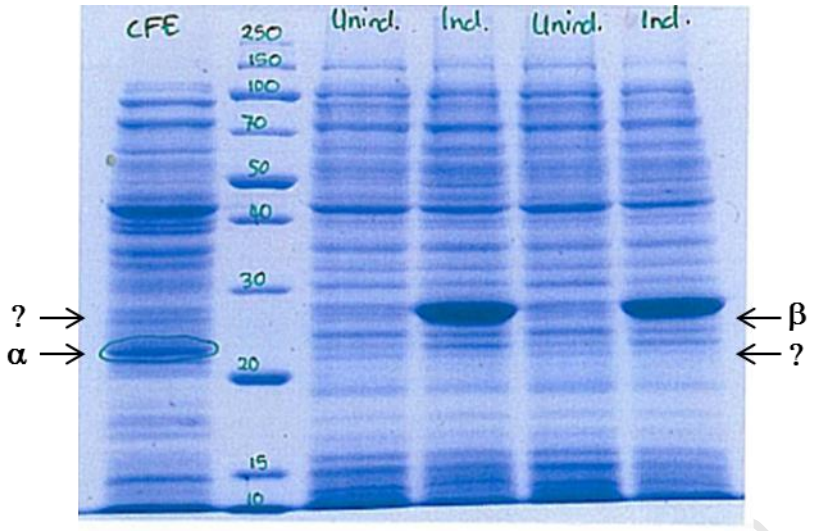


Figure 4.10: Induced and uninduced supernatant samples of the RBS-mutated clones of NHase.

Cell free extract (CFE) shows the soluble protein extracted from disrupted cells. Induced (Ind.) and uninduced (Unind.) samples, analysed by 12 % SDS-PAGE, are total protein samples from induced and uninduced culture medium, normalized for protein concentration. The question mark denotes the absence or low expression of alpha or beta subunit.

Relative to the total protein sample, the CFE sample showed very little beta protein, but appeared to show presence of the alpha subunit. Sequence analysis confirmed that the NHase cassette was inserted correctly and in frame. It was surmised that the beta subunit is relatively insoluble, compared to the expression of the alpha subunit.. SDS-PAGE analysis of the resuspended pellet (the insoluble fraction obtained from disrupted cells) showed a large proportion of insoluble beta subunit (Figure 4.11). The beta subunit was therefore considered to be unstable, misfolding, or expressed into inclusion bodies. Although it appeared that both the alpha and beta subunits were overexpressed, there was only very little activity of NHase as measured by HPLC.

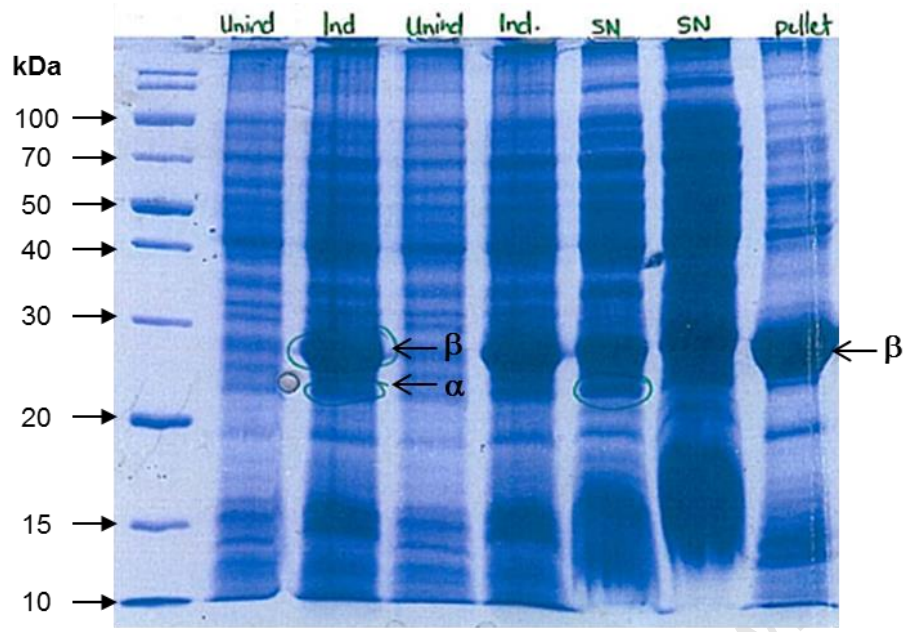


Figure 4.11: Uninduced and induced pellet samples of the RBS-mutated clone of NHase.

Total protein samples were analysed by 12 % SDS-PAGE. SN refers to supernatant, and only shows the soluble protein extracted from disrupted cells. Pellet is insoluble matter obtained from centrifugation after cell disruption. Uninduced (Unind) and induced (Ind) samples are total protein samples from induced and uninduced culture medium, normalized for protein concentration.

4.3.2. Cloning and expression for co-expression of NHase

The cloning and expression strategy for NHase was revisited. The RBS for the alpha subunit was mutated to a recognized *E. coli* RBS, however, the alpha and beta subunits were still not expressed in equal ratio, likely caused by ribosome stalling during translation. A co-expression strategy was used next to ensure that both the beta and alpha subunit expression was independent of each other and under the control of the expression plasmid promoter.

4.3.2.1. Cloning NHase for coexpression

The alpha subunit-chaperone sequence and beta subunit sequence were amplified separately from the genome of *R. rhodochrous* ATCC BAA-870 using PCR. The 1060 bp product consisting of the amplified alpha subunit and chaperon was purified and the size verified on an agarose gel (Figure 4.12). Furthermore, the insert was blunt cloned into the

pJet vector and verified by restriction digestion using the restriction enzyme *Bgl*II which releases blunt cloned inserts from the vector.

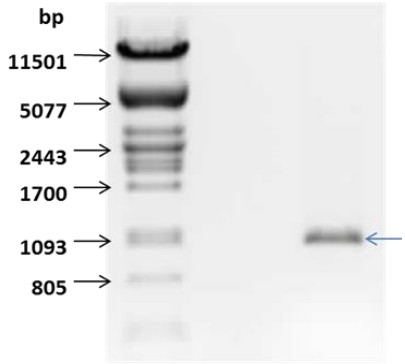


Figure 4.12: Purified alpha-chaperone insert amplified by PCR using BAA-870 genomic DNA template.

The 0.8 % agarose gel was run at 100V. The 1060 bp amplified insert is indicated by arrow. The DNA size marker is Pstλ-DNA.

After directional cloning into the expression vector pRSFDuet-1, the correct insert was again verified by restriction digestion using the directional cloning restriction sites incorporated in the original PCR primers. The clone containing the alpha-chaperone insert was linearised to ensure the correct vector size was obtained when the insert is present (Figure 4.13).

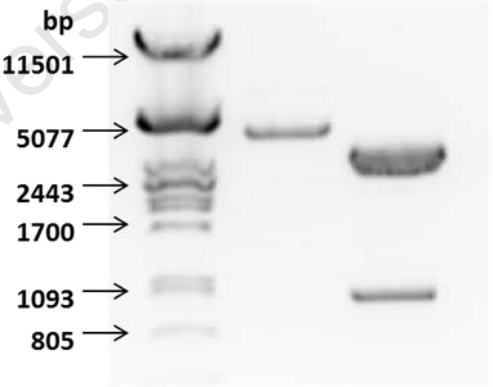


Figure 4.13: Alpha-chaperone insert verification in the expression vector pRSFDuet-1.

The vector was cut with only *Nde*I (left lane) to verify the total size was consistent with vector and insert. The vector was cut with *Nde*I and *Pac*I (right lane) releasing the 1060 bp alpha-chaperone insert. The 0.8 % agarose gel was run at 100 V. The DNA size marker is Pstλ-DNA.

Clones containing the correct insert were verified by sequencing and used for further protein expression studies. The strategy for cloning the beta subunit into pET21a(+) was followed exactly as for the alpha-chaperone, and verified at each step by restriction digestion (results not shown).

4.3.2.2. Coexpression of NHase

The 26 kDa beta subunit, and the 23 kDa alpha subunit together with the 16 kDa chaperone, were successfully expressed from their independent expression vectors. Trial inductions were done to test the separate expression of the beta, and alpha-chaperone constructs (Figure 4.14).

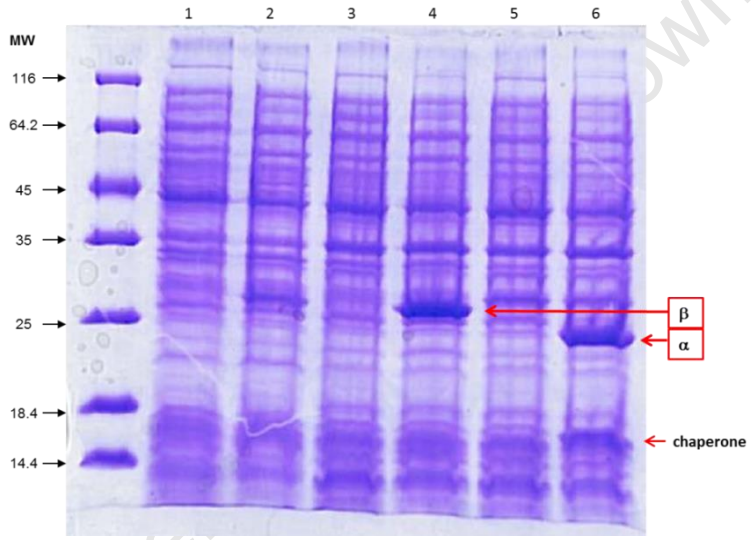


Figure 4.14: Separate expression of the NHase beta and alpha subunits using different expression constructs.

Beta and alpha-chaperone coexpression was induced in BL21(DE3) *E. coli* by addition of 0.1 mM IPTG and 0.1 mM $CoCl_2$ and expressed from pET21a(+)and pRSF-Duet vectors, respectively, for 4 hours at 37 °C. Lanes 1, 3 and 5: uninduced. Lane 4: Induction of the beta subunit. Lane 6: Induction of the alpha subunit and chaperone protein.

The pRSF-Duet construct expressing alpha-chaperone, and the pET21a(+) construct expressing the beta subunit, could be independently induced in separate *E. coli* BL21(DE3) cultures. The independent expression of each subunit was good from each expression plasmid. The plasmid constructs for beta and alpha-chaperone were transformed into BL21(DE3) using at first a 1:1 ratio of plasmid. Different ratios of plasmid were also transformed to find the optimum ratio for near to equal expression of beta and alpha

subunits. It was found that a ratio of 4:1 pRSFDuet-alpha chaperone: pET21a(+)-beta showed good coexpression of both subunits (Figure 4.15).

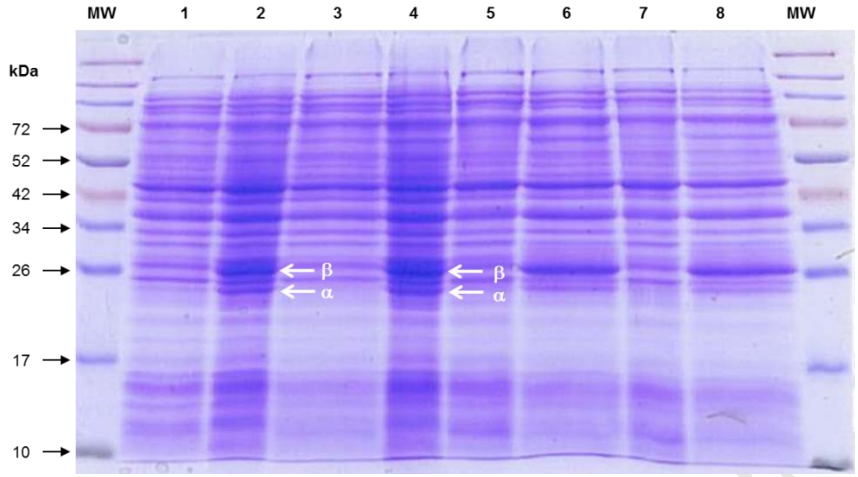


Figure 4.15: Coexpression of the NHase beta and alpha subunits using different concentrations of expression vector and different induction conditions.

Beta and alpha-chaperone coexpression was induced in BL21(DE3) *E. coli* by addition of 0.1 mM IPTG and 0.01 mM $CoCl_2$. The concentration of pET21a(+)-beta expression vector was 4X less than the concentration of the pRSF-Duet alpha-chaperone vector. Lanes 1, 3, 5 and 7: uninduced. Lanes 6 and 7: Induction at 37 °C for 4 hours. Lanes 2 and 4: Induction at 25 °C for 12 hours.

Reducing the temperature of expression, and increasing the expression time gave optimal coexpression of the alpha and beta subunits (Figure 4.16).

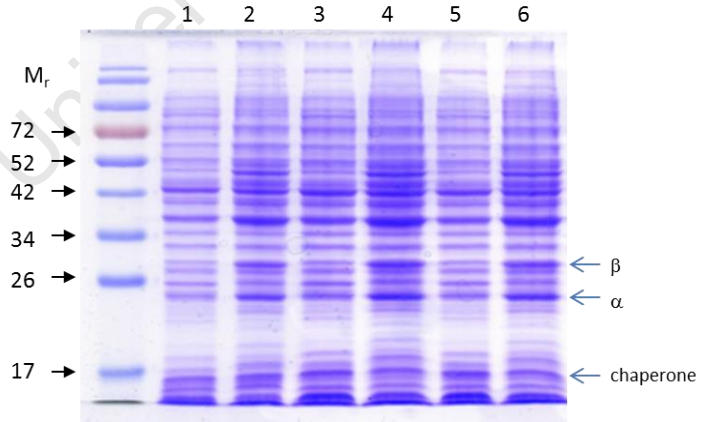


Figure 4.16: Coexpression of the NHase beta and alpha subunits using a 4:1 pRSFDuet-alpha chaperone:pET21a(+)-beta expression vector ratio.

Beta and alpha-chaperone coexpression were induced in BL21(DE3) *E. coli* containing kanamycin and ampicillin, by addition of 0.1 mM $CoCl_2$ 30 minutes prior to addition of 0.1 mM IPTG. The concentration of pET21a(+)-beta expression vector was 4X less than the concentration of the pRSF-Duet alpha-chaperone vector. Lanes 1, 3 and 5: uninduced. Lanes 2, 4 and 6: Different clones induced at 200 rpm at 20 °C for 12 hours.

Induced and uninduced fractions were measured for benzonitrile conversion by HPLC (Figure 4.17) to confirm the presence of active NHase.

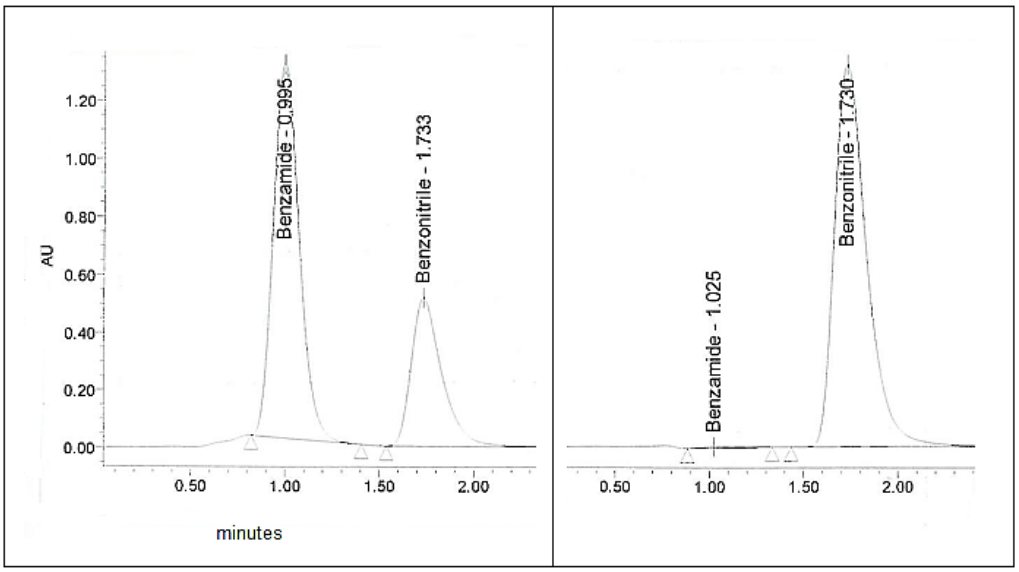


Figure 4.17: HPLC trace showing benzonitrile and benzamide elution over time.

Left panel: Reaction of NHase containing fraction showing conversion of benzonitrile to benzamide (eluting at 0.995 minutes). Right panel: benzonitrile control reaction, eluting at 1.73 minutes. Reactions were carried out by addition of benzamide assa solution to an aliquot of enzyme containing solution, and reacted at 37 °C for 5 minutes with shaking (200 rpm). Reactions were stopped by addition of acidified eluent.

4.3.3. Purification of Nitrile Hydratase

4.3.3.1. Purification of NHase

In following purification attempts, the stability of NHase upon purification was problematic. The un-coupling of subunits was common, with loss of the beta subunit a frequent occurrence. When pure enzyme was obtained, the protein was seen to precipitate from solution readily, with 'purer' enzyme preparations subject to more precipitation. Also, hydrophobic interaction chromatography was deemed a necessary purification step to remove 'stubborn' impurities from the enzyme preparation. However, purification steps involving ammonium sulphate (precipitation and chromatography) were particularly detrimental leading to the loss of the beta subunit.

First NHase purification trials involved a traditional apporach: ammonium sulphate fractionation (Figure 4.18) to remove most impurities, followed by hydrophobic interaction (Figure 4.19) and ion exchange chromatography. However, steps involving ammonium sulphate revealed instabilities in the interaction between alpha and beta, and the ratios of each subunit were disrupted (Figure 4.20).

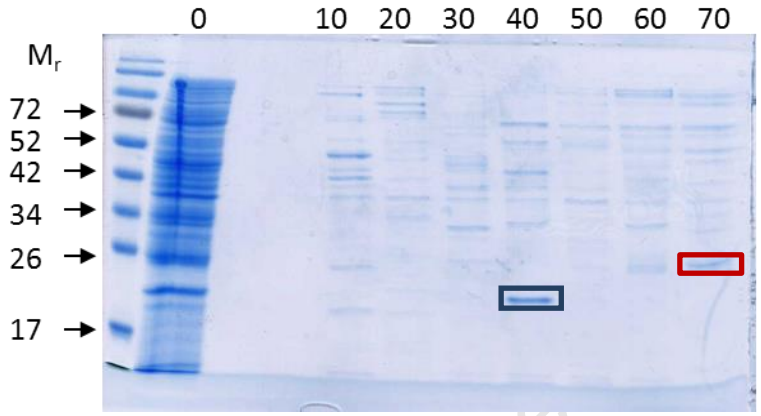


Figure 4.18: Example of ammonium sulfate precipitation fractions for purification of nitrile hydratase.

Ammonium sulphate fractionation decouples the alpha and beta subunit. Lane 1: starting material containing NHase alpha (23 kDa), blue box, and beta (26 kDa), red box, subunits. Lanes 10 – 70: 10-70% ammonium sulphate precipitated fractions.

Hydrophobic interaction chromatography of cell free extract showed NHase activity elution at 100 – 0 mM ammonium sulphate, indicating a relatively hydrophobic protein. However, subunit de-coupling was evident, and much protein was lost due to instability. Further purification steps using pooled fractions eluted by hydrophobic interaction column chromatography led to complete loss of NHase activity, and only the alpha subunit remained, indicating the beta subunit was lost (data not shown).

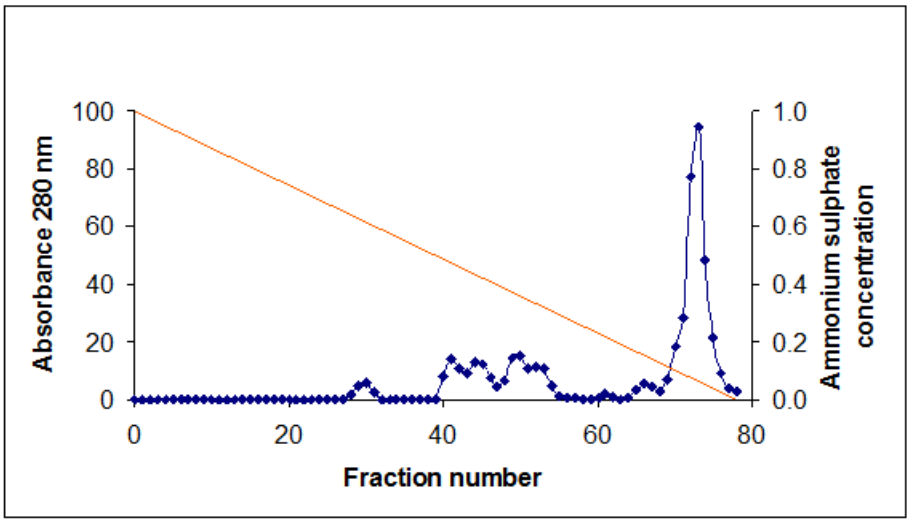


Figure 4.19: Hydrophobic interaction chromatography of cell free extract containing NHase.

Elution of protein (blue trace) was done using a 1 - 0 M ammonium sulphate gradient (red line).

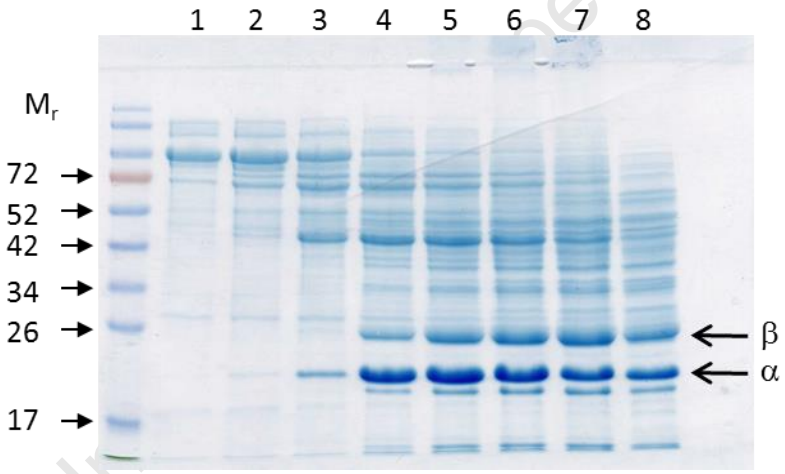


Figure 4.20: Ion exchange elution fractions containing NHase alpha and beta subunits at different concentrations.

Fractions containing NHase eluted from a Q-Sepharose ion exchange column. The concentration of alpha subunit is noted to be more than the beta subunit.

Since ammonium sulphate precipitation steps de-coupled the beta and alpha NHase subunits, further purifications avoided fractionation. Ion exchange chromatography of cell free extract containing NHase consistently eluted NHase between 320 and 375 mM NaCl, with peak fractions containing activity against benzonitrile at 350 mM NaCl (Figure 4.21).

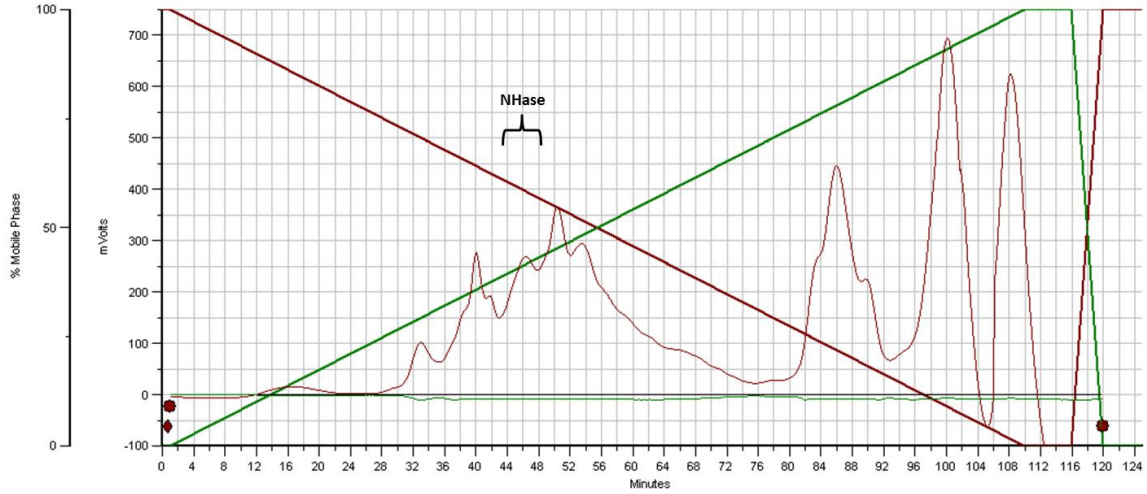


Figure 4.21: Elution by ion exchange chromatography of cell free extract containing nitrile hydratase.

NHase eluted at fractions 44 – 48 (corresponding to 44 – 48 minutes), at an KCl concentration of ~350 mM. Green and red diagonal lines represent increase and decrease in % mobile phase respectively.

Ion exchange fractions containing NHase were pooled and subjected to a further round of ion exchange chromatography with a reduced salt gradient (Figure 4.22).

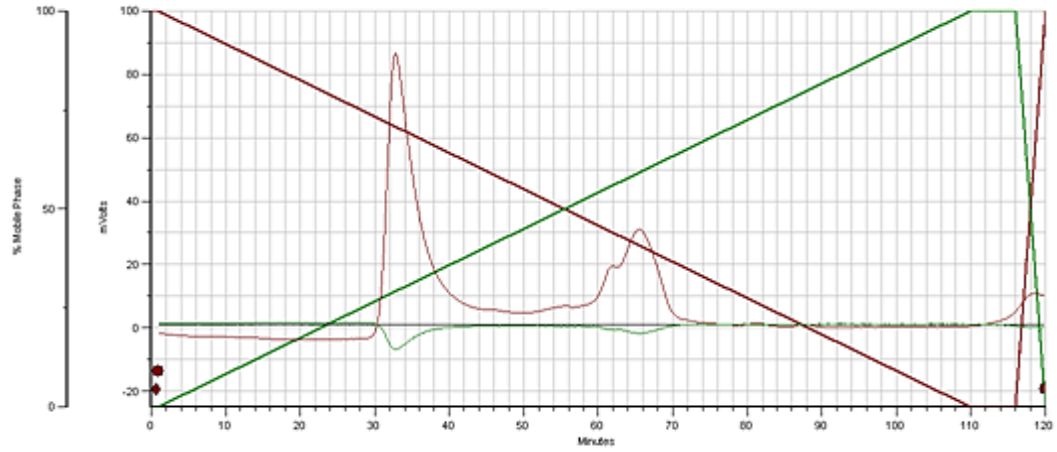


Figure 4.22: Elution by ion exchange chromatography using a 100-400 mM KCl gradient of pooled fractions containing nitrile hydratase eluted at ~350 mM.

Fractions 44 – 48 eluted by ion exchange chromatography at an NaCl concentration of ~350 mM, and with activity against benzonitrile, were pooled and concentrated, and re-applied to ion exchange column. NHase was eluted with a 100-400 mM NaCl gradient. Green and red diagonal lines represent increase and decrease in % mobile phase respectively.

Fractions were tested for NHase activity by reaction against benzonitrile, using the spectrophotometric assay to measure benzamide (See Appendix 7.4.1.1). The assay was developed specifically to be able to test for NHase activity qualitatively and quickly, without relying on SDS-PAGE analysis, and using very small volumes. Fractions containing NHase activity were pooled and eluted by size exclusion chromatography (Figure 4.23). Size exclusion fractions were measured for NHase activity against benzonitrile by absorbance assay (Figure 4.24), and active fractions were pooled and further purified by FPLC size exclusion chromatography (Figure 4.25).

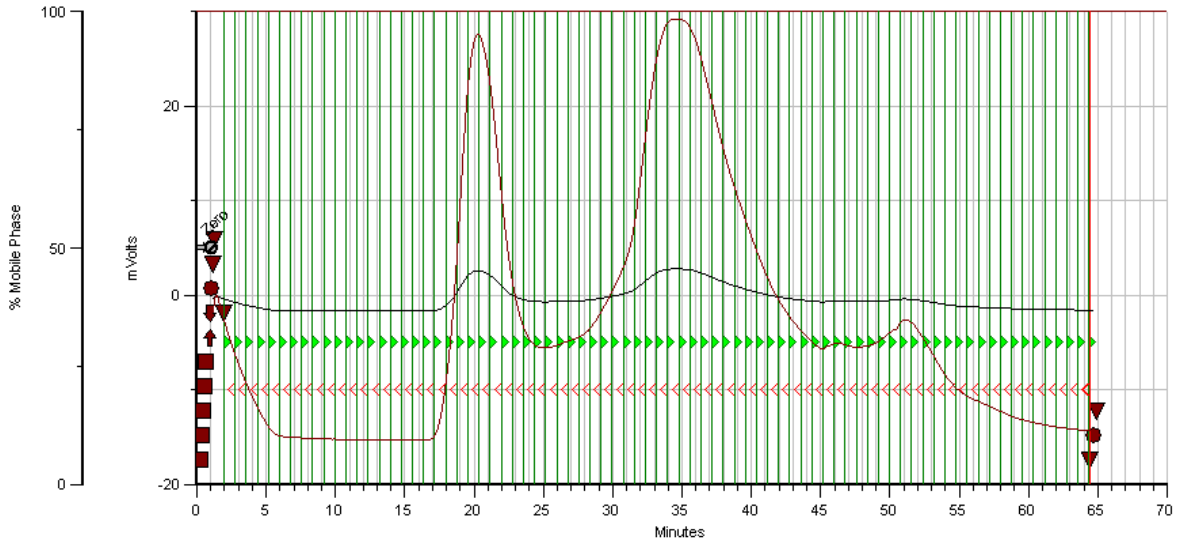


Figure 4.23: NHase elution by S200 Sephacryl size exclusion chromatography.

NHase elution (--) was monitored by absorbance at 280 nm, as well as absorbance assay for benzamide formation and SDS-PAGE. Fractions 30 – 40 (at 30 – 40 minutes elution time) corresponded with NHase activity.

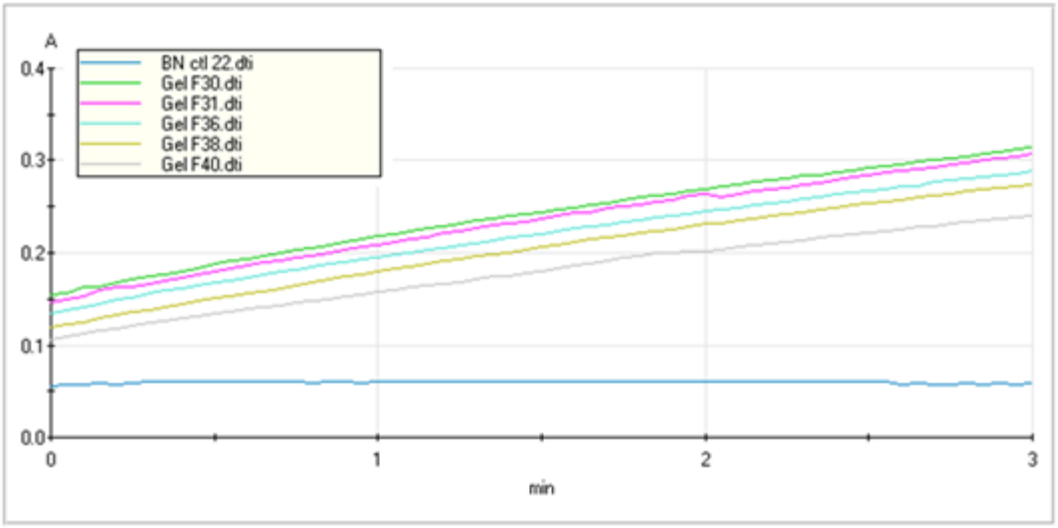


Figure 4.24: NHase activity measurement in size exclusion fractions using the absorbance assay for benzamide formation.

Gel filtration fractions containing NHase were tested for benzonitrile conversion by measuring the formation of benzamide spectrophotometrically (diluted samples) at 252 nm. The rate of benzamide formation was recorded over 3 minutes from addition of fraction containing NHase. If the rate of benzamide formation was not linear, samples were diluted appropriately. The blue line with zero gradient is control benzonitrile, while the coloured lines are fractions containing NHase.

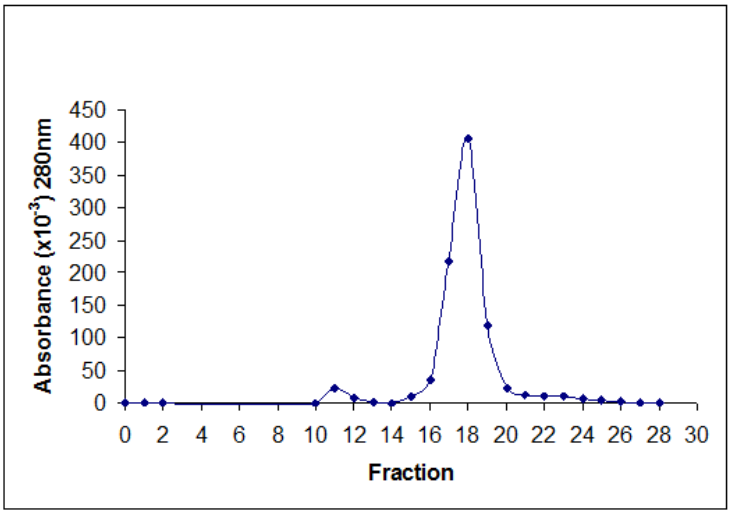


Figure 4.25: Nitrile hydratase elution by TSK FPLC size exclusion chromatography.

Fractions eluting from S200 size exclusion chromatography were pooled, concentrated, and a 0.5 mL aliquot added to a TSK4000 FPLC size exclusion chromatography column, equilibrated and run with 50 mM potassium phosphate buffer pH 7.4 containing 100 mM NaCl. NHase elution at 280 nm (--) was monitored by absorbance assay for benzamide formation.

4.3.4. Secondary structural prediction and molecular modelling of nitrile hydratase from *R. rhodochrous* ATCC BAA-870

The predicted secondary structure of the alpha subunit of L-NHase showed the active site residues ("CTLCSC", position 110) are in a flexible region of the protein. The alpha subunit consists mainly of alpha helices, with very few, short beta sheets (Figure 4.26).

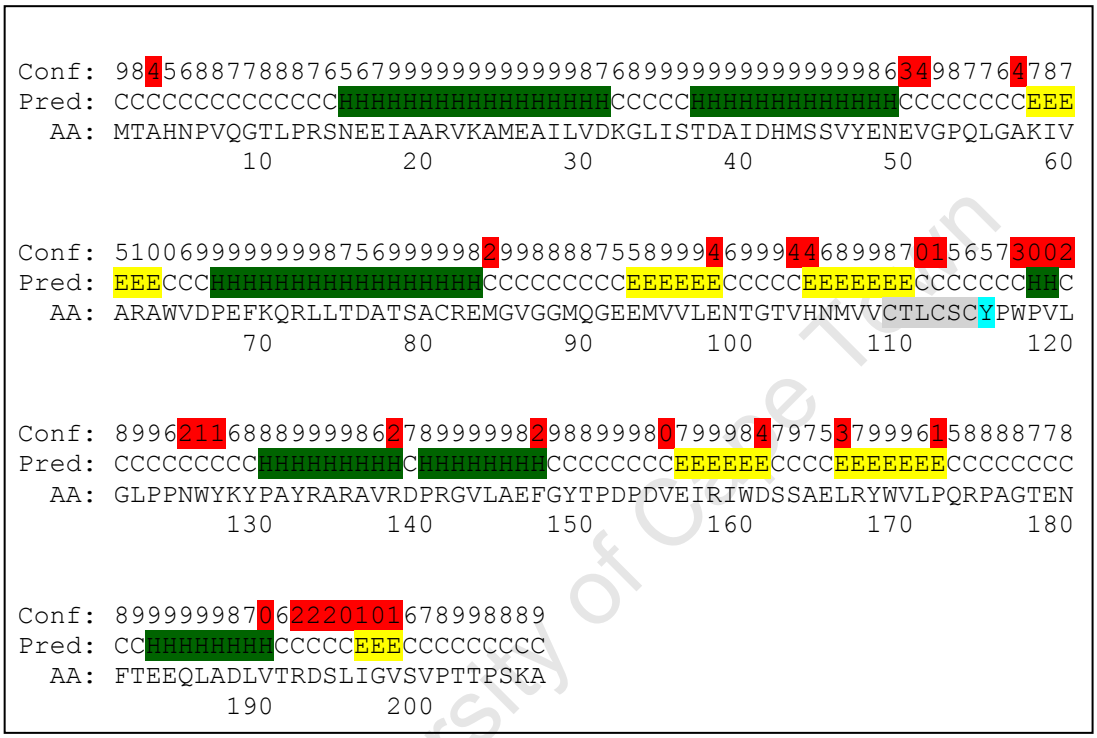


Figure 4.26: PSIPRED prediction results for the alpha subunit of L-NHase.

Secondary structure prediction was done using PSIPRED V2.6 by Jones *et al.* 2000 (Jones *et al.* 2000). The confidence of structure prediction for each amino acid (AA) is given as a score of 0 (low), highlighted in red, to 9 (high). Secondary structure is predicted as alpha helix (H) in green, beta strand (E) in yellow, or random coil (C). The active site conserved residues, "CTLCSC" on the alpha subunit, are highlighted in grey. The BAA-870 active site Tyrosine (αTyr115) is highlighted in blue.

4.3.4.1. Molecular modeling of nitrile hydratase from *R. rhodochrous* ATCC BAA-870

The nitrile hydratase alpha subunit from BAA-870 showed highest percentage identity at the amino acid level (58.9%) to the alpha subunit of the *B. pallidus* NHase (PDB code 3HHT) (Van Wyk *et al.* 2009), followed by 57.7%, 48.7% and 40.1% identity to the alpha subunits of

1UGP, 2QDY and 2ZZD respectively (Arakawa *et al.* 2009; Miyanaga *et al.* 2004; Song *et al.* 2007). Structural alignment with nitrile hydratase from *B. pallidus* showed only a 37% identity with the beta subunit. Based on the closest structural alignments for molecular modelling, the structure of the *B. pallidus* NHase alpha subunit, PDB code 3HHT (Van Wyk *et al.* 2009) and Co-type NHase beta subunit, PDB code 1UGP (Miyanaga *et al.* 2004) were used to generate the structural alignments. The modelled BAA-870 NHase alpha subunit structurally aligned with the alpha subunit from *B. pallidus* RAPc8 NHase revealed the alpha subunits are highly similar, with only a small N-terminus extension for the alpha subunit of NHase from BAA-870.

There is also no visible or obvious difference in the positioning of the active site cysteine residues or the cobalt ion in BAA-870 NHase and RAPc8 NHase. Similarly, a structural alignment of the beta subunits showed a highly conserved structure. Structurally, the NHase from BAA-870 appears to be almost fully conserved (Figure 4.27).

The BAA-870 NHase alpha subunit has a conserved active site Tyrosine (αTyr115) (Figure 4.26). As observed for other Co-type NHases, the conserved active site sequence is 'C**T**LCSC**Y**'. Three conserved residues that putatively determine the substrate specificity of *P. thermophila* NHase (βLeu48, βPhe51 and βTrp72) are also conserved in L-NHase from *R. rhodochrous* BAA-870 (which are identical to that in J1 L-NHase). The kinetic parameters for similar sized substrates were found to be similar for *P. thermophila* and *R. rhodochrous* J1 (Wieser *et al.* 1998).

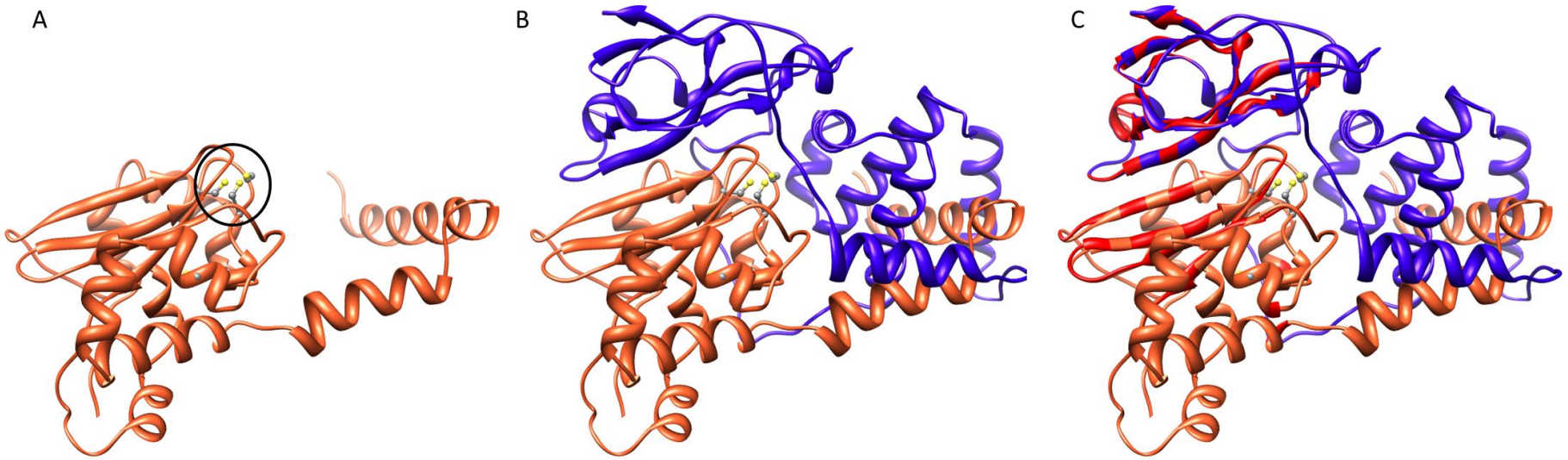


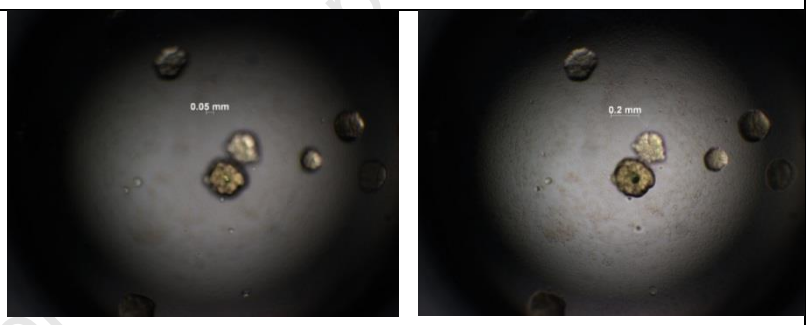
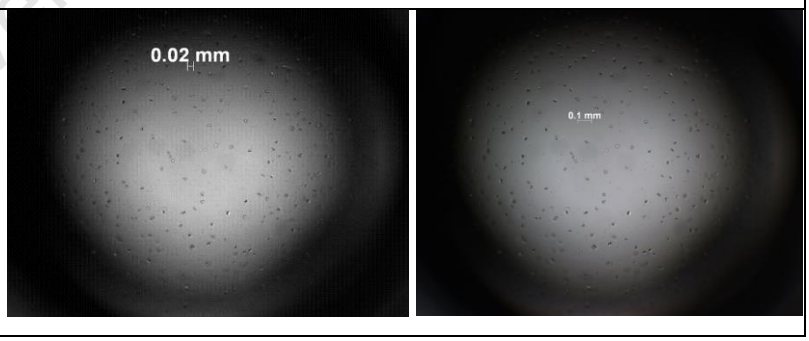
Figure 4.27: Molecular model of the predicted protein secondary structure of BAA-870 nitrile hydratase, and an alignment with its closest structural homologues.

The BAA-870 NHase molecular model was generated using its closest structural homologues: the alpha subunit of *Bacillus pallidus* NHase (PDB code 3HHT) (Van Wyk *et al.* 2009), and beta subunit of *Pseudonocardia thermophila* Co-type NHase (PDB code 1UGP) (Miyanaga *et al.* 2004). The alpha and beta subunit is depicted in orange and blue respectively, while only the structural alignment differences between the homologues are shown in red. Yellow and grey sticks show the active site cysteines of the alpha subunit. A: alpha subunit structural model of BAA-870 NHase showing the active site cysteines (circled). B: the full length NHase structural model of BAA-870. C: The full length modeled NHase of BAA-870 showing the structural alignment with its closest structural homologues. Figures were generated using UCSF Chimera version 1.5.3 (Pettersen *et al.* 2004).

4.3.4.2. Attempted crystallography of NHase

The most promising crystal hits were obtained in 0.2 M sodium acetate-trihydrate, 0.1 M Tris-HCl buffer, pH 8.5, containing 30% (w/v) PEG-4000 and 0.1 M Tris-HCl, pH 8.5, containing 8% (w/v) PEG-8000 (Table 4.4). However, in both cases, the crystals were too small (less than 0.5 mm) to be used for diffraction. The instability of purified NHase prevented availability of higher concentrations of protein for further crystal trials and optimisation. The NHase could be purified to ~1.8 mg/ml, but on further concentration for crystallography conditions, the protein precipitated.

Table 4.4: Crystallization conditions used for the most promising nitrile hydratase crystal hits

Condition	Images
0.2 M sodium acetate-trihydrate, 0.1 M Tris-HCl, pH 8.5 containing 30% (w/v) PEG-4000	
0.1 M Tris-HCl, pH 8.5 containing 8% (w/v) PEG-8000	

4.4. Discussion and conclusions

4.4.1. Cloning problems; instability of the subunit ratio during purifications; codon usage and proposed gene difficulties

Expression studies of the L-NHase showed that the beta subunit and the L-NHase chaperone were overexpressed, but not the alpha subunit. The L-NHase-containing clone was subsequently mutated at the ribosome binding site for the alpha subunit, as it was found to be a particularly rare site (RBS), likely unrecognised by *E. coli* ribosomes. L-NHase expression studies were performed using *E. coli* Single-Step and JM109(DE3) expression cells. Induction of L-NHase was done using IPTG for both strains, with rhamnose additionally used to induce the single-step expression cells.

Sequencing of the nitrile hydratase operon revealed that the gene order is the same as that of *R. rhodochrous* J1. The order of the gene cassette is beta subunit, alpha subunit, and then chaperone protein (the equivalent of p44k in *Geobacillus pallidus*). The alpha subunit of nitrile hydratase cloned from *R. rhodochrous* ATCC BAA-870 was aligned with that from *R. rhodochrous* J1, and the alpha subunits containing the active site are identical. There is less sequence conservation between the beta subunits, although they are still highly similar (97%).

4.4.2. Expression of nitrile hydratase

Nitrile hydratase from *Geobacillus pallidus* Rapc8 was used as an expression control during expression troubleshooting, as the expression conditions for this protein are known and optimised. The ribosome binding sites (RBS) of *R. rhodochrous* J1 and *Geobacillus pallidus* were compared. The RBS for the alpha subunit (Table 4.5) was shown to be a non-standard binding site for protein expression initiation, while the RBS for the beta subunit was easily recognizable for *E. coli* ribosomes. Expression of the construct in the pET system plasmids with directional cloning, however, means that the RBS for the beta subunit is under the control of the plasmid promoter, and therefore would not pose a problem for *E. coli* ribosomes. It is noted that the chaperone RBS, while theoretically recognised by *E. coli* ribosomes, may cause stalling in the translation of the protein since it is not a 'traditional' RBS.

There is evidence that rhodococcal genes may not express efficiently in *E. coli* cells (Denome *et al.* 1994), or that *Rhodococcus* proteins are inactive in *E. coli* cells (Denome *et al.* 1994;

Ikehata *et al.* 1989). There are, however, several examples of rhodococcal proteins expressed in *E. coli* host cells. These include the expression of enantioselective NHase from *R. erythropolis* AJ270 (alpha and beta subunits together with the chaperone protein, p44k), nitrilase from *R. rhodochrous* tgI-A6, the 4-gene operon of alkene monooxygenase from *R. rhodochrous* B-276 (formerly *Nocardia corallina*) and cytochrome P450 from *R. ruber* DSM 44319. In addition, the expression of the 3-gene operon of nitrile hydratase from other species has been successful. The beta homolog protein in *G. pallidus* is also naturally not in the same frame as the beta and alpha subunits, and this does not cause any protein expression problems (Tsekoa *et al.* 2004).

The order in which the alpha and beta subunits, and the chaperone, appear on the genome differs amongst organisms. In *R. rhodochrous* J1, as with ATCC BAA870, the beta subunit is transcribed before the alpha subunit, which is in turn before the chaperone protein (Figure 4.28). The alpha and beta subunits are in the same translation frame, whereas the chaperone protein is in a different frame to the alpha and beta subunit. The chaperone protein which is transcribed with the alpha and beta subunits of nitrile hydratase bears no similarity to any of the heat shock protein families, but rather shares the nitrile hydratase beta superfamily fold. It is therefore also referred to as the nitrile hydratase beta subunit homolog protein.

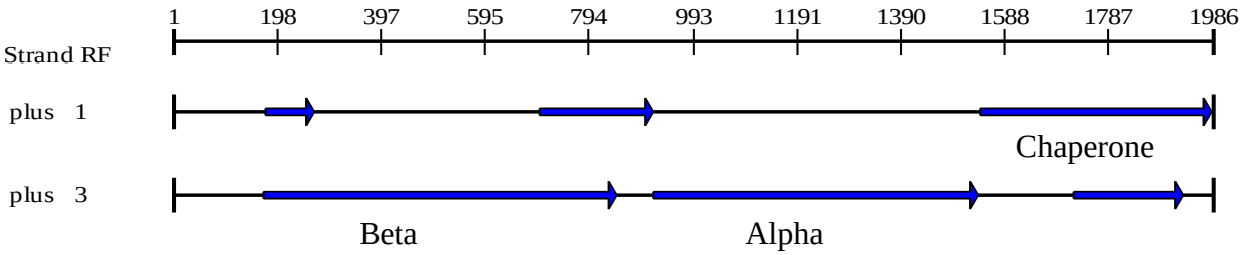


Figure 4.28: Translation overview of low molecular weight nitrile hydratase from *R. rhodochrous* J1 and BAA-870.

The gene cassette codes for the beta and alpha subunits in the same frame (plus 3 frame), with the chaperone (beta homolog protein) in a different frame (plus 1 frame). The NHase beta and alpha subunits, and chaperone protein were sequenced using the forward and reverse primers designed for amplifying the region from genomic DNA of *R. rhodochrous* ATCC BAA-870. The forward and reverse sequences were assembled using the CAP3 sequence assembly programme, and the translation overview created using DNAMan (Lynnon Corporation, USA).

The alpha and beta subunits have different ribosome binding sites, and the initiation sequence for chaperone translation falls within the N-terminus of the alpha subunit sequence. A consensus ribosome binding site sequence, or Shine-Dalgarno sequence in Prokaryotes, is usually followed by a spacer sequence of varying length. While the ribosome binding sites in *R. rhodochrous* are recognised sites for *E. coli* ribosomes, they are uncommon (Table 4.5).

Table 4.5: Examples of various Shine-Dalgarno sequences recognised by *E. coli* ribosomes

araB	-UUUGGAUGGAGUGAAACGAUGGCGAUU-
galE	-AGCCUAAUGGAGCGAAUUAUGAGAGUU-
lacI	-CAAUUCAGGGUGGUGAUUGUGAAACCA-
lacZ	-UUCACACAGGAAACAGCUAUGACCAUG-
Q β phage replicase	-UAACUAAGGAUGAAAUGCAUGUCUAAG-
ϕ X174 phage A protein	-AAUCUUGGAGGCUUUUUUAUGGUUCGU-
R17 phage coat protein	-UCAACCGGGGUUUGAAGCAUGGCUUCU-
Ribosomal protein S12	-AAAACCAGGAGCUAUUUAAUGGCAACA-
Ribosomal protein L10	-CUACCAGGAGCAAAGCUAAUGGCUUUA-
trpE	-CAAAAUUAGAGAAUAACAAUGCAAACA-
trpL leader	-GUAAAAAGGGUAUCGACAAUGAAAGCA-
3'-end of 16S rRNA	3' $_{HO}$AUUCCUCCACUAG – 5'

The initiation AUG codons (brown) lie approximately 10 nucleotides downstream of their respective recognition sequences (blue). The Shine-Dalgarno sequences are complementary to the UCCU core sequence element of *E. coli* 16S rRNA. Adapted from The Mechanics of Protein Synthesis (Figure 32.9, page 1049).

In order for translation initiation to occur, a translation initiation codon is required, in addition to a ribosomal binding site which includes the Shine-Dalgarno (SD) sequence. Translation is initiated from the translation initiation region (TIR) of the transcribed messenger RNA. The SD sequence is located 7 ± 2 nucleotides upstream of the initiation codon (AUG in most recombinant systems). Optimal translation initiation is obtained from mRNA with an SD sequence of UAAGGAGG. Translation initiation is largely dependent on the secondary structure of the RBS, and it was shown that efficiency of initiation can be improved by high adenine and thymine contents (Laursen *et al.* 2002). In particular, the translation initiation efficiency is influenced by the codon immediately following the initiation codon; in highly expressed genes it was shown that adenine content was high (Stenstrom *et al.* 2001).

The binding sites for *Geobacillus pallidus* NHase (Table 4.6) are almost identical to many of the ribosome binding sites recognised by *E. coli* ribosomes (Table 4.5), which may explain why expression in *E. coli* of *G. pallidus* proteins has not been problematic.

Table 4.6: Ribosome binding sites of *R. rhodochrous*, and the expression control organism, *G. pallidus*

	Ribosome binding site sequence		
	Alpha subunit	Beta subunit	Chaperone
G. pallidus	GAAGGAGGAA	GGGAGG	AGGAGGAAAA
R. rhodochrous J1	GAAACGAGCC	AAGGAG	AAGGCCTGAC

Translation can also result in errors when genes containing rare codons are expressed, since it cause stalling of ribosomes at positions where amino acids coupled to minor codon tRNAs are incorporated (McNulty *et al.* 2003). This codon bias is particularly problematic when rare codons occur in clusters and in doublets and triplets. Rare codon bias leads to translational errors such as premature translational termination, frame shifting events and mistranslation amino acid substitutions (Kurland and Gallant 1996; Sørensen *et al.* 2003). Kane *et al.* (1992) reported in-frame two amino acid 'hops' at the disfavoured AGA codon (Kane *et al.* 1992). The most problematic codons are those decoded by products of the genes argU (AGA and AGG), argX (CGG), argW (CGA and CGG), ileX (AUA), glyT (GGA), leuW (CUA), proL (CCC) and lys (AAG). In fact, AAG is a major *E. coli* codon and is decoded by $tRNA^{Lys}UUU$ which has a fair degree of 'wobble' due to its structure – therefore UUU reads AAG less efficiently than most, and efficiency of decoding by the tRNA is particularly problematic when there are consecutive AAG codons (Sørensen and Mortensen 2005). Despite the strategy of site-directed mutagenesis for eliminating codon bias, codon-optimized genes have been shown to lack mRNA transcription and stability in recombinant expression systems (Wu *et al.* 2004).

The codon frequency usage of *E. coli* and *R. rhodochrous* were compared. *E. coli* is accepted as the genetic code standard, and most expression strains are *E. coli* cells. It was proposed that perhaps the codon usage in *Rhodococcus* differed enough with *E. coli* so as to cause expression problems by becoming rate-limiting. Stretches of four or more rare codons can reduce expression significantly (Robinson *et al.* 1994) by causing ribosome stalling, which can uncouple translation and transcription thereby leading to premature termination of translation. In common *E. coli* expression strains, the 3 rarest expression codons are specifically those coding for arginine (AGA/ AGG), isoleucine (AUA) and leucine (CUA), hence the development of *E. coli* BL21 codon plus RIL strain by Stratagene, where the copy numbers of t-RNA's for arginine, isoleucine and leucine are increased. This adaptation of

codon usage is useful in the expression of high GC content templates. However, it was found that the frequency codon usage for *Rhodococcus* was quite similar to *E. coli*, and the ratios of these particularly rare codons frequencies to the codons coding for the same amino acid, were almost identical in both strains. *E. coli* has a slight bias towards preferentially using codons that finish in a U/T whereas *Rhodococcus* seems to prefer a G/C as the last codon. Problematic expression of rhodococcal proteins in *E. coli* may be attributed to the GC-rich nature of the rhodococcal genome.

For *R. rhodochrous* nitrile hydratase expression, a few expression cells were tested, including BL21 with a T7 promoter (OverExpress C41(DE3)pLysS) which is resistant to toxic protein expression. Parameters tested included temperature ranges, IPTG concentration ranges, and cobalt concentration ranges as well as addition at various points to optimize expression. Lack of protein expression was possible due to rare ribosome binding sites and/or rare codon usage in the *Rhodococcus* codon that are not recognised or commonly used by *E. coli*. Both *E. coli* and *Rhodococcus* use the bacterial translation table and supposedly have similar codon usage. Rare-codon usage is not usually a rate limiting problem during translation but if 4 or more happen to occur in a row, it can cause ribosomal pausing and subsequently uncouples transcription from translation. This can be solved by altering the codons to those used more favourably during high expression in *E. coli*. Without doing mutagenesis to change codons, it can be solved by modifying the expression codon for cloning into a pETBlue 2 vector and using it in RosettaBlue (DE3)pLacI competent *E. coli* strain for example (which is used for expressing codons which are otherwise rare in *E. coli*). Each subunit in our construct has its own ribosome binding site, the presence of which was confirmed by sequencing.

4.4.2.1. Rationale and success of the two plasmid cloning system

Nitrile hydratase coexpression cloning

There are a number of important factors to consider when selecting a two-plasmid vector system for expression in *E. coli*, including vector compatibility, vector copy number and antibiotic selectivity.

The copy number of each plasmid may differ, either due to the randomness of drawn plasmids into replication from the cellular pool, or the size of the plasmid (larger plasmids will take longer to replicate, and will be at a selection disadvantage in every cell population). Plasmids of similar size may not necessarily guarantee constant copy numbers since the efficiency of initiation is not necessarily balanced in individual cells. Imbalances in plasmid replication rates may lead to descendants of the cell containing either plasmid but very rarely both in the absence of selection. A plasmid's replicon contains the origin of replication and in some cases associated cis acting elements (del Solar *et al.* 1998). Most plasmids used in recombinant protein expression have a ColEi or p15A replicon. Plasmid copy number is controlled by the replicon, and it preferably replicates in a relaxed fashion (Baneyx 1999).

To further increase the maintenance of two different plasmids in the same cell, the antibiotic resistance of the two plasmids should be different. Different antibiotic resistance will ensure that both plasmids are selected for and maintained in daughter cells. The multicopy plasmids are stably replicated and maintained under selective conditions, and therefore plasmid-free daughter cells are rare (Summers 1998).

Considering all the requirements for protein co-expression, two plasmids, pRSFDuet-1 and pET-21a(+), were selected for directional cloning of the alpha-chaperone construct and beta subunit, respectively (Table 4.7). The efficiency of transcription, frequency of mRNA translation, and mRNA stability will determine the gene expression levels. Expression plasmids require a string transcriptional promoter to control high-level expression of the cloned gene. Basal transcription in the absence of inducer is minimized by the presence of a suitable repressor.

The T7 promoter is a 20-nucleotide sequence not recognized by *E. coli* RNA polymerase, and it transcribes genes five times faster than *E. coli* RNA polymerase (Sørensen and Mortensen 2005). Plasmid stability can be enhanced by preventing transcription through the origin of replication and from irrelevant promoters elsewhere in the plasmid, hence the presence of a transcription terminator downstream from the coding gene sequence. Transcription terminators stabilise the mRNA by forming a stem loop at the three prime end (Newbury *et al.* 1987).

For co-expression of multiple gene products from multiple plasmids in the same cell, different antibiotic resistance markers are required in addition to different plasmid incompatibility groups. For example, plasmids containing the ColE1 replicon and p15A replicon can be combined since they are compatible plasmids (Mayer 1995). The ColE1 replicon is derived from the pBR322 family of plasmids (copy number ~15-20) or pUC family of plasmids (copy number ~ 500-700), while the p15A replicon is derived from pACYC184 (copy number ~10-12).

Table 4.7: Plasmid characteristics for expression of beta and alpha-chaperone in a two plasmid system

Plasmid name	**pRSFDuet-1**	**pET-21a(+)**
Expression	Alpha subunit and chaperone	Beta subunit
Size	3829 bp	5443 bp
Promoter	T7lac	T7lac
Replicon	RSF 1030 (NTP1)	pBR322, derived from ColEI
Resistance	Kanamycin	Ampicillin
Copy number	>100	~ 40
Host strain	BL21(DE3) or BL21(DE3)pLysS	BL21(DE3)
Induction	IPTG	IPTG
Tag addition	None	C-terminus
Sequencing primers	DuetUP2 primer; T7 terminator	T7 promoter; T7 terminator
Insert size (nucleotide bases)	618 and 444 total: ~1062	675
Expressed protein size	206 amino acids (22.7 KDa) 148 amino acids (16.9 KDa)	225 amino acids (25.1 KDa)

The copy number of the pRSTDuet-1 vector is higher than that of pET-21a(+), but will be offset by the longer open reading frame expression for alpha and chaperone expression as opposed to beta subunit. Higher expression of either protein will not lead to increased expression overall since only a 1:1 ratio of each subunit will be purified. The C-terminal His-tag attached to the beta subunit will be used to purify the native NHase protein.

Formation of inclusion bodies

The structure of inclusion bodies and their mechanism of formation are poorly understood (Villaverde and Carrio 2003), and are largely assumed to be as a result of a stress response in the cell during high recombinant protein expression rates. It is yet to be determined whether inclusion bodies form as a result of unknown clustering mechanisms or due to hydrophobic interactions between unfolded proteins (Villaverde and Carrio 2003), but either way they result from an unbalance equilibrium between protein aggregation and solubilisation. The lowering of temperature and expression rate for example can reduce aggregation of expressed proteins, and so too can the co-expression of recombinant chaperones (Jonasson *et al.* 2002). Van den Berg *et al.* (1999) showed that macromolecular crowding in the cell cytoplasm (at protein concentrations of 200-300 mg/ml) could lead to unfavourable protein folding conditions when recombinant proteins are expressed at high rates (van den Berg *et al.* 1999).

There are several advantages to co-expression of two proteins in the same cell; versus expression in two separate systems with post-expression reconstitution of protein partners, in this case the reconstitution of the alpha and beta subunits. Many examples show the importance of the protein interaction partner during proper *in vivo* protein folding, and indeed co-expression showed increased amounts of properly folded proteins as well as decreased proteolytic degradation of the complexed protein (Li *et al.* 1997; Stebbins *et al.* 1999; Tan 2001).

Crystallization classification by XtalPred

The XtalPred sequence analysis tool (Slabinski *et al.* 2007) compares nine of the protein's biochemical and biophysical features with corresponding probability distributions obtained from TargetDB. A plot is automatically generated for each protein feature and reveals distributions of failures and successes in the learning sets extracted from TargetDB; interpolated empirical distributions of crystallization probability, and the positions of the protein in those distributions. The prediction is made by combining individual crystallization probabilities into a single crystallization score. Based on the score, the protein is assigned to one of five crystallization classes: optimal, suboptimal, average, difficult and very difficult.

Each class represents different crystallization success rate observed in TargetDB. The server shows the distribution of the crystallization probability in the crystallization classes and the position of the protein within this distribution.

Full length L-NHase from BAA-870 is classified as a difficult crystallography target (class 4) according to the XtalPred crystal prediction tool (Slabinski *et al.* 2007). Percentage coiled structure (50%) and instability index (relatively unstable) obtained from XtalPred are in agreement with results obtained from a ProtParam tool calculation (data not shown) (Gasteiger *et al.* 2005). The XtalPred classification is based on 41 protein homologs clustered to a 60% sequence identity and 34 homologs contained in the PDB.

Purification and stability of NHase

Bioinformatic analysis using the ProtParam tool showed that the beta subunit of NHase, by amino acid sequence analysis, has an apparent instability index of 43.5 (considered unstable), while the alpha subunit sequence is 35.8 (stable). The instability index, defined as the relative volume occupied by aliphatic side chains, gives a good estimation of the stability of the protein in vitro. A protein with an I.I. of more than 40 predicts that the protein may be unstable, whereas a value of less than 40 is predicted as stable. This is in agreement with the experimental findings, where the beta subunit was easily lost during purification steps. Also, the NHase is considered relatively hydrophobic, as deduced form the elution of the protein by hydrophobic interaction chromatography, a property which could encourage the precipitation of NHase upon purification and concentration.

Addition of a His-tag at the C-terminus of the NHase beta subunit increased the yield of NHase during purification, and decreased the de-coupling of subunits, possibly by stabilization of the beta subunit when attached to the resin. The purification protocol negated the use of ammonium sulphate fractionation or precipitation which was found to cause precipitation of the beta subunit. The construction of separate alpha-chaperone and beta expression vectors has several advantages: enhanced expression of the alpha and beta subunits by using the vector promoters recognised by *E. coli*, control of the subunit expression ratios by altering concentrations of vector in *E. coli*, and the possibility to trap the

chaperone-alpha association to elucidate the role of the chaperone. Incorporation of cobalt ion into L-NHase has been found to depend on the α-subunit exchange between cobalt-free L-NHase and its cobalt-containing mediator (Zhou *et al.* 2010). The interaction of the L-NHase chaperone with the NHase is presumed to be a novel mode of post-translational maturation, with the chaperone acting as a "self-subunit swapping" chaperone, but also perhaps a metallochaperone, as deduced by studies of the H-NHase interaction with the associated chaperone (Zhou *et al.* 2010).

It should be noted that differences in the ratio of the alpha and beta subunits observed by SDS-PAGE did not necessarily imply that the enzyme was inactive. A mismatched density of the alpha and beta subunits are a common feature of NHase analysed by SDS-PAGE and has been found by several other groups (Cramp and Cowan 1999; Wieser *et al.* 1998; Yu *et al.* 2006) although the subunits are very similar in size (23-26 kDa).

Suggestions for improved enzyme mechanism

Recent developments in the technologies used to improve enzymes, such as directed evolution techniques, have greatly influenced the use of nitrile metabolizing enzymes in applied biocatalysis. One such example is that of the engineered papain nitrile hydratase for use as a biocatalyst by Dufour and co-workers (Dufour *et al.* 1995). Papain, a cysteine protease, was successfully converted into an enzyme with nitrile hydratase activity through protein engineering. This was achieved by replacing Gln19 in the oxyanion hole of papain with a glutamic acid residue. The nitrile substrate forms a covalent thiomidate adduct with the active site thiol group of the enzyme. The altered Glu19 residue participates in the acid-catalysed hydrolysis of the thiomidate to an amide by providing a proton to form a more reactive protonated thiomidate. The hydrolysis of nitrile substrates therefore proceeds in a stepwise manner. First, the nitrile is hydrolysed to the corresponding amide. Since the Q19E mutant protein retains most of the natural amidase activity of natural papain, the amide is then further hydrolysed in the second step to the corresponding carboxylic acid. This mutant enzyme is considered a nitrilase in overall function, since it possesses both nitrile hydratase and amidase activities. The mutant papain is stable, active in organic solvents and at low pH, and is specific for peptide-based substrates. To illustrate the uses of this engineered protein,

the mutant enzyme has since been used to synthesise amidrazones (that serve as intermediates in the synthesis of key components of biologically active molecules) with a more than 4000-fold increase in rate of production over the wild-type activity, and with a much higher level of purity (Dufour *et al.* 1998).

The serine/threonine side chain differences in the conserved active site of NHase, "C(**T**/**S**)LCSC(**Y**/**T**)", are proposed to play an important role (Payne *et al.* 1997) in the activity of the enzyme, however no mutational studies on these residues have confirmed this. In Co-type NHase, T and Y correspond to αThr109 and αTyr114, while in Fe-type NHase, S and T correspond to αSer109 and αThr114. Studies of Fe substitution in a cobalt enzyme have not been performed. Substitution mutations of these conserved residues would give interesting insights into the role of these residues in Fe and Co-type NHase, as well as elucidate the role of the Fe and Co ions at the active site.

In *P. thermophila*, the NHase αTyr114 residue is thought to be involved in the interaction with the nitrile hydratase activator protein (Miyanaga *et al.* 2004). Given the conserved active site sequence, a mutational substitution of the corresponding conserved tyrosine (αTyr115) in BAA-870 NHase which follows the alpha subunit active site consensus motif (CTLSCS) would be a good test for activator protein dependence. In other Co-type NHases, the active site Tyr -OH group forms an H-bond with main-chain oxygen atoms of αLeu119 and αLeu121 *via* a water molecule, also has hydrophobic interactions with its environment.

The *P. thermophila* NHase crystal structure in complex with the inhibitor *n*-butyric acid revealed that the hydroxyl group of another structurally relevant Tyr residue, βTyr68, forms hydrogen bonds with both *n*-butyric acid and αSer112 located in the active centre. Presumably, this intersubunit interaction may play a role in stability of the protein, and is another subject for a future mutational study.

A successful strategy for improving the activity of NHase was highlighted by the increasing the thermostability of the NHase from *G. pallidus* RAPc8 (Goldschmidt *et al.* 2007), where error-prone PCR was used to generate random mutations in the sequence of the gene. Several thermostable mutants were selected using this technique, and new electrostatic

interactions identified which played a role in enhanced thermostability, and allowed a strategy for further site-directed mutagenesis of NHase to form a highly superior and thermostable NHase. The structure of Co-containing NHase from *G. pallidus* has also given insight into the determinants of the specificity of the enzyme (Zhou *et al.* 2010). The authors suggest that αPhe55 obstructs access of larger substrates to the active site cysteine claw complex, allowing the hydration of only smaller aliphatic nitriles. The close structural similarity to the RAPc8 NHase offers some starting points for site-directed mutagenesis of BAA-870 NHase for increased thermostability, and possibly the introduction of mutations which could allow better transformations of aliphatic nitriles.

A very recent study has presented a greatly improved NHase stability by introduction of mutations to induce salt-bridging in thermally sensitive regions; regions were identified through RMSF calculations of the thermophilic NHases from a *Bacillus* (1V29) and *Pseudonocardia* (1UGQ) strain (Chen *et al.* 2013). Three thermally sensitive areas were identified in the beta subunit. Stabilizing salt bridge interactions were transferred to the corresponding region of the *Rhodococcus* ruber TH NHase was identified as an industrially useful acrylamide producing biocatalyst (Ma *et al.* 2010), and the researchers wished to improve the NHase activity. Interestingly, one mutated thermally sensitive area caused an enhanced expression of the β-subunit, but a complete loss of activity. However, one of the area mutations resulted in greatly enhanced tolerance of the NHase to sonication during cell disruption, as well as increased product tolerance and a 160% increase in thermal stability. This enhanced stable NHase was revealed through dynamic simulation to have created 10 new internal and inter-subunit salt bridges. Supporting the finding herein, they concluded that C-terminal salt bridge modification was a powerful strategy for enhancing thermal stability of NHase. Addition of a C-terminal His-tag to the BAA-870 NHase allowed better expression and purification of the enzyme. Although the J1 and BAA-870 NHase alpha subunits are identical, the beta subunits differ. It is likely that the source of the relatively unstable BAA-870 NHase lies in the beta subunit.

4.5. Conclusions

A summary of the methods employed to isolate, purify and structurally characterize the BAA-870 NHase is presented in Table 4.8. The nitrile hydratase (NHase) gene from *R. rhodochrous* was amplified from the genome of the organism using primers containing the cloning restriction sites *Nde*I and *Hin*dIII. The gene amplified contains the beta subunit, alpha subunit and chaperone, as confirmed by DNA sequencing. The NHase from BAA-870 is a low-molecular mass, Co-containing NHase. The PCR product was ligated into the pET28a(+) expression vector. Expression studies of the L-NHase showed that the beta subunit and the L-NHase chaperone were overexpressed, but not the alpha subunit.

The L-NHase-containing clone was additionally mutated at the ribosome binding site for the alpha subunit, as it was found to be a particularly rare site (RBS), likely unrecognised by *E. coli* ribosomes. The mutation was created using the Phusion mutagenesis kit, and the desired point mutation was obtained and verified by DNA sequencing. The mutagenesis reaction, however, incorporated undesired mutations at the start codon of the gene insert. The mutated insert was amplified using the original cloning primers containing the NdeI and HindIII cloning sites, and re-ligated into pJET1.2 for cloning and screening of the correct clones, before the correct insert was re-cloned into the expression vector (pET28a). Induction of L-NHase was done using cobalt chloride and IPTG for both strains, with rhamnose additionally used to induce the single-step expression cells.

To have further control over the expression of stable NHase, a two construct expression system was employed. The pET21a(+) construct expressing the NHase beta subunit with a C-terminal His-tag could be independently induced in *E. coli* BL21(DE3) from the pRSF-Duet construct expressing NHase alpha-chaperone. The pRSFDuet construct containing the NHase alpha chaperone and the pET21a(+) construct containing NHase beta subunit could be optimally co-induced by transforming the expression strain with a plasmid ratio of 4:1 respectively. Addition of a His-tag at the C-terminus of the NHase beta subunit increased the yield of NHase and decreased the de-coupling of the subunits during purification. It also

allowed for easier purification by use of nickel column chromatography (and subsequent avoidance of destabilizing hydrophobic interaction chromatography).

Interestingly, the BAA-870 NHase could be modelled using two different NHase structures; the closest structural homologues being the *B. pallidus* Rapc8 NHase alpha subunit (Van Wyk *et al.* 2009) and *P. thermophila* NHase beta subunit (Miyanaga *et al.* 2001). Also, despite the intensive research done on NHase, there is still no rhodococcal cobalt-type NHase crystal structure determined. It was confirmed in this study that the enzyme is a difficult crystallography target, and indeed even crystallization prediction programs have suggested that the protein is an unlikely crystallization target. The protein instability index suggests the protein (especially the beta subunit) is inherently unstable and indeed this was noted in the difficulties encountered during purification steps. Many groups reported that NHase was less active as an isolated enzyme than when applied as a whole cell biocatalyst, lending support to the lack of reports on pure *Rhodococcus* cobalt-containing NHase as a whole. Clearly, the NHase requires some 'help' in the form of directed protein engineering if it is to be applied as a successful biocatalyst.

Despite conservation of the NHase active site sequence in the alpha subunit, the most significant sequence similarity exists between the beta subunits of both ferric and cobalt NHases, and also between the high and low molecular weight NHase beta subunits. Cramp *et al.* also found that the *Bacillus pallidus* Dac521 NHase beta shares significant sequence similarity with the beta subunit of several other NHases, but no significant similarity of the alpha subunit was found (Cramp and Cowan 1999). H-NHase is reported to be more stable than L-NHase, with the biggest different in sequence between the two lying in the beta subunits; the additional beta subunit sequence insertions appear to be responsible for the better stability of H-NHase overall. The beta subunit therefore appears to be critical in stabilizing the $\alpha\beta$ dimer of NHases in general.

Table 4.8. Summary of methods applied to isolate, purify and structurally characterize BAA-870 nitrile hydratase

	PURIFICATION	CLONING			MODELLING
Summary/ Brief description	**Wildtype L-NHase from fermentation**	**Cloned L-NHase : amplified by PCR from the genome**			**Using L-NHase sequence from the genome**
Method	Chapter 4.2.6 Ammonium sulfate precipitation Ion exchange Hydrophobic interaction Gel filtration Spectrophotometric assay for benzamide production	Strategy 1: Chapter 4.2.3 Beta-alpha-chaperone construct in pET20b(+) and pET28a(+)	Strategy 2: Chapter 4.2.4 Beta-alpha-chaperone construct in pET21 Alpha RBS mutated clone	Strategy 3: Chapter 4.2.5 Two plasmid construct. Beta subunit cloned into pET21a(+) and alpha-chaperone cloned into pRSFDuet-1	Chapter 4.2.9 Homology models built using MODELLER and secondary structural predictions from PSIPRED and the mGenThreader fold recognition tool.
Details	Combinations and/or different orders of the above purification methods were used.	Sequence as amplified from the genome. His-tag at N-terminus of beta.	Sequence amplified from the genome, with mutated RBS for the alpha subunit. His-tag at N-terminus of beta	Beta sequence and alpha-chaperone sequence amplified separately from the genome. Beta subunit under control of a plasmid promoter; alpha-chaperone under control of a plasmid promoter. His-tag at C-terminus of beta. No His-tag for the alpha chaperone construct	Insight into probable His-tag positioning problems
Problems	Stability problem, increasing precipitation with increased purity and concentration. Crystal trials gave micro crystals – problems to optimize due to poor protein concentration	Poor expression of alpha. Full protein was not expressed in *E. coli*; protein instability problem	Lost chaperone expression; protein instability	Individual expression of beta and alpha on separate plasmids	
Successes	Assay developed for spectral measurement of benzamide Crystal hit leads but with limited protein quantity for optimization	Expression of beta and chaperone	Some activity of NHase	Expression of both subunits individually and simultaneously in co-transformed *E. coli* Functional protein obtained with activity against benzonitrile and 3-cyanopyridine	Structural similarity to *B. pallidus* NHase identified

Comparative analysis has suggested that NHase has a poor chance of crystallizing (this study), and is also a relatively unstable protein, cited in several reports as having low thermal stability and noted in the reduced activity on application outside of the whole cell and increased stability on immobilization (Chen *et al.* 2013; Chiyanzu *et al.* 2010; Kopf *et al.* 1996; Liu *et al.* 2008; Nagasawa *et al.* 1986; Prasad and Bhalla 2010; Tucker *et al.* 2012). The BAA-870 NHase precipitated readily at high concentrations, and was easily subjected to subunit disruption in the presence of ammonium sulphate. However, the structural modelling has suggested that overall the BAA-870 NHase structure is conserved. The enzyme is therefore a strong candidate for mutation to enhance thermal stability. The successful random mutagenesis strategy (Van Wyk, 2009, not published) or directed C-terminal salt bridge modification strategy (Chen *et al.* 2013) employed by other researchers to enhance stability of NHase is highly recommended as future work.

Despite intensive research done on NHases with biocatalytic potential, the molecular details of the catalytic activity of NHases are not known comprehensively, and a *Rhodococcus* cobalt-containing NHase crystal structure is still lacking. Of the solved NHase crystal structures, none contains a substrate.

Industrial plants that currently utilize NHase for biocatalysis offer encouragement for the successful large-scale application green chemistry (as exampled by acrylamide production by Mitsubishi Rayon Company in Japan, SNF Floerger in France and Lonza Guangzhou Fine Chemicals in China). There is still a vacant but key role existing for an aromatic and stereoselective NHase in industry. The NHase from BAA-870 has extremely promising biocatalytic potential for the synthesis of peptidomimetic and pharmaceutical compounds given the activity found in previous research on the enzyme (Brady *et al.* 2006b; Chhiba *et al.* 2012; Kinfe *et al.* 2008; Kinfe *et al.* 2009). BAA-870 NHase is an ideal candidate for rational design to improve its stability and performance as an applied, pure biocatalyst. The current research presents, in addition to the NHase sequence information, various plasmid constructs for the induction and expression of the enzyme, as well as several starting points for further characterization (crystal lead hits, an easy absorbance assay for aromatic nitrile conversion and purification strategies), and has made possible better access to a promising NHase for further development.

5. Chapter 5

5.1. Overall Summary

5.1.1. Nitrile biocatalysis and a valuable biotechnological genome

Nitriles are important and versatile intermediates used in the organic synthesis of amines, amides, amidines, esters, carboxylic acids, aldehydes, ketones and heterocyclic compounds. Nitriles are particularly important sources (as precursors) of carboxyamides and carboxylic acids via hydrolysis. Biologically, we have the advantage of having two distinct enzyme systems, a nitrilase which converts nitriles directly to the carboxylic acid, and nitrile hydratase which provides the intermediate amide. By virtue of their ability to degrade a wide variety of nitrile compounds, nitrile-converting organisms have found extensive uses in the biotechnology industry, and applications of their activity have afforded useful industrial processes for large scale production of chemicals, drug intermediates and enantioselective targets.

A *Rhodococcus* organism was isolated from soil and tested for nitrile biocatalysis. The organism could convert several aromatic and aliphatic nitrile substrates, including benzonitrile, chlorobenzonitrile and naproxen nitrile, and also displayed fair activity towards phenylglycinonitrile, 1,2-cyanonitrile, 3-hydroxy-3-phenylpropionitrile and benzylnitrile. The organism showed excellent capacity to enantioselectively hydrolyze racemic compounds, as outlined by the biocatalytic resolution of β-hydroxy nitriles and synthesis of β-amino nitriles for example. The BAA-870 genome is rich with potential synthetic targets and biocatalytic tools. The sequencing of the genome of BAA-870 has expanded the inventory of putative protein sequences and enabled a 'reverse-genetics' approach to biocatalyst discovery, wherein prior knowledge of the gene itself expedites the screening process. Using bioinformatics methods, it was possible to identify classes of genes which can be cloned, expressed, purified and evaluated for their activity.

Rhodococcus rhodochrous ATCC BAA-870, identified as a promising candidate for biotechnology as both a whole cell biocatalyst as well as for its nitrilase and nitrile hydratase

enzymes, was subsequently studied further to evaluate its potential synthetic applications. The genome was sequenced in order to use the genome as a resource for gene fishing and specific gene identification. The optimal assembly and annotation of the BAA-870 genome was explored. The genome content, and specifically the nitrile converting genes, were summarized with respect to the protein complement and genome organization, and compared to other rhodococcal genomes. The information extracted from the genome of BAA-870, including potential targets of interest for future biotechnology applications are identified, and an outline of the benefits of having the genome sequence of *R. rhodochrous* ATCC BAA-870 were presented herein.

Solexa sequencing output data gave a total of 5,643,967 reads, 36-bp in length (4,273,289 unique sequences). The sequence coverage of the BAA-870 genome was 74%, with an apparent raw coverage depth of 27X, given a total genome size of 5.8 Mbp. The assembled genome sequence of *R. rhodochrous* ATCC BAA-870 was submitted to the Bacterial Annotation System web server, BASys, for automated, in-depth annotation of the chromosomal sequence. BAA-870 contains 6997 predicted protein-coding sequences (CDS). Of this, 0.9% encodes proteins of unknown function, and includes 318 conserved hypothetical proteins and 604 hypothetical proteins. Additionally, 49 proteins belong to known protein families, but their function is unknown. The average GC% of the coding sequence of BAA-870 is 67.46%. The complete biosynthetic pathways for all nucleotides, nucleosides and natural amino acids are also contained in the genome of BAA-870. Out of a total of 6874 genes annotated using BASys, 1342 genes were assigned an EC number.

The BAA-870 genome contains 501 oxidoreductases (37% of the enzymes). EC 1, oxidoreductases, make up the major enzyme class (37%), followed by EC 2 transferases (25%), and EC 3 hydrolases (17%). EC 4 and 6, lyases and ligases, both contribute 9% of the enzyme classes, while the EC 5 isomerases are the least at 3% of the total. The oxidoreductases and transferases together make up 62% of the enzyme class fraction. BAA-870 has 59 annotated oxygenases, but the EC classification suggests there are at least 114 oxygenases present within the genome. BAA-870 has one annotated monooxygenase but 46 annotated dioxygenases (59 oxygenases in total). It is unclear which oxygenases in BAA-870 are catabolic and which are involved in secondary metabolism, but their abundance is

consistent with BAA-870's ability to degrade an exceptional range of aromatic compounds. There are at least 35 and 23 various esterases and lipases in the genome of BAA-870, enzymes with immense applications in industry. Rhodococcal genomes are important models for studying actinomycete physiology and can facilitate the exploitation of these industrially important microorganisms. A number of potential genes which could be used for further biocatalysis have also been identified in the genome of *R. rhodochrous* ATCC BAA-870, including epoxide hydrolases, enantioselective enzymes with a range of chemistries, Baeyer-Villiger enzymes, shikimate pathway enzymes, and enzymes which may be useful in biofuel production, drug design targets and kinetic resolutions. The generally large and complex genomes of *Rhodococcus* afford a wide range of genes attributed to extensive secondary metabolic pathways which are presumably responsible for the array of biotransformations and bioremediations demonstrated. These secondary metabolic pathways have yet to be characterised, and offer numerous targets for drug design as well as synthetic chemistry applications.

The locations and numbers of nitrile converting enzymes in the available genomes of *Rhodococcus* were identified. BAA-870 was isolated on a nitrile as the sole source of nitrogen (Brady *et al.* 2004b) and was determined to have nitrile hydrolysis capabilities. However, the genes responsible for the conversions were not known until the genome was sequenced. Now, it is evident that there is one nitrilase and one nitrile hydratase, numerous amidases (mostly of the nitrilase superfamily), and a number of associated regulator elements.

The genome sequence of *R. rhodochrous* ATCC BAA-870 is the first *rhodochrous* subtype to be sequenced and added Genbank. It will be the seventh rhodococcal genome to be submitted to the NCBI, but the fifth subtype addition. The sequence of the *R. rhodochrous* ATCC BAA-870 genome will facilitate the further exploitation of Rhodococci for industrial uses and provide numerous sources for new biotechnology projects and applications, as well as enable further characterisation of a model biotechnologically relevant organism. The genome sequence data has been deposited to GenBank (Registration date 26 November 2011), and given the accession number PRJNA78009 (ID 78009 in BioProject).

5.1.2. Nitrilase and nitrile hydratase from *Rhodococcus rhodochrous* ATCC BAA-870

Nitrilase enzymes catalyse the direct hydrolysis of aliphatic and aromatic nitriles into the corresponding carboxylic acid and ammonia. Applications of nitrilases have advanced in recent years due to ongoing studies detailing their structural and catalytic properties. Industrial nitrilases are used in several biotransformations, including the manufacture of biologically active enantiomers such as (*R*)-mandelic acid, (*S*)-phenyllactic acid, and (*R*)-3-hydroxy-4-cyano-butyric acid, a key intermediate in the synthesis of Lipitor (the blockbuster drug by Pfizer Inc.). Nitrilases are also used in the production of several other high-value compounds, including pyrazinoic acid, nicotinic acid, optically active amino acids and ibuprofen. Several nitrilases are applied by DuPont, Lonza, BASF and Mitsubishi Rayon.

All members of the superfamily with significant nitrilase activity form helical oligomers. The nitrilase gene from BAA-870 was amplified and directionally cloned into a pET28a(+) expression vector. The nitrilase gene consists of 1107 nucleotides, and equates to a 41 kDa translated protein. Nitrilase from BAA-870 was successfully over-induced using a standard IPTG induction in *E. coli* BL21 cells. The nitrilase proteins from *R. rhodochrous* BAA-870, tg1-A6 and J1 share a 98.6% identity. Few major differences are observed, and amino acid changes are mostly conserved. It is likely that the BAA-870 nitrilase also undergoes autolysis before polymerization into fibres as suggested by Thuku *et al.* Purified nitrilase is active against the aromatic nitriles benzonitrile and 3-cyanopyridine. Microscopy of recombinant expressed nitrilase from *R. rhodochrous* ATCC BAA-870 revealed that the protein forms 'c'-shaped oligomers in the pre-fibre form, and that the 'c'-shaped oligomer consists of 8 nitrilase subunits arranged in a left-handed helix, as deduced by docking of the model into an 8-mer volume from the reconstructed nitrilase structure of the nitrilase from *Rhodococcus rhodochrous* J1. When left for an extended time, the octamers assemble into long fibres, as observed with the nitrilase from J1, to form complexes of more than 600 kDa. The purification of the homogeneous pre-fibre form of nitrilase offers the possibility of crystallizing the nitrilase

The L-NHase gene, encoding a non-corrin cobalt containing metalloenzyme, was also isolated from the genome of BAA-870. Induction and expression systems for NHase were optimized. The gene encoding the full L-NHase cassette (beta and alpha subunits and chaperone) were successfully amplified from the genome, and cloned into various cloning and expression vectors. The total size of the NHase gene cassette is 1.8 kb. The NHase gene cassette was expressed using clones contained in both the pET20b(+) and pET28a(+) vectors. However, only expression of the beta subunit and beta subunit homolog (chaperone) was noted.

The ribosome binding sites for each gene in the L-NHase cassette were investigated to improve the expression of the problematic, GC-rich *Rhodococcus* gene targets in *E. coli*. The new RBS-mutated clone did not appear to influence the expression of the alpha subunit. The expression of the chaperone was reduced compared to the beta subunit. Also, stability of NHase upon purification was problematic. The un-coupling of subunits was common, with loss of the beta subunit a frequent occurrence. When pure enzyme was obtained, the protein was seen to precipitate from solution readily. The beta subunit is relatively insoluble, and the alpha subunit is expressed extra-cellularly. It was not clear if the beta subunit was degrading or being collected in exclusion bodies. The beta subunit is considered to be unstable, and was misfolding, or expressed into inclusion bodies during expression. The cloning and expression strategy for NHase was revisited and a two plasmid co-expression strategy was used to ensure that both the beta and alpha subunit expression was independent of each other and under the control of the expression plasmid promoter. The alpha subunit-chaperone sequence and beta subunit sequence were amplified separately from the genome of *R. rhodochrous* ATCC BAA-870 using PCR. The 1060 bp product consisting of the amplified alpha subunit and chaperone and the 675 bp beta subunit gene product were cloned into two separate cloning vectors for independent expression.

The 26 kDa beta subunit, and the 23 kDa alpha subunit together with the 16 kDa chaperone, were successfully expressed from their independent expression vectors. It was found that a ratio of 4:1 pRSFDuet-alpha chaperone: pET21a(+)-beta showed good coexpression of both subunits. The pRSF-Duet construct expressing alpha-chaperone, and the pET21a(+) construct expressing the beta subunit, could be independently induced in separate *E. coli* BL21(DE3)

cultures. Addition of a His-tag at the C-terminus of the NHase beta subunit increased the yield of NHase during purification, and decreased the de-coupling of subunits, possibly by stabilization of the beta subunit when attached to the resin. The purification protocol negated the need for ammonium sulphate fractionation, which was found to cause precipitation of the beta subunit.

The modelling of NHase gave insights into the positioning of the His-tag, with a suggested instability as a consequence of His tag addition to the C-terminus of the beta subunit of NHase. The C-terminus is somewhat near the internal structure, whereas modelling shows the N-terminus is a flexible stretch of protein. The purification of NHase was re-designed to ensure beta subunit binding to the Ni resin during affinity purification by the flexible N-terminus – it was theorized that the chaperone protein could perform its function with either subunit, and the protein could be assembled into the native form with beta bound to the column resin.

Structural alignment with nitrile hydratase from *B. pallidus* showed only a 37% identity with the beta subunit. Based on the closest structural alignments for molecular modeling, the structure of the *B. pallidus* NHase alpha subunit, PDB code 3HHT (Van Wyk *et al.* 2009) and Co-type NHase beta subunit, PDB code 1UGP (Miyanaga *et al.* 2004) were used to generate the structural alignments. The modeled BAA-870 NHase alpha subunit was structurally aligned with the alpha subunit from *B. pallidus* RAPc8 NHase. Structurally, the alpha subunits are highly similar, with only a small extension of the N-terminus visible for the alpha subunit of NHase from BAA-870. There is no visible or obvious difference in the positioning of the active site cysteine residues or the cobalt ion. Similarly, a structural alignment of the beta subunits showed a highly conserved structure. The closest structural homologues of BAA-870 NHase are the *B. pallidus* Rapc8 NHase alpha subunit (Van Wyk *et al.* 2009) and *P. thermophila* Co-type NHase beta subunit (Miyanaga *et al.* 2001). Structurally, the NHase from BAA-870 appears to be fully conserved. The close structural similarity to the *B. pallidus* RAPc8 NHase, in addition to recent error-prone PCR experiments which increased thermostability of the RAPc8 NHase, offer some starting points for site-directed mutagenesis of BAA-870 NHase for increased thermostability, and possibly the introduction of mutations which could allow better transformations of aliphatic nitriles.

The BAA-870 nitrilase was found to be similar, but not identical to, the J1 nitrilase. Like the J1 nitrilase, it is likely that BAA-870 nitrilase also undergoes a C-terminal truncation via autolysis. This study presents a full 3D reconstruction of the *Rhodococcus rhodochrous* BAA-870 nitrilase c-shaped oligomer, whereas in the case of the *Rhodococcus rhodochrous* J1 nitrilase structure only 2-D averages of two different projections have been reported (Thuku *et al.* 2007). The BAA-870 nitrilase pre-fibre form was found to consist of 9-10 subunits (~5 dimers) forming a 1-start, left-handed, c-shaped oligomer. The nitrilase further matures over time to form long oligomeric fibres.

5.1.3. Conclusion

In 2004 the organism BAA-870 was earmarked as a potentially useful nitrile biocatalyst; However, initial results suggested it contained at least two different nitrile hydrolyzing enzyme systems consisting of nitrilase and nitrile hydratase/amidase (Brady *et al.* 2004b; Chhiba *et al.* 2012; Kinfe *et al.* 2009). It was not known if there were multiple isoenzymes or what the copy number was. The assembled and annotated BAA-870 genome has now provided the blueprint for nitrile metabolizing potential, as well as several other biocatalytic targets of interest. For the nitrilase and nitrile hydratase to be of specific use as biocatalysts, it was preferable to remove the enzymes from contaminating influences and downstream enzymatic interference common in whole cell systems. Two nitrile-converting enzymes, a nitrilase and nitrile hydratase, were identified in the genome and successfully amplified from the genome, and cloned into expressions vectors for their convenient expression. The enzymes can now be independently expressed and purified for further characterization studies. This allows the facilitation of further structural and mechanistic studies, and permits future modification of the enzymes at the genetic level. Table 5.1 summarises the intellectual property generated and contributions made during this thesis.

In conclusion, this study has presented the first genome sequence of the biotechnologically relevant organism, *Rhodococcus rhodochrous* ATCC BAA-870. In addition to offering more insight into *Rhodococcus* phylogeny, the organism presents a host of biotechnologically

important protein targets. The nitrilase and nitrile hydratase expression systems facilitate their structural and mechanistic study, offering several future possible enzyme modification strategies.

Table 5.1: Summary of intellectual property generated during the course of this thesis

The Genome Sequence of *Rhodococcus rhodochrous* ATCC BAA-870
The genome of BAA-870 was sequenced and raw output data was assembled by a novel pipeline of programs. Sequence statistics, sequence assemblies and gene annotations were explored and optimised. Several biocatalytic gene targets have been identified on the genome. Specifically, the locations and numbers of nitrile converting enzymes in the genome of BAA-870 were identified. The genome sequence data has been deposited in GenBank (Registration date 26 November 2011; accession number PRJNA78009; ID 78009 in BioProject). This is the first *rhodochrous* subtype to be sequenced and added to Genbank. It is the seventh rhodococcal genome to be submitted to the NCBI.
Plasmid constructs made
The nitrilase gene (1107 bp) in pET28a(+): for standard IPTG induction in *E. coli* BL21(DE3) cells. The full length NHase gene cassette (1.8kb) encoding the L-NHase cassette (beta and alpha subunits and chaperone) in pET20b(+) and pET28a(+): NHase in pET20b(+) for expression in *E. coli* BL21(DE3) containing ampicillin, and induced in the presence of 0.1 mM $CoCl_2$ and 0.1 mM IPTG, and NHase in pET28a(+) for expression in *E. coli* BL21(DE3) containing kanamycin, and induced in the presence of 0.1 mM $CoCl_2$ and 0.1 mM IPTG. NHase in pRSFDuet-1 and pET21a(+): The NHase alpha subunit and chaperone (1060 bp) and the beta (675 bp) subunit genes were cloned into two separate cloning vectors for independent expression. NHase beta subunit, restricted with *Nde*I and *Xho*I, was ligated into pET21a(+). The NHase alpha-chaperone, restricted with *Nde*I and *Pac*I, was ligated into pRSFDuet-1
Continuous spectrophotometric assay for nitrile hydratase activity
Methods Section 4.2.7. A method for monitoring benzamide production from benzonitrile was adapted, optimized and the volume miniaturized for high throughput formats (96-well plate). The production of nicotinamide from 3-cyanopyridine from was also measured using this novel spectrophotometric assay.
Working expression systems
The pET28(+) construct containing nitrilase was induced using L-rhamnose in single step cells. The pET28(+) construct containing nitrilase was induced using IPTG in *E. coli* BL21(DE3). The pET21a(+) construct expressing the NHase beta subunit with C-terminal His-tag was independently induced in *E. coli* BL21(DE3) The pRSF-Duet construct expressing NHase alpha-chaperone, was independently induced in *E. coli* BL21(DE3). The pRSFDuet construct containing the NHase alpha chaperone and the pET21a(+) construct containing NHase beta subunit was co-expressed (optimally at a plasmid ratio of 4:1 respectively).
Nitrilase purification

	Incorporation of an N-terminal His-tag in the pET28(+) construct containing nitrilase allows for an easy two-step chromatographic purification protocol: by nickel column and size exclusion chromatography.
Nitrilase sequence and EM structure	
	BAA-870 nitrilase was found to have a 98 % identity with that from *R. rhodochrous* J1 by mass spectroscopic analysis of wild-type purified nitrilase and by sequence analysis of the cloned nitrilase. Nitrilase was expressed in the native organism as a 600 kDa complex which matured over time to form long fibres, the outer diameter of which was found to be 120 Å on average. The BAA-870 nitrilase pre-fibre form was found to be a left-handed, c-shaped particle. Final 3D reconstruction showed that there are 9-10 subunits (~5 dimers) forming the 1-start, left-handed, c-shaped oligomer, suggesting that the c-shaped nitrilase could form either an octamer or a decamer. The particle outer diameter was measured to be ~11.4 nm and the resolution of the final negative stain density was determined at $FSC_{0.5}$ to be 16 Å. A lead condition for possible nitrilase crystallization of the pre-fibre form was found: 10% w/v PEG 6000, 100 mM bicine, pH 9.
Nitrile hydratase model	
	NHase from BAA-870 is a low-molecular mass, Co-containing NHase. The closest structural homologues of BAA-870 NHase are the *B. pallidus* Rapc8 NHase alpha subunit (Van Wyk *et al.* 2009) and *P. thermophila* Co-type NHase beta subunit (Miyanaga *et al.* 2001). Structurally, the NHase from BAA-870 appears to be almost fully conserved.

References

6. References

Abu-Ruwaida, A. S., Banat, I. M., Haditirto, S., and Khamis, A. (1991), 'Nutritional requirements and growth characteristics of a biosurfactant producing *Rhodococcus* bacterium', *World J. Microb. Biot.,* 7, 53-61.

Agarkar, V. B., Kimani, S. W., Cowan, D. A., Sayed, M. F.-R., and Sewell, B. T. (2006), 'The quaternary structure of the amidase from *Geobacillus pallidus* RAPc8 is revealed by its crystal packing', *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.,* 62, 1171-78.

Aguilar, M.-I., Purcell, A. W., Devi, R., Lew, R., Rossjohn, J., Smith, A. I., and Perlmutter, P. (2007), 'α,β-Amino Acid-Derived Hybrid Peptides – new opportunities in Peptidomimetics', *Org. Biomol. Chem.,* 5, 2884-90.

Ahmad, M., Roberts, J. N., Hardiman, E. M., Singh, R., Eltis, L. D., and Bugg, T. D. H. (2011), 'Identification of DypB from *Rhodococcus jostii* RHA1 as a Lignin Peroxidase', *Biochemistry,* 50 (23), 5096-107.

Almatawah, Q. A. and Cowan, D. A. (1999), 'Thermostable nitrilase catalysed production of nicotinic acid from 3-cyanopyridine', *Enzyme Microb. Technol.,* 25, 718-24.

Almatawah, Q. A., Cramp, R., and Cowan, D. A. (1999), 'Characterization of an inducible nitrilase from a thermophilic bacillus', *Extremophiles,* 3 (4), 283-91.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990), 'Basic Local Alignment Search Tool', *J. Mol. Biol.,* 215, 403-10.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997), 'Gapped BLAST and PSI-BLAST: a new generation of protein database', *Nucleic Acids Res.,* 25, 3389-402.

Amini, S. and Tavazoie, S. (2011), 'Antibiotics and the post-genome revolution', *Current Opinion in Microbiology,* 14 (5), 513-18.

Andrade, J., Karmali, A., Carrondo, M. A., and Frazão, C. (2007a), 'Structure of an amidase from *Pseudomonas aeruginosa* showing a trapped acyl transfer reaction intermediate state', *J. Biol. Chem.,* 282, 19598-605.

Andrade, J., Karmali, A., Carrondo, M. A., and Frazão, C. (2007b), 'Crystallization, diffraction data collection and preliminary crystallographic analysis of hexagonal crystals of *Pseudomonas aeruginosa* amidase', *Acta Crystallogr.,* F63, 214-16.

Arakawa, T., Kawano, Y., Kataoka, S., Katayama, Y., Kamiya, N., Yohda, M., and Odaka, M. (2007), 'Structure of Thiocyanate Hydrolase: A New Nitrile Hydratase Family Protein with a Novel Five-coordinate Cobalt(III) Center', *J. Mol. Biol.,* 366, 1497-509.

Arakawa, T., Kawano, Y., Katayama, Y., Nakayama, H., Dohmae, N., Yohda, M., and Odaka, M. (2009), 'Structural Basis for Catalytic Activation of Thiocyanate Hydrolase Involving Metal-Ligated Cysteine Modification', *Journal of the American Chemical Society,* 131 (41), 14838-43.

Asturias, J. A. and Timmis, K. N. (1993), 'Three different 2,3-dihydroxybiphenyl-1,2-dioxygenase genes in the Gram-positive polychlorobiphenyl-degrading bacterium *Rhodococcus globerulus* P6', *J. Bacteriol.,* 175, 4631-40.

Baeyer, A. and Villiger, V. (1899), 'Einwirkung des Caro'schen Reagens auf Ketone', *Ber. Dtsch. Chem. Ges.,* 32, 3625-33.

Baldwin, P. R. and Penczek, P. A. (2007), 'The Transform Class in SPARX and EMAN2', *J. Struct. Biol.,* 157, 250-61.

Banerjee, A. (2000), 'Stereoselective microbial Baeyer-Villiger oxidations', in Patel, R. N. (ed.), *Stereoselective biocatalysis* (New York: Marcel Dekker, Inc.), 867-76.

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

Banerjee, A., Chase, M., Clayton, R. A., and Landis, B. (2005), 'Methods for the Stereoselective Synthesis and Enantiomeric Enrichment of β-Amino Acids', (US).

Baneyx, F. (1999), 'Recombinant protein expression in *Escherichia coli*', *Curr. Opin. Biotechnol.,* 10, 411-21.

Barglow, K. T., Saikatendu, K. S., Bracey, M. H., Huey, R., Morris, G. H., Olson, A. J., Stevens, R. C., and Cravatt, B. F. (2008), 'Functional Proteomic and Structural Insights into Molecular Recognition in the Nitrilase Family Enzymes', *Biochemistry,* 47, 13514-23.

Barrett, J. F. (2007), 'Impact of Genomics-Emerging Targets for Antibacterial Therapy', in John, B. T. and David, J. T. (eds.), *Comprehensive Medicinal Chemistry II* (7; Oxford: Elsevier), 731-48.

Barth, P. T., Bolton, L., and Thomson, J. C. (1992), 'Cloning and partial sequencing of an operon encoding two *Pseudomonas putida* haloalkanoate dehalogenases of opposite stereospecificity', *J. Bacteriol.,* 174 (8), 2612-19.

Bartling, D., Seedorf, M., Schmidt, R. C., and Weiler, E. W. (1994), 'Molecular characterization of two cloned nitrilases from *Arabidopsis thaliana*: key enzymes in biosynthesis of the plant hormone indole-3-acetic acid', *Proc. Natl. Acad. Sci. USA,* 91, 6021-25.

Beard, C. B., Mason, P. W., Aksoy, S., Tesh, R. B., and Richards, F. F. (1992), 'Transformation of an insect symbiont and expression of a foreign gene in the Chagus Disease vector *Rhodnius prolixus*', *Am. J. Trop. Med. Hyg.,* 46, 195-200.

Behki, R. M., Topp, E. E., and Blackwell, B. A. (1994), 'Ring hydroxylation of N-methylcarbamate insecticides by *Rhodococcus* TE1', *J. Agr. Food Chem.,* 42, 1375-78.

Beierlein, J. M., Deshmukh, L., Frey, K. M., Vinogradova, O., and Anderson, A. C. (2009), 'The Solution Structure of *Bacillus anthracis* Dihydrofolate Reductase Yields Insight into the Analysis of Structure–Activity Relationships for Novel Inhibitors', *Biochemistry,* 48 (19), 4100-08.

Bell, K. S., Philp, J. C., Aw, D. W. J., and Christofi, N. (1998), 'The genus *Rhodococcus*', *J. Appl. Microbiol.,* 85, 195-210.

Ben-Yakhir, D. (1987), 'Growth retardation of *Rhodnius prolixus* symbionts by immunizing host against *Nocardia* (*Rhodococcus*) *rhodnii*', *J. Insect Physiol.,* 33, 379-83.

Bengtsson, O., Hahn-Hägerdal, B., and Gorwa-Grauslund, M. F. (2009), 'Xylose reductase from *Pichia stipitis* with altered coenzyme preference improves ethanolic xylose fermentation by recombinant *Saccharomyces cerevisiae*', *Biotechnology for Biofuels,* 2, 9.

Bennett, B. C., Wan, Q., Ahmad, M. F., Langan, P., and Dealwis, C. G. (2009), 'X-ray structure of the ternary MTX·NADPH complex of the anthrax dihydrofolate reductase: A pharmacophore for dual-site inhibitor design', *J. Struct. Biol.,* 166 (2), 162-71.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000), 'The Protein Data Bank', *Nucleic Acids Res.,* 28, 235-42.

Bhalla, T. C., Miura, A., Wakamoto, A., Ohba, Y., and Furuhashi, K. (1992), 'Asymmetric hydrolysis of alpha aminonitriles to optically active amino acids by a nitrilase of *Rhodococcus rhodochrous* PA-34', *Appl. Microbiol. Biotechnol.,* 37, 184-90.

Bigey, F., Janbon, G., Arnaud, A., and Galzy, P. (1995), 'Sizing of the *Rhodococcus* sp. R312 genome by pulsed-field gel electrophoresis. Localization of genes involved in nitrile degradation', *Antonie Van Leeuwenhoek,* 68, 173-79.

Blackall, L. L. (1994), 'Molecular identification of activated-sludge foaming bacteria', *Water Sci. Technol.,* 29, 35-42.

Blakey, A. J., Colby, J., Williams, E., and O'Reilly, C. (1995), 'Regio-specific and stereo-specific nitrile hydrolysis by the nitrile hydratase from *Rhodococcus* AJ270', *FEMS Microbiol. Lett.,* 129, 57-61.

Bochner, B. R. (2009), 'Global phenotypic characterization of bacteria', *FEMS Microbiol Rev,* 33, 191-205.

Boehm, R., Li, S.-M., Melzer, M., and Heide, L. (1997), '4-hydroxybenzoate prenyltransferases in cell-free extracts of Lithospermum erythrorhizon cell cultures', *Phytochemistry,* 44 (3), 419-24.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011), 'Scaffolding pre-assembled contigs using SSPACE', *Bioinformatics,* 27 (4), 578-79.

Bonnet, D., Artaud, I., Moali, C., Petre, D., and Mansuy, D. (1997), 'Highly efficient control of iron-containing nitrile hydratases by stoichiometric amounts of nitric oxide and light', *FEBS Lett.,* 409, 216-20.

Bork, P. (2000), 'Powers and Pitfalls in Sequence Analysis: The 70% Hurdle', *Genome Res.,* 10, 398-400.

Bosello, M., Mielcarek, A., Giessen, T. W., and Marahiel, M. A. (2012), 'An enzymatic pathway for the biosynthesis of the formylhydroxyornithine required for rhodochelin iron coordination', *Biochemistry,* 51 (14), 3059-66.

Bosello, M., Robbel, L., Linne, U., Xie, X., and Marahiel, M. A. (2011), 'Biosynthesis of the siderophore rhodochelin requires the coordinated expression of three independent gene clusters in Rhodococcus jostii RHA1', *J Am Chem Soc,* 133 (12), 4587-95.

Boyle, A. W., Silvin, C. J., Hassett, J. P., Nakas, J. P., and Tanenbaum, S. W. (1992), 'Bacterial PCB degradation', *Biodegradation,* 3, 285-98.

Bozdemir, T. O., Dursoy, T., Erincin, E., and Yurum, Y. (1996), 'Biodesulfurization of Turkish lignites: Optimization of the growth parameters of *Rhodococcus rhodochrous*, a sulfur-removing bacterium', *Fuel,* 75, 1596-600.

Bradford, M. M. (1976), 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Anal. Biochem.,* 72, 248-54.

Brady, D., Dube, N., and Petersen, R. (2006a), 'Green chemistry: Highly selective biocatalytic hydrolysis of nitrile compounds', *S. Afr. J. Sci.,* 102, 339-44.

Brady, D., Dube, N., and Peterson, R. (2006b), 'Green Chemistry: Highly Selective Biocatalytic Hydrolysis of Nitrile Compounds', *S. Afr. J. Sci.,* in print.

Brady, D., Steenkamp, L., Reddy, S., Skein, E., and Chaplin, J. (2004a), 'Optimisation of the enantioselective biocatalytic hydrolysis of naproxen ethyl ester using ChiroCLEC CR', *Enzyme and Microbial Technology,* 34, 283-91.

Brady, D., Beeton, A., Zeevaart, J., Kgaje, C., van Rantwijk, F., and Sheldon, R. A. (2004b), 'Characterisation of nitrilase and nitrile hydratase biocatalytic systems', *Appl. Microbiol. Biotechnol.,* 64 (1), 76-85.

Brandão, P. F. B., Clapp, J. P., and Bull, A. T. (2002), 'Diversity and taxonomy of geographically diverse strains of nitrile-metabolising actinomycetes using chemometric and molecular sequencing techniques', *Environ. Microbiol.,* 4, 262-76.

Brennan, B. A., Cummings, J. G., Chase, D. B., Turner, I. M. J., and Nelson, M. J. (1996), 'Resonance Raman spectroscopy of nitrile hydratase, a novel iron-sulfur enzyme', *Biochemistry,* 35, 10068-77.

Brenner, C. (2002), 'Catalysis in the nitrilase superfamily', *Curr. Opin. Struct. Biol.,* 12, 775-82.

Brenner, S. E. (1999), 'Errors in genome annotation', *Trends Genet.,* 15, 132-33.

Briglia, M., Middeldorp, P. J. M., and Salkinoja-Salonen, M. S. (1994), 'Mineralization performance of *Rhodococcus chlorophenolicus* strain PCP-1 in contaminated soil simulating on site conditions', *Soil Biol. Biochem.,* 26, 377-85.

Briglia, M., Rainey, F. A., Stackebrandt, E., Schraa, G., and Salkinoja-Salonen, M. S. (1996), '*Rhodococcus percolatus* sp. nov., a bacterium degrading 2,4,6-trichlorophenol', *Int. J. Syst. Bacteriol.,* 46, 23-30.

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), 4923-35.

Bugg, T. D. H., Abell, C., and Coggins, J. R. (1988), 'Specificity of *E. Coli* shikimate dehydrogenase towards analogues of 3-dehydroshikimic acid', *Tetrahedron Letters,* 29 (51), 6779-82.

Bugg, T. D. H., Ahmad, M., Hardiman, E. M., and Singh, R. (2011), 'The emerging role for bacteria in lignin degradation and bio-product formation', *Current Opinion in Biotechnology,* 22 (3), 394-400.

Burton, S. G. and Cowan, D. A. (2002), 'Development of Biotechnology in South Africa', *Electronic J. Biotechnol.,* 5 (1).

Canutescu, A. A., Shelenkov, A. A., and Dunbrack Jr, R. L. (2003), 'A graph-theory algorithm for rapid protein side-chain prediction', *Protein Sci.,* 12 (9), 2001-14.

Chaisson, M., Pevzner, P., and Tang, H. (2004), 'Fragment assembly with short reads', *Bioinformatics,* 20 (13), 2067-74.

Chaisson, M. J. and Pevzner, P. A. (2008), 'Short read fragment assembly of bacterial genomes', *Genome Res.,* 18, 324-30.

Chalker, A. F. and Lunsford, R. D. (2002), 'Rational identification of new antibacterial drug targets that are essential for viability using a genomics-based approach', *Pharmacology & Therapeutics,* 95 (1), 1-20.

Chefson, A. and Auclair, K. (2006), 'Progress towards the easier use of P450 enzymes', *Molecular BioSystems,* 2 (10), 462-69.

Chefson, A. and Auclair, K. (2007), 'CYP3A4 Activity in the Presence of Organic Cosolvents, Ionic Liquids, or Water-Immiscible Organic Solvents', *ChemBioChem,* 8 (10), 1189-97.

Chefson, A., Zhao, J., and Auclair, K. (2006), 'Replacement of Natural Cofactors by Selected Hydrogen Peroxide Donors or Organic Peroxides Results in Improved Activity for CYP3A4 and CYP2D6', *ChemBioChem,* 7 (6), 916-19.

Chen, G., Kayser, M. M., Mihovilovic, M. D., Mrstik, M. E., Martinez, C. A., and Stewart, J. D. (1999), 'Asymmetric oxidations at sulfur catalyzed by engineered strains that overexpress cyclohexanone monooxygenase', *New J. Chem.,* 23 (827-832).

Chen, H.-P., Chow, M., Liu, C. C., Lau, A., Liu, J., and Eltis, L. D. (2012), 'Vanillin Catabolism in *Rhodococcus jostii* RHA1', *Appl. Environ. Microbiol.,* 78 (2), 586-88.

Chen, J., Yu, H., Liu, C., Liu, J., and Shen, Z. (2013), 'Improving stability of nitrile hydratase by bridging the salt-bridges in specific thermal-sensitive regions', *J Biotechnol*.

Chen, S. L., Hung, C. S., Xu, J., Reigstad, C. S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R. R., Ozersky, P., and al, e. (2006), 'Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach', *Proc. Natl. Acad. Sci. USA,* 103, 5977-82.

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, 68 - 74.

Chin, K. H., Tsai, Y. D., Chan, N. L., Huang, K. F., Wang, A. H., and Chou, S. H. (2007), 'The crystal structure of XC1258 from *Xanthomonas campestris*: a putative procaryotic Nit protein with an arsenic adduct in the active site', *Proteins,* 69, 665-71.

Chiu, W.-C., You, J.-Y., Liu, J.-S., Hsu, S.-K., Hsu, W.-H., Shih, C.-H., Hwang, J.-K., and Wang, W.-C. (2006), 'Structure-Stability-Activity Relationship in Covalently Corss-linked *N*-Carbamoyl D-Amino acid Amidohydrolase and *N*-Acylamino acid Racemase', *J. Mol. Biol.,* 359, 741-53.

Chiyanzu, I., Cowan, D. A., and Burton, S. G. (2010), 'Immobilization of *Geobacillus pallidus* RAPc8 nitrile hydratase (NHase) reduces substrate inhibition and enhances thermostability', *J. Mol. Catal. B: Enzymatic,* 63, 109-15.

Chowdhury, E. K., Akaishi, Y., Nagata, S., and Misono, H. (2003), 'Cloning and overexpression of the 3-hydroxyisobutyrate dehydrogenase gene from *pseudomonas putida* E23', *Biosci Biotechnol Biochem.,* 67 (2), 438-41.

Christian, H.-J., Hollergeschwandner, C., Peitzsch, M., Ress-Loeschke, M., Hauer, B., Brandao, P. F., Bunch, A., Robinson, G., Bull, A. T., and Syldatk, C. (2001), *5th Int. Symp. on Biocatalysis and Biotransformation* (Darmstadt), 140.

Christodoulou, S., Hung, T. V., Trewhell, M. A., and Black, R. G. (1994), 'Enzymatic degradation of egg-yolk cholesterol', *J. Food Protect.,* 57, 908-12.

Christofi, N., Ivshina, I. B., Kuyukina, M. S., and Philp, J. C. (1998), 'Biological treatment of crude oil contaminated soil in Russia', in Lerner, D. N. and Walton, N. R. G. (eds.), *Contaminated Land and Groundwater: Future Directions* (Engineering Geology Special Publication edn., 14; London: Geological Society), 45-51.

Coggins, J. R., Abell, C., Evans, L. B., Frederickson, M., Robinson, D. A., Roszak, A. W., and Lapthorn, A. P. (2003), 'Experiences with the shikimate-pathway enzymes as targets for rational drug design', *Biochem. Soc. Trans.,* 31, 548-52.

Colarusso, S., Koch, U., Gerlach, B., Steinkühler, C., De Francesco, R., Altamura, S., Matassa, V. G., and Narjes, F. (2003), 'Phenethyl Amides as Novel Noncovalent Inhibitors of Hepatitis C Virus NS3/4A Protease: Discovery, Initial SAR, and Molecular Modeling', *J. Med. Chem.,* 46 (3), 345-48.

Colonna, S., Gaggero, N., Carrea, G., Ottolina, G., Pasta, P., and Zambianchi, F. (2002), 'First asymmetric epoxidation catalysed by cyclohexanone monooxygenase', *Tetrahedron Letters,* 43 (10), 1797-99.

Cong, Y. and Ludtke, S. J. (2010), 'Single particle analysis at high resolution', *Methods Enzymol,* 482, 211-35.

Conn, E. E. (1981), 'Biosynthesis of cyanogenic glycosides', in Vennesland, B., Conn, E. E., Knowles, C. J., Westley, J., and Wissing, F. (eds.), *Cyanide in Biology* (London: Academic Press), 183-96.

Cowan, D., Cramp, R., Pereira, R., Graham, D., and Almatawah, Q. (1998), 'Biochemistry and biotechnology of mesophilic and thermophilic nitrile metabolizing enzymes', *Extremophiles,* 2 (3), 207-16.

Cramp, R. A. and Cowan, D. A. (1999), 'Molecular characterisation of a novel thermophilic nitrile hydratase', *Biochim. Biophys. Acta,* 1431, 249-60.

Crosby, J., Parratt, J. S., and Turner, N. J. (1992), 'Enzymic hydrolysis of prochiral dinitriles', *Tetrahedron: Asymmetry,* 3, 1547-50.

Crosby, J., Moilliet, J., Parratt, J. S., and Turner, N. J. (1994), 'Regioselective hydrolysis of aromatic dinitriles using a whole cell catalyst', *J. Chem. Soc. Perkin Trans.,* I, 1679-87.

Czernecki, F., Franco, S., and Valery, J.-M. (1997), 'Further Study toward Amipurimycin: Synthesis of the Northern Part', *J. Org. Chem.,* 62 (14), 4845-47.

Davies, S. G., Smyth, G. D., and Chippindale, A. M. (1999), 'Syntheses of derivatives of L-daunosamine and its C-3 epimer employing as the key step the asymmetric conjugate addition of a homochiral lithium amide to *tert*-butyl (*E*,*E*)-hexa-2,4-dienoate', *J. Chem. Soc. Perkin Trans.,* 1, 3089-104.

Davies, S. G., Ichihara, O., Lenoir, I., and Walters, I. A. S. (1994), 'Asymmetric-synthesis of (-)-(1*R*,2*S*)-cispentacin and related cis-2-amino and trans-2-amino cyclopentane-1-carboxylic and cyclohexane-1-carboxylic acids', *J. Chem. Soc. Perkin Trans.,* 1, 1411-15.

de Carvalho, C. C. C. R. and da Fonseca, M. M. R. (2005), 'The remarkable *Rhodococcus erythropolis*', *Appl. Microbiol. Biotechnol.,* 67, 715-26.

de los Reyes, F. L., Ritter, W., and Raskin, L. (1997), 'Group-specific small-subunit rRNA hybridization probes to characterize filamentous foaming in activated sludge systems', *Appl. Environ. Microbiol.,* 63, 1107-17.

del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M., and Diaz-Orejas, R. (1998), 'Replication and control of circular bacterial plasmids', *Microbiol. Mol. Biol. Rev.,* 62, 434-64.

DeLong, E. F. (2009), 'The microbial ocean from genomes to biomes', *Nature,* 459, 200-06.

Denome, S. A., Oldfield, C., Nash, L. J., and Young, K. D. (1994), 'Characterization of the desulfurization genes from *Rhodococcus* sp. strain ICTS8', *J. Bacteriol.,* 176, 6707-16.

Dent, K., Weber, B., Benedik, M., and Sewell, B. (2009), 'The cyanide hydratase from *Neurospora crassa* forms a helix which has a dimeric repeat', *Applied Microbiology and Biotechnology,* 82 (2), 271-78.

Desai, L. V. and Zimmer, M. (2004), 'Substrate selectivity and conformational space available to bromoxynil and acrylonitrile in iron nitrile hydratase', *Dalton Trans.,* 6, 872-77.

Devos, D. and Valencia, A. (2001), 'Intrinsic errors in genome annotation', *Trends Genet.,* 17, 429-31.

Dhillon, J., Chhatre, S., Shanker, R., and Shivaraman, N. (1999), 'Transformation of aliphatic and aromatic nitriles by a nitrilase from *Pseudomonas* sp.', *Can. J. Microbiol.,* 45 (10), 811-15.

Diez, A., Prieto, M. I., Alvarez, M. J., Bautista, J. M., Garrido, A., and Puyet, A. (1996), 'Improved Catalytic Performance of a 2-Haloacid Dehalogenase from Azotobacter sp. by Ion-Exchange Immobilisation', *Biochemical and Biophysical Research Communications,* 220 (3), 828-33.

Doan, P. E., Nelson, M. J., Jin, H., and Hoffman, B. (1996), 'An Implicit TRIPLE Effect in Mims Pulsed ENDOR: A Sensitive New Technique for Determining Signs of Hyperfine Couplings', *J. Am. Chem. Soc.,* 118, 7014-15.

Dravis, B. C., Swanson, P. E., and Russell, A. J. (2001), 'Haloalkane hydrolysis with an immobilized haloalkane dehalogenase', *Biotechnology and Bioengineering,* 75 (4), 416-23.

Dufour, É., Storer, A. C., and Ménard, R. (1995), 'Engineering nitrile hydratase activity into a cysteine protease by a single mutation', *Biochemistry,* 34, 16382-88.
Dufour, É., Tam, W., Nägler, D. K., Storer, A. C., and Ménard, R. (1998), 'Synthesis of amidrazones using an engineered papain nitrile hydratase', *FEBS Lett.,* 433, 78-82.
Duz, M., Whittaker, A. G., Love, S., Parkin, T. D., and Hughes, K. J. (2009), 'Exhaled breath condensate hydrogen peroxide and pH for the assessment of lower airway inflammation in the horse', *Res Vet Sci,* 87 (2), 307-12.
Eason, J. R., Jameson, P. E., and Bannister, P. (1995), 'Virulence assessment of *Rhodococcus fascians* strains on pea cultivars', *Plant Pathol.,* 44, 141-47.
Effenberger, F. and Böhme, J. (1994), 'Enzyme-catalysed enantioselective hydrolysis of racemic naproxen nitrile', *Bioorg. Med. Chem.,* 2, 715-21.
Effenberger, F. and Oβwald, S. (2001), 'Enantioselective hydrolysis of (*RS*)-2-fluoroarylacetonitriles using nitrilase from *Arabidopsis thaliana*', *Tetrahedron: Asymmetry,* 12 (2), 279-85.
Eichhorn, E., Roduit, J. P., Shaw, N., Heinzmann, K., and Kiener, A. (1997), 'Preparation of (S)-piperazine-2-carboxylic acid, (R)-piperazine-2-carboxylic acid, and (S)-piperidine-2-carboxylic acid by kinetic resolution of the corresponding racemic carboxamides with stereoselective amidases in whole bacterial cells', *Tetrahedron: Asymmetry,* 8, 2533-36.
El Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., Williamson, J. R., Schimmel, P., Swairjo, M. A., and de Crécy-Lagard, V. (2009), 'The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA', *Nucleic Acids Res.,* 37 (9), 2894-909.
Embley, T. M. and Stackebrandt, E. (1994), 'The molecular phylogeny and systematics of the actinomycetes', *Annu. Rev. Microbiol.,* 48, 257-89.
Endo, I. and Odaka, M. (2000), 'What evidences were elucidated about photoreactive nitrile hydratase?', *J. Mol. Catal. B: Enzymatic,* 10, 81-86.
Endo, I., Odaka, M., and Yohda, M. (1999a), 'An enzyme controlled by light: the molecular mechanism of photoreactivity in nitrile hydratase', *TRENDS Biotechnol.,* 17, 244-49.
Endo, I., Odaka, M., and Yohda, M. (1999b), 'An enzyme controlled by light: the molecular mechanism of photoreactivity in nitrile hydratase', *TRENDS Biotechnol.,* 17 (6), 244-48.
Endo, I., Nojiri, M., Tsujimura, M., Nakasako, M., Nagashima, S., Yohda, M., and Odaka, M. (2001), 'Fe-type nitrile hydratase', *J. Inorg. Biochem.,* 83, 247-53.
Erb, A., Weiß, H., Härle, J., and Bechthold, A. (2009), 'A bacterial glycosyltransferase gene toolbox: Generation and applications', *Phytochemistry,* 70 (15-16), 1812-21.
Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-y., Pieper, U., and Sali, A. (2001), 'Comparative Protein Structure Modeling Using MODELLER', *Current Protocols in Protein Science* (John Wiley & Sons, Inc.).
Faber, K. (1995), *Biotransformations in Organic Chemistry* (2nd edn.; Berlin: Springer-Verlag).
Faber, K. (1997), 'Biotransformation of non-natural compounds: State of the art and future development', *Pure and Applied Chemistry,* 69 (8), 1613-32.
Fawcett, J. K. and Scott, J. E. (1960), 'A rapid and precise method for the determination of urea', *J. Clin. Pathol.,* 13, 156-59.
Ferrari, R., Ceconi, C., Curello, S., Ghielmi, S., and Albertini, A. (1989), 'Superoxide dismutase: Possible therapeutic use in cardiovascular disease', *Pharmacological Research,* 21 (Supplement 2), 57-65.

Fiechter, A. (1992), 'Biosurfactants: moving towards industrial application', *Trends Biotechnol.,* 10, 208-17.
Fink, M. J., Rudroff, F., and Mihovilovic, M. D. (2011), 'Baeyer–Villiger monooxygenases in aroma compound synthesis', *Bioorg. Medicinal Chem. Lett.,* 21 (20), 6135-38.
Finnerty, W. M. (1992), 'The biology and genetics of the genus *Rhodococcus*', *Annu. Rev. Biochem.,* 46, 193-218.
Finnerty, W. M. (1994), 'Biosurfactants in environmental biotechnology', *Curr. Opin. Biotech.,* 5, 291-95.
Fougias, E. and Forster, C. F. (1994), '*Rhodococcus rubra* in relation to stable foams in activated sludge', *Process Biochem.,* 29, 553-57.
Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and al, e. (1996), 'SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields', *J. Struct. Biol.,* 116, 190-99.
Fredrickson, J. K., Zachara, J. M., Balkwill, D. L., Kennedy, D., Li, S. M., Kostandarithes, H. M., Daly, M. J., Romine, M. F., and Brockman, F. J. (2004), 'Geomicrobiology of high-level waste-contaminated vadose sediments at the Hanford site, Washington state', *Appl. Environ. Microbiol.,* 70, 4230-41.
Fretland, A. J. and Omiecinski, C. J. (2000), 'Epoxide hydrolases: biochemistry and molecular biology', *Chemico-Biological Interactions,* 129 (1-2), 41-59.
Frey, P. A. and Magnusson, O. T. (2003), 'S-adenosylmethionine: A wolf in sheep's clothing, or a rich man's adenosylcobalamin?', *Chemical Rev.,* 103 (6), 2129-48.
Frishman, D. (2007), 'Protein annotation at genomic scale: the current status', *Chem. Rev.,* 107 (8), 3448-66.
Frishman, D., Mokrejs, M., Kosykh, D., Kastenmüller, G., Kolesov, G., Zubrzycki, I., Gruber, C., Geier, B., Kaps, A., Albermann, K., Volz, A., Wagner, C., Fellenberg, M., Heumann, K., and Mewes, H.-W. (2003), 'The PEDANT genome database', *Nucleic Acids Res.,* 31 (1), 207-11.
Fulco, A. J. and Ruettinger, R. T. (1987), 'Occurrence of a barbiturate-inducible catalytically self-sufficient 119,000 Dalton cytochrome P-450 monooxygenase in bacilli', *Life Sciences,* 40 (18), 1769-75.
Furuya, T. and Kino, K. (2009), 'Discovery of 2-Naphthoic Acid Monooxygenases by Genome Mining and their Use as Biocatalysts', *ChemSusChem,* 2 (7), 645-49.
Gademann, K., Hintermann, T., and Schreiber, J. V. (1999), 'Beta-peptides: twisting and turning', *Curr. Med. Chem.,* 6, 905-25.
Galperin, M. Y. (2005), 'A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts', *BMC Microbiol.,* 5, 35.
Galperin, M. Y. and Koonin, E. V. (2004), ''Conserved hypothetical' proteins: prioritization of targets for experimental study', *Nucleic Acid Res.,* 32 (18), 5452-63.
Galperin, M. Y., Moroz, O. V., Wilson, K. S., and Murzin, A. G. (2006), 'House cleaning, a part of good housekeeping', *Mol. Microbiol.,* 59 (1), 5-19.
Gardner, M. J., Shallom, S. J., Carlton, J. M., Salzberg, S. L., Nene, V., Shoaibi, A., Ciecko, A., Lynn, J., Rizzo, M., Weaver, B., Jarrahi, B., Brenner, M., Parvizi, B., Tallon, L., Moazzez, A., Granger, D., Fujii, C., Hansen, C., Pederson, J., Feldblyum, T., Peterson, J., Suh, B., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., White, O., Cummings, L. M., Smith, H. O., Adams, M. D., Venter, J. C., Carucci, D. J., Hoffman, S. L., and Fraser, C. M. (2002), 'Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11, and 14', *Nature,* 419, 531-34.

Garner, P. and Ramakanth, S. (1986), 'Stereodivergent Syntheses of *threo*- and *erythro*-6-Amino-6-deoxyheptosulose Derivatives via an Optically Active Oxazolidine Aldehyde', *J. Org. Chem.,* 51, 2609-12.

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

Ge, W., Clifton, I. J., Stok, J. E., Adlington, R. M., Baldwin, J. E., and Rutledge, P. J. (2008), 'Isopenicillin N Synthase Mediates Thiolate Oxidation to Sulfenate in a Depsipeptide Substrate Analogue: Implications for Oxygen Binding and a Link to Nitrile Hydratase?', *Journal of the American Chemical Society,* 130 (31), 10096-102.

Gedey, S., Liljeblad, A., Lázár, L., Fülöp, F., and Kanerva, L. T. (2001), 'Preparation of highly enantiopure [beta]-amino esters by Candida antarctica lipase A', *Tetrahedron: Asymmetry,* 12 (1), 105-10.

Geueke, B. and Kohler, H. P. (2007), 'Bacterial β-peptidyl aminopeptidases: on the hydrolytic degradation of β-peptides', *Applied Microbiology and Biotechnology,* 74 (6), 1197-204.

Goldschmidt, L., Cooper, D., Derewenda, Z., and Eisenberg, D. (2007), 'Toward rational protein crystallization: A Web server for the design of crystallizable protein variants', *Protein Science,* 16, 1569-76.

Gong, J.-S., Lu, Z.-M., Li, H., Shi, J.-S., Zhou, Z.-M., and Xu, Z.-H. (2012), 'Nitrilases in nitrile biocatalysis: recent progress and forthcoming research', *Microbial Cell Factories,* 11 (1), 142.

Goodfellow, M. (1989a), 'Genus *Rhodococcus*', in Williams, S. T., Sharpe, M. E., and Holt, J. G. (eds.), *Bergey's Manual of Systematic Bacteriology* (4; Baltimore: Williams and WIlkins), 2362-71.

Goodfellow, M. (1989b), 'Supragenic classification of actinomycetes', in Williams, S. T., Sharpe, M. E., and Holt, J. G. (eds.), *Bergey's Manual of Systematic Bacteriology* (4; Baltimore: Williams and Wilkins), 2333-39.

Goodfellow, M. and Alderson, G. (1977), 'The actinomycete-genus *Rhodococcus*: a home for the '*rhodochrous*' complex', *J. Gen. Microbiol.,* 100, 99-122.

Goodfellow, M., Thomas, E. G., Ward, A. C., and James, A. L. (1990), 'Classification and identification of the rhodococci', *Zentralblatt für Bakteriologie,* 274, 299-315.

Goodfellow, M., Davenport, R., Stainsby, F. M., and Curtis, T. P. (1997), 'Actinomycete diversity associated with foaming in activated sludge plants', *J. Ind. Microbiol.,* 17, 268-80.

Gradley, M. L., Deverson, C. J. F., and Knowles, C. J. (1994), 'Asymmetric hydrolysis of R-(–),S(+)-2-methylbutyronitrile by *Rhodococcus rhodochrous* NCIMB 11216', *Arch. Microbiol.,* 161 (3), 246-51.

Gray, K. A., Pogrebinsky, O. S., Mrachko, G. T., Xi, L., Monticello, D. J., and Squires, C. H. (1996), 'Molecular mechanisms of biocatalytic desulfurization of fossil fuels', *Nature Biotechnol.,* 14, 1705-09.

Grayhack, E. J. and Phizicky, E. M. (2001), 'Genomic analysis of biochemical function', *Curr. Opin. Chem. Biol.,* 5, 34-39.

Green, M. L. and Karp, P. D. (2004), 'A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases', *BMC Bioinformatics,* 5, 76.

Greene, S. N. and Richards, N. G. J. (2006), 'Electronic structure, bonding, spectroscopy and energetics of Fe-dependent nitrile hydratase active-site models', *Inorg. Chem.,* 45, 17-36.

Grifantini, R., Pratesi, C., Galli, G., and Grandi, G. (1996a), 'Topological mapping of the cysteine residues of N-carbamyl-D-amino-acid amidohydrolase and their role in enzymatic activity', *J. Biol. Chem.,* 271, 9326-31.

Grifantini, R., Pratesi, C., Galli, G., and Grandi, G. (1996b), 'Topological mapping of the cysteine residues of N-carbamyl-D-amino-acid amidohydrolase and their role in enzymatic activity', *J. Biol. Chem.,* 271 (16), 9326-31.

Grogan, G. (2011), 'Cytochromes P450: exploiting diversity and enabling application as biocatalysts', *Current Opinion in Chemical Biology,* 15 (2), 241-48.

Gröger, H., Trauthwein, H., Buchholz, S., Drauz, K., Sacherer, C., Godfrin, S., and Werner, H. (2004), 'The first aminoacylase-catalyzed enantioselective synthesis of aromatic β-amino acids', *Org. Biomol. Chem.,* 2 (14), 1977-78.

Gromiha, M. M. and Selvaraj, S. (1998), 'Protein secondary structure prediction in different structural classes', *Protein Engineering,* 4, 249-51.

Guo, Y., Ribeiro, J. M. C., Anderson, J. M., and Bour, S. (2009), 'dCAS: a desktop application for cDNA sequence annotation', *Bioinformatics,* 25 (9), 1195-96.

Gutiérrez-García, V. M., López-Ruiz, H., Reyes-Rangel, G., and Juaristi, E. (2001), 'Enantioselective synthesis of [beta]-amino acids. Part 11: Diastereoselective alkylation of chiral derivatives of [beta]-aminopropionic acid containing the [alpha]-phenethyl group', *Tetrahedron,* 57 (30), 6487-96.

Häggblom, M. M., Nohynek, L. J., Palleroni, N. J., Kronqvist, K., Nurmiaho-Lassila, E. L., Salkinoja-Salonen, M. S., Klatte, S., and Kroppenstedt, R. M. (1994), 'Transfer of polychlorophenol-degrading *Rhodococcus chlorophenolicus* (Apajalahti *et al*. 1986) to the genus *Mycobacterium* as *Mycobacterium chlorophenolicum* comb. nov.', *Int. J. Syst. Bacteriol.,* 44, 485-93.

Halden, R. U., Peters, E. G., Halden, B. G., and Dwyer, D. F. (2000), 'Transformation of mono- and dichlorinated phenoxybenzoates by phenoxybenzoate-dioxygenase in pseudomonas pseudoalcaligenes POB310 and a modified diarylether-metabolizing bacterium', *Biotechnol. Bioeng.,* 69 (1), 107-12.

Hall, T. A. (1999), 'BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT', *Nucl. Acids Symp. Ser,* 41, 95-98.

Hammond, R. J., Poston, B. W., Ghiviriga, I., and Feske, B. D. (2007), 'Biocatalytic synthesis towards both antipodes of 3-hydroxy-3-phenylpropanitrile a precursor to fluoxetine, atomoxetine and nisoxetine', *Tetrahedron Lett.,* 48, 1217-19.

Hancock, R. D., Galpin, J. R., and Viola, R. (2000), 'Biosynthesis of L-ascorbic acid (vitamin C) by *Saccharomyces cerevisiae*', *FEMS Microbiol. Lett.,* 186 (2), 245-50.

Hao, J., Ma, C., Gao, C., Qiu, J., Wang, M., Zhang, Y., Cui, X., and Xu, P. (2007), '*Pseudomonas stutzeri* as a novel biocatalyst for pyruvate production from DL-lactate', *Biotechnol. Lett.,* 29 (1), 105-10.

Harrop, T. C. and Mascharak, P. K. (2004), 'Fe(III) and Co(III) Centers with Carboxamido Nitrogen and Modified Sulfur Coordination: Lessons Learned from Nitrile Hydratase', *Acc. Chem. Res.,* 37, 253-60.

Hashimoto, K., Suzuki, H., Taniguchi, K., Noguchi, T., Yohda, M., and Odaka, M. (2008), 'Catalytic Mechanism of Nitrile Hydratase Proposed by Time-resolved X-ray

Crystallography Using a Novel Substrate, *tert*-Butylisonitrile', *J. Biol. Chem.,* 283 (52), 36617-23.

Heald, S. C., Brandao, P. F., Hardicre, R., and Bull, A. T. (2001), 'Physiology, biochemistry and taxonomy of deep-sea nitrile metabolising *Rhodoccocus* strains', *Antonie Van Leeuwenhoek,* 80, 169-83.

Heck, T., Geueke, B., Seebach, D., Osswald, S., Ter Wiel, M. K. J., and Kohler, H. P. E. (2009), '[beta]-Aminopeptidases: enzymatic degradation and synthesis of [beta]-amino acid containing peptides and kinetic resolution of [beta]-amino acid amides', *New Biotechnology,* 25 (Supplement 1), S62-S62.

Heinrich, L., Mary-Verla, A., Li, Y., Vassermann, J., and Chottard, J. C. (2001), 'Cobalt(III) complexes with carboxyamide-N and sulfenato-S or sulfinato-S ligands suggest that a coordinated sulfenato-S is essential for the catalytic activity of nitrile hydratases', *Eur. J. Inorg. Chem.,* 9, 2203-06.

Heiss, G. S., Gowan, B., and Dabbs, E. R. (1992), 'Cloning of DNA from a *Rhodococcus* strain conferring the ability to decolorize sulfonated azo dyes', *FEMS Microbiol. Lett.,* 99, 221-26.

Hernandez, D., François, P., Farinelli, L., Osteras, M., and J., S. (2008), 'De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer', *Genome Res.,* 18, 802-09.

Hjort, C. M., Godtferedson, S. E., and Emborg, C. (1990), 'Isolation and characterisation of a nitrile hydratase from a *Rhodococcus* sp.', *J. Chem. Technol. Biotechnol.,* 48, 217-26.

Hohn, M., Tang, G., Goodyear, G., Baldwin, P. R., Huang, Z., Penczek, P. A., Yang, C., Glaeser, R. M., Adams, P. D., and Ludtke, S. J. (2007), 'SPARX, a new environment for Cryo-EM image processing', *J. Struct. Biol.,* 157, 47-55.

Holder, J. W., Ulrich, J. C., DeBono, A. C., Godfrey, P. A., Desjardins, C. A., Zucker, J., Zeng, Q., Leach, A. L. B., Ghiviriga, I., Dancel, C., Abeel, T., Gevers, D., Kodira, C. D., Desany, B., Affourtit, J. P., Birren, B. W., and Sinskey, A. J. (2011), 'Comparative and Functional Genomics of *Rhodococcus opacus* PD630 for Biofuels Development', *PLoS Genet.,* 7 (9), e1002219.

Holton, R. A., Somoza, C., Kim, H.-B., Liang, F., Biediger, R. J., Boatman, P. D., Shindo, M., Smith, C. C., Kim, S., Nadizadeh, H., Suzuki, Y., Tao, C., Vu, P., Tang, S., Zhang, P., Murthi, K. K., Gentile, L. N., and Liu, J. H. (1994), 'First Total Synthesis of Taxol', *J. Am. Chem. Soc.,* 116, 1597-98.

Honda, J., Kandori, H., Okada, T., Nagamune, T., Shichida, Y., Sasabe, H., and Endo, I. (1994), 'Spectroscopic observation of the intramolecular electron transfer in the photoactivation processes of nitrile hydratase', *Biochemistry,* 33, 3577-83.

Hopmann, K. H. and Himo, F. (2008), 'Theoretical Investigation of the Second-Shell Mechanism of Nitrile Hydratase', *Eur. J. Inorg. Chem.,* 2008, 1406-12.

Hopmann, K. H., Guo, J. D., and Himo, F. (2007), 'Theoretical investigation of the first-shell mechanism of nitrile hydratase', *Inorg. Chem.,* 46, 4850-56.

Hourai, S., Miki, M., Takashima, Y., Mitsuda, S., and Yanagi, K. (2003), 'Crystal structure of nitrile hydratase from a thermophilic *Bacillus smithii*', *Biochem. Biophys. Res. Comm.,* 312, 340-45.

Hsiao, Y., Rivera, N. R., Rosner, T., Krska, S. W., Njolita, E., Wang, F., Sun, Y., Armstrong, J. D., Grabowski, E. J. J., Tillyer, R. D., Spindler, F., and Malan, C. (2004), 'Highly Efficient Synthesis of β-Amino Acid Derivatives via Asymmetric Hydrogenation of Unprotected Enamines', *J. Am. Chem. Soc.,* 126, 9918-19.

Huang, W., Jia, J., Cummings, J., Nelson, M., Schneider, G., and Lindqvist, Y. (1997), 'Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold', *Structure,* 5, 691-99.

Hugenholtz, P. (2002), 'Exploring prokaryotic diversity in the genomic era', *Genome Biology,* 3, REVIEWS0003.

Hummel, W. (1997), 'Screening and characterization of new enzymes for biosensing and analytics', in Scheller, F. W., Scubert, F., and Fedrowitz, J. (eds.), *Frontiers in Biosensorics I, Fundamental Aspects* (Basel: Birkhauser erlag), 49-61.

Hummel, W., Schutte, H., Schmidt, E., Wandrey, C., and Kula, M.-R. (1987), 'Isolation of L-phenylalanine dehydrogenase from *Rhodococcus* sp. M4 and its application for the production of L-phenylalanine', *Appl. Microbiol. Biotechnol.,* 26, 409-16.

Hung, C. L., Liu, J. H., Chiu, W.-C., Huang, S. W., Hwang, J.-K., and Wang, W.-C. (2007), 'Crystal structure of *Helicobacter pylori* formamidase AmiF reveals a cysteine-glutamate-lysine catalyic triad', *J. Biol. Chem.,* 282, 12220-29.

Iborra, J. L., Manjon, A., Canovas, M., Lozano, P., and Martinez, C. (1994), 'Continuous limonin degradation by immobilizes *Rhodococcus fascians* cells in kappa-carrageenan', *Appl. Microbiol. Biotechnol.,* 41, 487-93.

Ikehata, O., Nishiyama, M., Horinouchi, S., and Beppu, T. (1989), 'Primary structure of nitrile hydratase deduced from the nucleotide sequence of a *Rhodococcus* species and its expression in *Escherichia coli*', *Eur. J. Biochem.,* 181, 563-70.

Ivshina, I. B., Oborin, A. A., Nesterenko, O. A., and Kasumova, S. A. (1994), 'Bacteria of the *Rhodococcus* genus from the ground water of oil-bearing deposits in the Perm region near the Urals', *Mikrobiolgiya,* 50, 709-17.

Izumi, Y., Ohshiro, T., Ogino, H., Hine, Y., and Shimao, M. (1994), 'Selective desulfurization of dibenzothiopene by *Rhodococcus erythropolis* D1', *Appl. Environ. Microbiol.,* 60, 223-26.

Jandhyala, D., Berman, M., Meyers, P. R., Sewell, B. T., Willson, R. C., and Benedik, M. J. (2003), 'CynD, the Cyanide Dihydratase from *Bacillus pumilus*: Gene Cloning and Structural Studies', *Appl. Environ. Microbiol.,* 69 (8), 4794-805.

Janowitz, T., Trompetter, I., and Piotrowski, M. (2009), 'Evolution of nitrilases in glucosinolate-containing plants', *Phytochemistry,* 70, 1680-86.

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., and von Mering, C. (2009), 'STRING 8 - a global view on proteins and their functional interactions in 630 organisms', *Nucleic Acids Res.,* 37, D412-D16.

Jiménez, J. I., Minambres, B., García, J. L., and Díaz, E. (2002), 'Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440', *Environ. Microbiol.,* 4 (12), 824-41.

Jiménez, J. I., Minambres, B., García, J. L., and Díaz, E. (2004), 'Genomic insights in the metabolism of aromatic compounds in Pseudomonas', in Ramos, J.-L. (ed.), *Pseudomonas* (3; New York: Kluwer Academic/Plenum), 425-62.

Jin, H., Turner, I. M. J., Nelson, M. J., Gurbiel, R. J., Doan, P. E., and Hoffman, B. (1993), 'Coordination sphere of the ferric ion in nitrile hydratase', *J. Am. Chem. Soc.,* 115, 5290-91.

Jonasson, P., Liljeqvist, S., Nygren, P. A., and Stahl, S. (2002), 'Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*', *Biotechnol. Appl. Biochem.,* 35, 91-105.

Jones, C. E., Brown, A. L., and Baumann, U. (2007), 'Estimating the annotation error rate of curated GO database sequence annotations', *BMC Bioinformatics,* 8, 170.
Jones, D. T. (1999), 'GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences', *J. Mol. Biol.,* 287, 797-815.
Jones, D. T., McGuffin, L. J., and Bryson, K. (2000), 'The PSIPRED protein structure prediction server', *Bioinformatics,* 16, 404-05.
Joska, T. M. and Anderson, A. C. (2006), 'Structure-Activity Relationships of *Bacillus cereus* and *Bacillus anthracis* Dihydrofolate Reductase: toward the Identification of New Potent Drug Leads', *Antimicrob Agents Chemother.,* 50 (10), 3435-43.
Joyeux, L. and Penczek, P. A. (2002), 'Efficiency of 2D alignment methods', *Ultramicroscopy,* 92, 33-46.
Juaristi, E., Balderas, M., López-Ruiz, H., Jiménez-Pérez, V. M., Kaiser-Carril, M. L., and Ramírez-Quirós, Y. (1999), 'Enantioselective synthesis of [beta]-amino acids. Part 10: Preparation of novel [alpha],[alpha]- and [beta],[beta]-disubstituted [beta]-amino acids from (S)-asparagine[]', *Tetrahedron: Asymmetry,* 10 (18), 3493-505.
Kakeya, H., Sakai, N., Sano, A., Yokoyama, M., Sugai, T., and Ohta, H. (1991), 'Microbial hydrolysis of 3-substituted glutaronitriles', *Chem. Lett.,* 20, 1823-24.
Kamal, A. and Ramesh Khanna, G. B. (2001), 'A facile preparation of (±)-β-hydroxy nitriles and their enzymatic resolution with lipases', *Tetrahedron: Asymmetry,* 12, 405-10.
Kamal, A., Ramesh Khanna, G. B., and Ramu, R. (2002), 'Chemoenzymatic synthesis of both enantiomers of fluoxetine, tomoxetine and nisoxetine: lipase-catalyzed resolution of 3-aryl-3-hydroxypropanenitriles', *Tetrahedron: Asymmetry,* 13, 2039-51.
Kamal, A., Ramesh Khanna, G. B., Krishnaji, T., and Ramu, R. (2005a), 'A new facile chemoenzymatic synthesis of levamisole', *Bioorg. Med. Chem. Lett.,* 15, 613-15.
Kamal, A., Ramesh Khanna, G. B., Krishnaji, T., Tekumalla, V., and Ramu, R. (2005b), 'New chemoenzymatic pathway for β-adrenergic blocking agents', *Tetrahedron: Asymmetry,* 16, 1485-94.
Kamerbeek, N. M., Janssen, D. B., van Berkel, W. J. H., and Fraaije, M. W. (2003), 'Baeyer–Villiger Monooxygenases, an Emerging Family of Flavin-Dependent Biocatalysts', *Advanced Synthesis & Catalysis,* 345 (6-7), 667-78.
Kane, J. F., Violand, B. N., Curran, D. F., Staten, N. R., Duffin, K. L., and Bogosian, G. (1992), 'Novel in-frame two codon translational hop suring synthesis of bovine placental lactogen in a recombinant strain of *Escherichia coli*', *Nuc. Acids Res.,* 20, 6707-12.
Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004), 'The KEGG resource for deciphering the genome', *Nucleic Acids Res.,* 32, D277-D80.
Kaplan, Bezouska, K., Malandra, A., Veselá, A. B., Petříčková, A., Felsberg, J., Rinágelová, A., and Křen, V. (2011a), 'Genome mining for the discovery of new nitrilases in filamentous fungi', *Biotechnol. Lett.,* 33, 309-12.
Kaplan, O., Bezouška, K., Plíhal, O., Ettrich, R., Kulik, N., Vaněk, O., Kavan, D., Benada, O., Malandra, A., Šveda, O., Veselá, A. B., Rinágelová, A., Slámová, K., Cantarella, M., Felsberg, J., Dušková, J., Dohnálek, J., Kotik, M., Křen, V., and Martínková, L. (2011b), 'Heterologous expression, purification and characterization of nitrilase from *Aspergillus niger* K10', *BMC Biotechnol.,* 11 (2).
Karp, P. D. (2004), 'Call for an enzyme genomics initiative', *Genome Biol.,* 5 (8), 401.
Kashiwagi, M., Fuhshuku, K.-I., and Sugai, T. (2004), 'Control of the nitrile-hydrolyzing enzyme activity in *Rhodococcus rhodochrous* IFO 15564: preferential action of nitrile hydratase and amidase depending on the reaction condition factors and its

application to the one-pot preparation of amides from aldehydes', *J. Mol. Biocat. B: Enz,* 29, 249-58.
Kato, Y., Ooi, R., and Asano, Y. (1998), 'Isolation and characterization of a bacterium possessing a novel aldoxime-dehydration activity and nitrile-degrading enzymes', *Arch. Microbiol.,* 170, 85-90.
Kato, Y., Tsuda, T., and Asano, Y. (1999), 'Nitrile hydratase involved in aldoxime metabolism from *Rhodococcus* sp. strain YH3-3: Purification and characterization', *Eur. J. Biochem.,* 263, 662-70.
Kaul, P., Banerjee, A., Mayilraj, S., and Banerjee, U. C. (2004), 'Screening for enantioselective nitrilases: kinetic resolution of racemic mandelonitrile to (*R*)-(–)-mandelic acid by new bacterial isolates', *Tetrahedron: Asymmetry,* 15 (2), 207-11.
Kayser, K. J., Bielaga-Jones, B. A., Jackowski, K., Odusan, O., and Kilbane, J. J. (1993), 'Utilization of organosulphur compounds by axenic and mixed cultures of *Rhodococcus rhodochrous* IGTS8', *J. Gen. Microbiol.,* 139, 3123-29.
Kayser, M. M. (2009), '‘Designer reagents’ recombinant microorganisms: new and powerful tools for organic synthesis', *Tetrahedron,* 65 (5), 947-74.
Kayser, M. M., Drolet, M., and Stewart, J. D. (2005), 'Application of newly available bio-reducing agents to the synthesis of chiral hydroxy-[beta]-lactams: model for aldose reductase selectivity', *Tetrahedron: Asymmetry,* 16 (24), 4004-09.
Ken, C.-F., Lin, C.-T., Shaw, J.-F., and Wu, J.-L. (2003), 'Characterization of Fish Cu/Zn–Superoxide Dismutase and Its Protection from Oxidative Stress', *Marine Biotechnology,* 5 (2), 167-73.
Kent, B. N., Salichos, L., Gibbons, J. G., Rokas, A., Newton, I. L. G., Clark, M. E., and Bordenstein, S. R. (2011), 'Complete Bacteriophage Transfer in a Bacterial Endosymbiont (*Wolbachia*) Determined by Targeted Genome Capture', *Genome Biol. Evol.,* 3, 209-18.
Kieslich, K. (1991), 'Biotransformations of industrial use', *Acta Biotechnologica,* 11, 559-70.
Kilbane, J. J. (1989), 'Desulfurization of coal: the microbial solution', *Trends Biotechnol.,* 7, 97-101.
Kim, S.-Y., Jung, J., Lim, Y., Ahn, J.-H., Kim, S.-I., and Hur, H.-G. (2003), 'Cis-2’, 3’-Dihydrodiol production on flavone B-Ring by Biphenyl Dioxygenase from *Pseudomonas pseudoalcaligenes* KF707 expressed in *Escherichia coli*', *Antonie van Leeuwenhoek,* 84 (4), 261-68.
Kim, Y. M., Bergonia, H. A., Muller, C., Pitt, B. R., Watkins, W. D., and Lancaster, J. R., Jr. (1995), 'Nitric oxide and intracellular heme', *Adv Pharmacol,* 34, 277-91.
Kimani, S. W., Agarkar, V. B., and Cowan, D. A. (2007), 'Structure of an aliphatic amidase from *Geobacillus pallidus* RAPc8', *Acta Crystallogr., Sect D: Biol. Crystallogr.,* 63, 1048-58.
Kinfe, H. H., Chhiba, V. P., Frederick, J., Mathiba, K., and Brady, D. (2008), 'Application of stereoselective biocatalysts for the enantiomeric resolution of beta-hydroxynitriles', *J. Biotechnol.,* 136S, S356-S401.
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', *J. Mol. Catal. B: Enzymatic,* 59 (4), 231-36.
Kirby, R. (2011), 'Chromosome diversity and similarity within the Actinomycetales', *FEMS Microbiology Letters,* 319 (1), 1-10.

Klatte, S., Kroppenstedt, R. M., and Rainey, F. A. (1994), '*Rhodococcus opacus* sp. nov., an unusually nutritionally versatile *Rhodococcus* sp.', *Syst. Appl. Microbiol.,* 17, 355-60.

Kobayashi, M. and Shimizu, S. (1994), 'Versatile nitrilases: nitrile-hydrolyzing enzymes', *FEMS Microbiol. Lett.,* 120, 217-24.

Kobayashi, M. and Shimizu, S. (1998), 'Metalloenzyme nitrile hydratase: structure, regulation, and application to biotechnology', *Nat. Biotechnol.,* 16, 733-36.

Kobayashi, M., Nagasawa, T., and Yamada, H. (1989), 'Nitrilase of *Rhodococcus rhodochrous* J1: Purification and characterization', *Eur. J. Biochem.,* 182, 349-56.

Kobayashi, M., Nagasawa, T., and Yamada, H. (1992a), 'Enzymatic synthesis of acrylamide: a success story not yet over', *TRENDS Biotechnol.,* 10, 402-08.

Kobayashi, M., Goda, M., and Shimizu, S. (1998), 'Nitrilase Catalyzes Amide Hydrolysis as Well as Nitrile Hydrolysis', *Biochem. Biophys. Res. Comm.,* 253 (3), 662-66.

Kobayashi, M., Yanaka, N., Nagasawa, T., and Yamada, H. (1990a), 'Nitrilase-catalyzed production of pyrazinoic acid, an antimycobacterial agent, from cyanopyrazine by resting cells of Rhodococcus rhodochrous J1', *J Antibiot (Tokyo),* 43 (10), 1316-20.

Kobayashi, M., Yanaka, N., Nagasawa, T., and Yamada, H. (1990b), 'Purification and characterization of a novel nitrilase of *Rhodococcus rhodochrous* K22 that acts on aliphatic nitriles', *J. Bacteriol.,* 172, 4807-15.

Kobayashi, M., Yanaka, N., Nagasawa, T., and Yamada, H. (1992b), 'Primary structure of an aliphatic nitrile-degrading enzyme, aliphatic nitrilase from *Rhodococcus rhodocrous* K22 and expression of its gene and identification of its active site residue', *Biochemistry,* 31, 9000-07.

Kobayashi, M., Izui, H., Nagasawa, T., and Yamada, H. (1993a), 'Nitrilase in biosynthesis of the plant hormone indole-3-acetic acid from indole-3-acetonitrile: cloning of the *Alcaligenes* gene and site directed mutagenesis of cysteine residues', *Proc. Natl. Acad. Sci. USA,* 90, 247-51.

Kobayashi, M., Komeda, H., Yanaka, N., Nagasawa, T., and Yamada, H. (1992c), 'Nitrilase from *Rhodococcus rhodochrous* J1: sequencing and overexpression of the gene and identification of an essential cysteine residue', *J. Biol. Chem.,* 267, 20746-51.

Kobayashi, M., Fujiwara, Y., Goda, M., Komeda, H., and Shimizu, S. (1997), 'Identification of active sites in amidase: Evolutionary relationship between amide bond- and peptide bond-cleaving enzymes', *Proc. Natl. Acad. Sci. USA,* 94, 11986-91.

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', *Biochim. Biophys. Acta,* 1129 (1), 23-33.

Kobayashi, M., Komeda, H., Nagasawa, T., Nishiyama, M., Horinouchi, S., Beppu, T., Yamada, H., and Shimizu, S. (1993b), 'Amidase coupled with low-molecular-mass nitrile hydratase from *Rhodococcus rhodochrous* J1. Sequencing and expression of the gene and purification and characterization of the gene product', *Eur. J. Biochem.,* 217 (1), 327-36.

Kolker, E. and al, e. (2005), 'Global profiling of *Shewanella oneidensis* MR-1: expression of hypothetical genes and improved functional annotations', *Proc. Natl. Acad. Sci. USA,* 102, 2099-104.

Koma, D., Yamanaka, H., Moriyoshi, K., Ohmoto, T., and Sakai, K. (2012), 'Production of Aromatic Compounds by Metabolically Engineered Escherichia coli with an Expanded Shikimate Pathway', *Applied and Environmental Microbiology,* 78 (17), 6203-16.

Komeda, H., Kobayashi, M., and Shimizu, S. (1996a), 'A novel gene cluster including the *Rhodococcus rhodocrous* J1 *nhIBA* genes encoding a low molecular mass nitrile hydratase (L-NHase) induced by its reaction product', *J. Biol. Chem.,* 271, 15796-802.
Komeda, H., Kobayashi, M., and Shimizu, S. (1996b), 'Characterization of the gene cluster of high-molecular-mass nitrile hydratase (H-NHase) induced by its reaction product in *Rhodococcus rhodocrous* J1', *Proc. Natl. Acad. Sci. USA,* 93, 4267-72.
Komeda, H., Hori, Y., Kobayashi, M., and Shimizu, S. (1996c), 'Transcriptional regulation of the *Rhodococcus rhodochrous* J1 *nitA* gene encoding a nitrilase', *Proc. Natl. Acad. Sci. USA,* 93, 10572-77.
Konstantinidis, K. T. and Tiedje, J. M. (2004), 'Trends between gene content and genome size in prokaryotic species with larger genomes', *Proc. Natl. Acad. Sci. USA,* 101, 3160-65.
Koonin, E. V. and Wolf, Y. I. (2008), 'Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world', *Nucleic Acids Res.,* 36, 6688-719.
Kopf, M. A., Bonnet, D., Artaud, I., Petre, D., and Mansuy, D. (1996), 'Key role of alkanoic acids on the spectral properties, activity, and active-site stability of iron-containing nitrile hydratase from *Brevibacterium* R312', *Eur. J. Biochem.,* 240, 239-44.
Koronelli, T. V. (1996), 'Principles and methods for raising the efficiency of biological degradation of hydrocarbons in the environment: review', *Appl. Biochem. Microbiol.,* 32, 519-25.
Kostichka, K., Thomas, S. M., Gibson, K. J., Nagarajan, V., and Cheng, Q. (2001), 'Cloning and Characterization of a Gene Cluster for Cyclododecanone Oxidation in *Rhodococcus ruber* SC1 ', *J. Bacteriol.,* 183 (21), 6478-86.
Kovacs, J. A. (2004), 'Synthetic analogues of cysteinate-ligated non-heme iron and non-corrinoid cobalt enzymes', *Chem. Rev.,* 104, 825-48.
Kreit, J., Lefebvre, G., and Germain, P. (1994), 'Membrane-bound cholesterol oxidase from *Rhodococcus* sp. cells - production and extraction', *J. Biotechnol.,* 33, 271-82.
Kubiak, K. and Nowak, W. (2008a), 'Molecular Dynamics Simulations of the Photoactive Protein Nitrile Hydratase', *Biophys. J.,* 94, 3824-38.
Kubiak, K. and Nowak, W. (2008b), 'Molecular dynamics simulations of the photoactive protein nitrile hydratase', *Biophys J,* 94 (10), 3824-38.
Kumar, S., Tamura, K., and Nei, M. (2004), 'MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment', *Briefings in Bioinformatics,* 5, 150-53.
Kumaran, D., Eswaramoorthy, S., Gerchman, S. E., Kycia, H., Studier, F. W., and Swaminathan, S. (2003), 'Crystal structure of a putative CN hydrolase from yeast', *Proteins,* 52, 283-91.
Kurland, C. and Gallant, J. (1996), 'Errors of heterologous protein expression', *Curr. Opin. Biotechnol.,* 7, 489-93.
Kuznetsova, E., Proudfoot, M., Sanders, S. A., Reinking, J., Savchenko, A., Arrowsmith, C. H., Edwards, A. M., and Yakunin, A. F. (2005), 'Enzyme genomics: Application of general enzymatic screens to discover new enzymes', *FEMS Microbiol. Rev.,* 29, 263-79.
Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V., Borozan, I., Carmel, L., Wolf, Y. I., Mori, H., Savchenko, A., Arrowsmith, C. H., Koonin, E. V., and Edwards, A. M. (2006), 'Genome-wide Analysis of Substrate Specificities of the Escherichia coli Haloacid Dehalogenase-like Phosphatase Family', *J. Biol. Chem.,* 281, 36149-61.

Kwon, N., Chae, J.-C., Choi, K., Yoo, M., Zylstra, G., Kim, Y., Kang, B., and Kim, E. (2008), 'Identification of functionally important amino acids in a novel indigo-producing oxygenase from *Rhodococcus* sp. strain T104', *Applied Microbiology and Biotechnology,* 79 (3), 417-22.

Laemmli, U. K. (1970), 'Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4', *Nature (London),* 227, 680-85.

Langdahl, B. R., Bisp, P., and Ingvorsen, K. (1996), 'Nitrile hydrolysis by *Rhodococcus erythropolis* BL1, an acetonitrile-tolerant strain isolated from a marine sediment', *Microbiology,* 142, 145-54.

Larkin, M. J., Kulakov, L. A., and Allen, C. C. R. (2006), 'Biodegradation by Members of the Genus *Rhodococcus*: Biochemistry, Physiology, and Genetic Adaptation', in Allen I. Laskin, S. S. and Geoffrey, M. G. (eds.), *Advances in Applied Microbiology* (59: Academic Press), 1-29.

Larkin, M. J., Kulakov, L. A., and Allen, C. (2010), 'Genomes and Plasmids in *Rhodococcus*', in Alvarez, H. M. (ed.), *Biology of Rhodococcus* (Microbiology Monographs, 16: Springer Berlin / Heidelberg), 73-90.

LaRonde-LeBlanc, N., Resto, M., and Gerratana, B. (2009), 'Regulation of active site coupling in glutamine-dependent NAD+ synthetase', *Nat Struct Mol Biol,* 16 (4), 421-29.

Laursen, B. S., Steffensen, S., Hedegaard, J., Moreno, J. M., Mortensen, K. K., and Sperling-Petersen, H. U. (2002), 'Structural requirements of the mRNA for intracistronic translation initiation of the enterobacterial *infB* gene', *Gene Cells,* 7, 901-10.

Layh, N., Parratt, J. S., and Willetts, A. (1998), 'Characterization and partial purification of an enantioselective arylacetonitrilase from *Pseudomonas fluorescens* DSM7155', *J. Mol. Catal. B: Enzymatic,* 5, 467-74.

Layh, N., Hirrlinger, B., Stolz, A., and Knackmuss, H.-J. (1997), 'Enrichment strategies for nitrile-hydrolysing bacteria', *Appl. Microbiol. Biotechnol.,* 47, 668-74.

Layh, N., Stoltz, A., Böhme, J., Effenberger, F., and Knackmuss, H.-J. (1994), 'Enantioselective hydrolysis of racemic naproxen nitrile and naproxen amide to S-naproxen by new bacterial isolates', *J. Biotechnol.,* 33, 175-82.

Lechavalier (1989), 'Nocardioform actinomycetes', in Williams, S. T., Sharpe, M. E., and Holt, J. G. (eds.), *Bergey's Manual of Sytematic Bacteriology* (4; Baltimore: Williams and Wilkins).

Lechavalier, M. P. and Lechavalier, H. (1970), 'Chemical composition as criterion in the classification of aerobic actinomycetes', *Int. J. Syst. Bacteriol.,* 20, 435-43.

Lee, S., Kim, H., Kim, S., Park, S., Kim, B., Shuler, M., and Lee, E. (2007), 'Cloning, expression and enantioselective hydrolytic catalysis of a microsomal epoxide hydrolase from a marine fish, *Mugil cephalus*', *Biotechnology Letters,* 29 (2), 237-46.

Lee, S. Y. (1996), 'Bacterial Polyhydroxyalkanoates', *Biotechnol. Bioeng.,* 49, 1-14.

Lespinet, O. and Labedan, B. (2006a), 'ORENZA: a web resource for studying ORphan ENZyme activities', *BMC Bioinformatics,* 7, 436.

Lespinet, O. and Labedan, B. (2006b), 'Orphan enzymes could be an unexplored reservoir of new drug targets', *Drug Discovery Today,* 11, 300-05.

Letek, M., Ocampo-Sosa, A. A., Sanders, M., Fogarty, U., Buckley, T., Leadon, D. P., González, P., Scortti, M., Meijer, W. G., Parkhill, J., Bentley, S., and Vázquez-Boland, J. A. (2008), 'Evolution of the *Rhodococcus equi* vap pathogenicity island seen through comparison of host-associated vapA and vapB virulence plasmids', *J. Bacteriol.,* 190 (17), 5797-805.

Letek, M., González, P., Macarthur, I., Rodríguez, H., Freeman, T. C., Valero-Rello, A., Blanco, M., Buckley, T., Cherevach, I., Fahey, R., Hapeshi, A., Holdstock, J., Leadon, D. P., Navas, J., Ocampo, A., Quail, M. A., Sanders, M., Scortti, M. M., Prescott, J. F., Fogarty, U., Meijer, W. G., Parkhill, J., Bentley, S. D., and Vázquez-Boland, J. A. (2010), 'The genome of a pathogenic *Rhodococcus*: cooptive virulence underpinned by key gene acquisitions', *PLoS Genet.,* 6 (9), e1001145.

Li, C., Schwabe, J. W., Banayo, E., and Evans, R. M. (1997), 'Coexpression of nuclea receptor partners increases their solubility and biological activities', *Proc. Natl. Acad. Sci. USA,* 94, 2278-83.

Li, C., Liu, Q., Song, X., Ding, D., Ji, A., and Qu, Y. (2003), 'Epoxide hydrolase-catalyzed resolution of ethyl 3-phenylglycidate using whole cells of *Pseudomonas* sp.', *Biotechnol. Lett.,* 25 (24), 2113-16.

Li, D., Cheng, S., Wei, D., Ren, Y., and Zhang, D. (2007), 'Production of enantiomerically pure (S)-beta-phenylalanine and (R)-beta-phenylalanine by penicillin G acylase from Escherichia coli in aqueous medium', *Biotechnol. Lett.,* 29 (12), 1825-30.

Li, S., Chaulagain, M. R., Knauff, A. R., Podust, L. M., Montgomery, J., and Sherman, D. H. (2009), 'Selective oxidation of carbolide C–H bonds by an engineered macrolide P450 mono-oxygenase', *Proceedings of the National Academy of Sciences,* 106 (44), 18463-68.

Liang, P., Labedan, B., and Riley, M. (2002), 'Physiological genomics of *Escherichia coli* protein families', *Physiol. Genomics,* 9, 15-26.

Liese, A., Seelbach, K., and Wandrey, C. (2000), *Industrial Biotransformations* (Weinheim: Wiley-Vch).

Liljeblad, A. and Kanerva, L. T. (2006), 'Biocatalysis as a profound tool in the preparation of highly enantiopure ß-amino acids', *Tetrahedron,* 62 (25), 5831-54.

Lin, Z., Johnson, L. C., Weissbach, H., Brot, N., Lively, M. O., and Lowther, W. T. (2007), 'Free methionine-(R)-sulfoxide reductase from *Escherichia coli* reveals a new GAF domain function', *Proc. Natl. Acad. Sci. USA,* 104 (23), 9597-602.

Liu, J., Yu, H., and Shen, Z. (2008), 'Insights into thermal stability of thermophilic nitrile hydratases by molecular dynamics simulation', *J Mol Graph Model,* 27 (4), 529-35.

Liu, M. and Sibi, M. P. (2002), 'Recent advances in the stereoselective synthesis of [beta]-amino acids', *Tetrahedron,* 58 (40), 7991-8035.

Lu, J., Zheng, Y., Yamagishi, H., Odaka, M., Tsujimura, M., Maeda, M., and Endo, I. (2003), 'Motif CXCC in nitrile hydratase activator is critical for NHase biogenesis in vivo', *FEBS Lett.,* 553 (3), 391-96.

Lu, W., Li, L., Chen, M., Zhou, Z., Zhang, W., Ping, S., Yan, Y., Wang, J., and Lin, M. (2013), 'Genome-wide transcriptional responses of Escherichia coli to glyphosate, a potent inhibitor of the shikimate pathway enzyme 5-enolpyruvylshikimate-3-phosphate synthase', *Molecular BioSystems,* 9 (3), 522-30.

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

Lugo-Mas, P., Dey, A., Xu, L. K., Davin, S. D., Benedict, J., Kaminsky, W., Hodgson, K. O., Hedman, B., Solomon, E. I., and Kovacs, J. A. (2001), 'How does single oxygen atom addition affect the properties of an Fe-nitrile hydratase analogue? The compensatory role of the unmodified thiolate', *J. Am. Chem. Soc.,* 128, 11211-21.

Lundgren, S., Lohkamp, B., Andersen, B., Piškur, J., and Dobritzsch, D. (2008), 'The Crystal Structure of β-Alanine Synthase from *Drosophila melanogaster* Reveals a Homooctameric Helical Turn-Like Assembly', *J. Mol. Biol.,* 377, 1544-59.

Luo, H., Fan, L., Chang, Y., Ma, J., Yu, H., and Shen, Z. (2008), 'Gene Cloning, Overexpression, and Characterization of the Nitrilase from *Rhodococcus rhodochrous* tg1-A6 in *E. coli*', *Appl. Biochem. Biotechnol.*

Ma, Y., Yu, H., Pan, W., Liu, C., Zhang, S., and Shen, Z. (2010), 'Identification of nitrile hydratase-producing Rhodococcus ruber TH and characterization of an amiE-negative mutant', *Bioresour Technol,* 101 (1), 285-91.

Maddrell, S. J., Turner, N. J., Kerridge, A., Willetts, A. J., and Crosby, J. (1996), 'Nitrile hydratase enzymes in organic synthesis: enantioselective synthesis of the lactone moiety of the mevinic acids', *Tetrahedron Lett.,* 37, 6001-04.

Maeda, M., Chung, S. Y., Song, E., and Kudo, T. (1995), 'Multiple genes encoding 2,3-dihydroxybiphenyl 1,2-dioxygenase in the Gram-positive polychlorinated biphenyl-degrading bacterium *Rhodococcus erythropolis* TA421, Isolated from a termite ecosystem', *Appl. Environ. Microbiol.,* 61, 549-55.

Maier-Greiner, U. H., Obermaier-Skrobranek, B. M. M., Estermaier, L. M., Kammerloher, W., Freund, C., Wülfing, C., Burkert, U. I., Matern, D. H., Breuer, M., Eulitz, M., Küfrevioglu, Ö. I., and Hartmann, G. R. (1991), 'Isolation and properties of a nitrile hydratase from the soil fungus *Myrothecium verrucaria* that is highly specific for the fertilizer cyanamide and cloning of its gene', *Proc. Natl. Acad. Sci. USA,* 88, 4260-64.

Mandelbaum, R. T., Sadowsky, M. J., and Wackett, L. P. (2008), 'Microbial Degradation of s-Triazine Herbicides', in Homer, M. L., Janis, E. M., and Orvin, C. B. (eds.), *The Triazine Herbicides* (San Diego: Elsevier), 301-28.

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Šali, A. (2000), 'Comparative protein structure modeling of genes and genomes', *Annu. Rev. Biophys. Biomol. Struct.,* 29 (1), 291-325.

Martínková, L. and Kren, V. (2002), 'Nitrile- and Amide-converting Microbial Enzymes: Stereo-, Regio- and Chemoselectivity', *Biocatal. Biotrans.,* 20 (2), 73-93.

Martínková, L. and Kren, V. (2010), 'Biotransformations with nitrilases', *Curr. Opin. Microbiol.,* 14, 130-37.

Martínková, L., Vejvoda, V., and Křen, V. (2008), 'Selection and screening for enzymes of nitrile metabolism', *J. Biotechnol.,* 133 (3), 318-26.

Martínková, L., Uhnáková, B., Pátek, M., Nešvera, J., and Křen, V. (2009), 'Biodegradation potential of the genus *Rhodococcus*', *Environ. Int.,* 35, 162-77.

Martiny, J. B. H. and Field, D. (2005), 'Ecological perspectives on our complete genome collection', *Ecol. Lett.,* 8, 1334-45.

Martzen, M. R., McCraith, S. M., Spinelli, S. L., Torres, F. M., Fields, S., Grayhack, E. J., and Phizicky, E. M. (1999), 'A biochemical genomics approach for identifying genes by the activity of their products', *Science,* 286, 1153-55.

Marwaha, S. S., Puri, M., Bhullar, M. K., and Kothari, R. M. (1994), 'Optimization of parameters for hydrolysis of limonin for debitterring of kinnow mandarine juice by *Rhodococcus fascians*', *Enzyme Microb. Technol.,* 16, 723-25.

Masai, E., Yamada, H., Healy, J. M., Hatta, T., Kimbara, K., Fukuda, M., and Yano, K. (1995), 'Characterization of biphenyl catabolic genes of gram-positive polychlorinated biphenyl degrader *Rhodococcus* sp. strain RHA1', *Appl. Environ. Microbiol.,* 61 (6), 2079-85.

Mathew, C. D., Nagasawa, T., Kobayashi, M., and Yamada, H. (1988), 'Nitrilase-Catalyzed Production of Nicotinic Acid from 3-Cyanopyridine in Rhodococcus rhodochrous J1', *Appl Environ Microbiol,* 54 (4), 1030-2.

Mayer, M. P. (1995), 'A new set of useful cloning and expression vectors derived from pBlueScript', *Gene,* 163, 41-46.

McClelland, M., Jones, R., Patel, Y., and Nelson, M. (1987), 'Restriction endonucleases for pulsed field mapping of bacterial genomes', *Nucleic Acids Res.,* 15, 5985-6005.

McGuffin, L. J. and Jones, D. T. (2003), 'Improvement of the GenTHREADER method for genomic fold recognition', *Bioinformatics,* 19, 874-81.

McIver, A. M., Garikipati, S. V., Bankole, K. S., Gyamerah, M., and Peeples, T. L. (2008), 'Microbial oxidation of naphthalene to cis-1,2-naphthalene dihydrodiol using naphthalene dioxygenase in biphasic media', *Biotechnol. Prog.,* 24 (3), 593-98.

McLeod, M. P., Warren, R. L., Hsiao, W. W., Araki, N., Myhre, M., Fernandes, C., Miyazawa, D., Wong, W., Lillquist, A. L., Wang, D.-X., M., D., Hara, H., Petrescu, A., Morin, R. D., Yang, G., Stott, J. M., Schein, J. E., Shin, H., Smailus, D., Siddiqui, A. S., Marra, M. A., Jones, S., J. M., Holt, R., Brinkman, F. S., Miyauchi, K., Fukuda, M., Davies, J. E., Mohn, W. W., and Eltis, L. D. (2006), 'The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse', *PNAS,* 103 (42), 15582-87.

McNeil, M. M. and Brown, J. M. (1994), 'The medically important aerobic actinomycetes: epidemiology and microbiology', *Clin. Microbiol. Rev.,* 7, 357-417.

McNulty, D. E., Claffee, B. A., Huddleston, M. J., and Kane, J. F. (2003), 'Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*', *Protein Expres. Purif.,* 27, 365-74.

Meth-Cohn, O. and Wang, M.-X. (1997), 'An in-depth study of the biotransformation of nitriles into amides and/or acids using *Rhodococcus rhodochrous* AJ270', *J. Chem. Soc. Perkin Trans.,* 1, 1099-104.

Miethling, R. and Karlson, U. (1996), 'Accelerated mineralization of pentachlorophenol in soil upon inoculation with *Mycobacterium chlorophenolicum* PCP-1 and *Sphingomonas chlorophenolica* RA2', *Appl. Environ. Microbiol.,* 62, 4361-66.

Milmo, S. (2000), 'BASF Plant for R-Mandelic Acid', *Chemical Market Reporter,* 258 (16), 8.

Mirza, I. A., Yachnin, B. J., Wang, S., Grosse, S., Bergeron, H., Imura, A., Iwaki, H., Hasegawa, Y., Lau, P. C., and Berghuis, A. M. (2009), 'Crystal structures of cyclohexanone monooxygenase reveal complex domain movements and a sliding cofactor', *J. Am. Chem. Soc.,* 131 (25), 8848-54.

Mitra, S. and Holz, R. C. (2007), 'Unraveling the Catalytic Mechanism of Nitrile Hydratases', *J. Biol. Chem.,* 282 (10), 7397-404.

Miyaji, M., Ueda, K., Kobayashi, Y., Hata, H., and Kondo, S. (2008), 'Fiber digestion in various segments of the hindgut of horses fed grass hay or silage', *Animal Science Journal,* 79 (3), 339-46.

Miyanaga, A., Fushinobu, S., Ito, K., and Wakagi, T. (2001), 'Crystal struture of cobalt-containing nitrile hydratase', *Biochem. Biophys. Res. Commun.,* 288, 1169-74.

Miyanaga, A., Fushinobu, S., Ito, K., Shoun, H., and Wakagi, T. (2004), 'Mutational and structural analysis of cobalt-containing nitrile hydratase on substrate and metal binding', *Eur. J. Biochem.,* 271, 429-38.

Mueller, P., Egorova, K., Vorgias, C. E., Boutou, E., Trauthwein, H., Verseck, S., and Antranikian, G. (2006), 'Cloning, overexpression, and characterization of a

thermoactive nitrilase from the hyperthermophilic archaeon *Pyrococcus abyssi*', *Protein Expres. Purif.,* 47, 672-81.

Mulbry, W. W. (1994a), 'Purification and Characterization of an Inducible *s*-Triazine Hydrolase from *Rhodococcus corallinus* NRRL B-15444R', *Appl. Environ. Microbiol.,* 60 (2), 613-18.

Mulbry, W. W. (1994b), 'Purification and characterization of an inducible S-triazine hydrolase from *Rhodococcus corralinus* NRRLB-1554R', *Appl. Environ. Microbiol.,* 60, 613-18.

Murad, F. (1994), 'Regulation of cytosolic guanylyl cyclase by nitric oxide: the NO-cyclic GMP signal transduction system', *Adv Pharmacol,* 26, 19-33.

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 Sci.,* 9 (5), 1024-30.

Na, K. S., Kuroda, A., Takiguchi, N., Ikeda, T., Ohtake, H., and Kato, J. (2005), 'Isolation and characterization of benzene-tolerant Rhodococcus opacus strains', *J Biosci Bioeng,* 99 (4), 378-82.

Nagasawa, T. and Yamada, H. (1990), 'Large-scale bioconversion of nitriles into useful amides and acids', in Abramowicz, D. A. (ed.), *Biocatalysis* (New York: Van Nostrand Reinhold), 277-318.

Nagasawa, T., Ryuno, K., and Yamada, H. (1986), 'Nitrile hydratase of *Brevibacterium* sp. R312. Purification and characterization', *Biochem. Biophys. Res. Commun.,* 139, 1305-12.

Nagasawa, T., Takeuchi, K., and Yamada, H. (1988), 'Occurrence of a cobalt-induced and cobalt-containing nitrile hydratase in *Rhodococcus rhodochrous* J1', *Biochem. Biophys. Res. Comm.,* 155, 1008-16.

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', *Eur. J. Biochem.,* 196, 581-89.

Nagasawa, T., Shimizu, S., and Yamada, H. (1993), 'The superiority of the third-generation catalyst, *Rhodococcus rhodochrous* J1 nitrile hydratase, for industrial production of acrylamide', *Appl. Microbiol. Biotechnol.,* 40, 189-95.

Nagasawa, T., Nanba, H., Ryuno, K., Takeuchi, K., and Yamada, H. (1987), 'Nitrile hydratase of *Pseudomonas chlororaphis* B23. Purification and characterization', *Eur. J. Biochem.,* 162, 691-98.

Nagasawa, T., Wieser, M., Nakamura, T., Iwahara, H., Yoshida, T., and Gekko, K. (2000), 'Nitrilase of *Rhodococcus rhodochrous* J1, conversion into the active form by subunit association', *Eur. J. Biochem.,* 267, 138-44.

Nagashima, S., Nakasako, M., Dohmae, N., Tsujimura, M., Takio, K., Odaka, M., Yohda, M., Kamiya, N., and Endo, I. (1998a), 'Novel non-heme iron center of nitrile hydratase with a claw setting of oxygen atoms', *Nat. Struct. Biol.,* 5, 347-51.

Nagashima, S., Nakasako, M., Dohmae, N., Tsujimura, M., Takio, K., Odaka, M., Yohda, M., Kamiya, N., and Endo, I. (1998b), 'Novel non-heme iron center of nitrile hydratase with a claw setting of oxygen atoms', *Nat. Struct. Biol.,* 5 (5), 347-51.

Nakai, T., Hasegawa, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Ueki, T., Nanba, H., Ikenaka, Y., Takahashi, S., Sato, M., and Tsukihara, T. (2000), 'Crystal structure of N-carbamyl-D-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases', *Structure,* 8, 729-37.

Nakao, A., Yoshihama, M., and Kenmochi, N. (2004), 'RPG: the Ribosomal Protein Gene database', *Nucl. Acids Res.,* 32, D168-D70.

Nel, A. J., Tuffin, I. M., Sewell, B. T., and Cowan, D. A. (2011), 'Unique aliphatic amidase from a psychrotrophic and haloalkaliphilic nesterenkonia isolate', *Appl. Environ. Microbiol.,* 77 (11), 3696-702.

Nelson, M. J., Jin, H., Turner, I. M. J., Grove, G., Scarrow, R. C., Brennan, B. A., and Que, L. J. (1991), 'A novel iron-sulfur center in nitrile hydratase from *Brevibacterium* sp.', *J. Am. Chem. Soc.,* 113, 7072-73.

Neu, T. R. (1996), 'Significance of bacterial surface-active compounds in interaction of bacteria with interfaces', *Microbiological Reviews,* 60, 151-66.

Newbury, S. F., Smith, N. H., Robinson, E. C., Hiles, I. D., and Higgins, C. F. (1987), 'Stabilization of translationally active mRNA by prokaryotic REP sequences', *Cell,* 48, 297-310.

Nishiyama, M., Horinouchi, S., Kobayashi, M., Nagasawa, T., Yamada, H., and Beppu, T. (1991), 'Cloning and characterization of genes responsible for metabolism of nitrile compounds from *Pseudomonas chlororaphis* B23', *J. Bacteriol.,* 173 (8), 2465-72.

Noguchi, T., Nojiri, M., Takei, K., Odaka, M., and Kamiya, N. (2003), 'Protonation structures of Cys-sulfinic and Cys-sulfenic acids in the photosensitive nitrile hydratase revealed by Fourier transform infrared spectroscopy', *Biochemistry,* 42, 11642-50.

Noguchi, T., Honda, J., Nagamune, T., Sasabe, H., Inoue, Y., and Endo, I. (1995), 'Photosensitive nitrile hydratase intrinsically possesses nitric oxide bound to the non-heme iron center: evidence by Fourier transform infrared spectroscopy', *FEBS Lett.,* 358, 9-12.

Noguchi, T., Hoshino, M., Tsujimura, M., Odaka, M., Inoue, Y., and Endo, I. (1996), 'Resonance Raman evidence that photodissociation of nitric oxide from the non-heme iron center activates nitrile hydratase from *Rhodococcus* sp. N-771', *Biochemistry,* 35 (51), 16777-81.

Nojiri, M., Kawano, Y., Hashimoto, K., and Kamiya, N. (To be published), 'X-ray snap shots of inhibitor binding process in photo-reactive nitrile hydratase'.

Nojiri, M., Nakayama, H., Odaka, M., Yohda, M., Takio, K., and Endo, I. (2000), 'Cobalt-substituted Fe-type nitrile hydratase of *Rhodococcus* sp. N-771', *FEBS Lett.,* 465 (2-3), 173-77.

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', *J. Biochem.,* 125 (4), 696-704.

Noverson, J. C., Olmstead, M. M., and Mascharak, P. K. (1999), 'Co(III) Complexes with Carboxamido N and Thiolato S Donor Centers: Models for the Active Site of Co-Containing Nitrile Hydratases', *J. Am. Chem. Soc.,* 121, 3553-54.

Novo, C., Tata, R., Clemente, A., and Brown, P. R. (2003), '*Burkholderia* genome analysis reveals new enzymes belonging to the nitrilase superfamily: The amidase of *Burkholderis cepacia* (hospital isolate)', *Int. J. Biol. Macromol.,* 33, 175-82.

Novo, C., Farnaud, S., Tata, R., Clemente, A., and Brown, P. R. (2002), 'Support for a three-dimensional structure predicting a Cys-Glu-Lys catalytic triad for *Pseudomonas aeruginosa* amidase comes from site-directed mutagenesis and mutations altering substrate specificity', *Biochem. J.,* 365, 731-38.

Nowak, W., Ohtsuka, Y., Hasegawa, J., and Nakatsuji, H. (2002), 'Density functional study on geometry and electronic structure of nitrile hydratase active site model', *Int. J. Quantum Chem.,* 90, 1174-87.

O'Reilly, C. and Turner, P. D. (2003), 'The nitrilase family of CN hydrolysing enzymes - a comparative study', *J. Appl. Microbiol.,* 95, 1161-74.

Ochi, K. (1995), 'Phylogenetic analysis of mycolic acid-containing wall-chemotype IV actinomycetes and allied taxa by partial sequencing of ribosomal protein AT-L30', *Int. J. Syst. Bacteriol.,* 45, 653-60.

Odaka, M., Fujii, K., Hoshino, M., Noguchi, T., Tsujimura, M., Nagashima, S., Yohda, M., Nagamune, T., Inoue, Y., and Endo, I. (1997), 'Activity regulation of photoreactive nitrile hydratase by nitric oxide', *J. Am. Chem. Soc.,* 119, 3785-91.

Okada, M., Noguchi, T., Nagashima, T., Yohda, M., Yabuki, S., Hoshino, M., Inoue, Y., and Endo, I. (1996), 'Location of the non-heme iron center on the α-subunit of photoreactive nitrile hydratase from *Rhodococcus* sp. N-771', *Biochem. Biophys. Res. Commun.,* 221, 146-50.

Omar, B. A., Flores, S. C., and McCord, J. M. (1992), 'Superoxide Dismutase: Pharmacological Developments and Applications', *Adv. Pharmacol.,* 23, 109-61.

Orru, R., Dudek, H. M., Martinoli, C., Torres Pazmiño, D. E., Royant, A., Weik, M., Fraaije, M. W., and Mattevi, A. (2011), 'Snapshots of Enzymatic Baeyer-Villiger Catalysis: oxygen activation and intermediate stabilization', *Journal of Biological Chemistry,* 286 (33), 29284-91.

Ottolina, G., Bianchi, S., Belloni, B., Carrea, G., and Danieli, B. (1999), 'First asymmetric oxidation of tertiary amines by cyclohexanone monooxygenase', *Tetrahedron Letters,* 40 (48), 8483-86.

Pace, H. C. and Brenner, C. (2001), 'The nitrilase superfamily: classification, structure and function', *Genome Biology,* 2 (1), 1.1-1.9.

Pace, H. C., Hodawadekar, S. C., Draganescu, A., Huang, J., Bieganowski, P., Pekarsky, Y., Croce, C. M., and Brenner, C. (2000), 'Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers', *Curr. Biol.,* 10, 907-17.

Palani, K., Burley, S. K., and Swaminathan, S. (2009), 'Crystal strucure of a transcriptional regulator, merr family from bacillus cereus', *NCBI deposit, PDB ID 3HHO*.

Panke, S. and Wubbolts, M. (2005), 'Advances in biocatalytic synthesis of pharmaceutical intermediates', *Curr. Opin. Chem. Biol.,* 9 (2), 188-94.

Pao, S. S., Paulsen, I. T., and Saier, M. H. J. (1998), 'Major facilitator superfamily', *Microbiol. Mol. Biol. Rev.,* 62 (1), 1-34.

Parales, R. E., Resnick, S. M., Yu, C.-L., Boyd, D. R., Sharma, N. D., and Gibson, D. T. (2000), 'Regioselectivity and Enantioselectivity of Naphthalene Dioxygenase during Arene *cis*-Dihydroxylation: Control by Phenylalanine 352 in the α Subunit', *J. Bacteriol.,* 182 (19), 5495-504.

Parekh, N. R., Walker, A., Roberts, S. J., and Welch, S. J. (1994), 'Rapid degradation of the triazinone herbicide metamitron by a *Rhodococcus* sp. isolated from treated soil', *J. Appl. Bacteriol.,* 77, 467-75.

Payne, D. J., Wallis, N. G., Gentry, D. R., and Rosenberg, M. (2000), 'The impact of genomics on novel antibacterial targets', *Curr. Opin. Drug Discov. Devel.,* 3 (2), 177-90.

Payne, D. J., Warren, P. V., Holmes, D. J., Ji, Y., and Lonsdale, J. T. (2001), 'Bacterial fatty-acid biosynthesis: a genomics-driven target for antibacterial drug discovery', *Drug Discovery Today,* 6 (10), 537-44.

Payne, M. S., Wu, S., Fallon, R. D., Tudor, G., Stieglitz, B., Turner, I. M., Jr., and Nelson, M. J. (1997), 'A stereoselective cobalt-containing nitrile hydratase', *Biochemistry,* 36 (18), 5447-54.

Pazos, F. and Sternberg, M. J. E. (2004), 'Automated prediction of protein function and detection of functional sites from structure', *Proc. Natl. Acad. Sci. USA,* 101 (41), 14754-59.

Peng, Y., Leung, H., Yiu, S., and Chin, F. (2010), 'IDBA – A Practical Iterative de Bruijn Graph De Novo Assembler', in Berger, B. (ed.), *Research in Computational Molecular Biology* (Lecture Notes in Computer Science, 6044: Springer Berlin / Heidelberg), 426-40.

Peplowski, L., Kubiak, K., and Nowak, W. (2007), 'Insights into catalytic activity of industrial enzyme Co-nitrile hydratase. Docking studies of nitriles and amides', *J. Mol. Model*.

Peplowski, L., Kubiak, K., and Nowak, W. (2008), 'Mechanical aspects of nitrile hydratase enzymatic activity. Steered molecular dynamics simulations of *Pseudonocardia thermophila* JCM 3095', *Chem. Phys. Lett.,* 467, 144-49.

Peters, J., Zelinski, T., Minuth, T., and Kula, M. R. (1993), 'Synthetic applications of the carbonyl reductases isolated from *Candida parapsilosis* and *Rhodococcus erythropolis*', *Tetrahedron: Asymmetry,* 4, 1683-92.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004), 'UCSF Chimera - a visualization system for exploratory research and analysis', *J. Comput. Chem.,* 25, 1605-12.

Phillips, G., Chikwana, V. M., Maxwell, A., El Yacoubi, B., Swairjo, M. A., Iwata-Reuyl, D., and de Crécy-Lagard, V. (2010), 'Discovery and characterization of an amidotransferase involved in the modification of archaeal tRNA', *J. Biol. Chem.,* 285, 12706-13.

Philp, J. C. and Bell, K. S. (1995), 'Applied studies with bacteria of the genus *Rhodococcus* relevant to treatment of land and water contaminated by industrial processes', *Proc. Int. Symp. Indust. Environ*, 105-26.

Phizicky, E. M., Bastiaens, P. I., Zhu, H., Snyder, M., and Fields, S. (2003), 'Protein analysis on a proteomic scale', *Nature,* 422, 208-15.

Pieper, U. and Steinbuchel, A. (1992), 'Identification, cloning and sequence analysis of the poly (3-hydroxyalkanoic acid) synthase gene of the Gram-positive bacterium *Rhodococcus ruber*', *FEMS Microbiol. Lett.,* 96, 73-79.

Piersma, S. R., Nojiri, M., Tsujimura, M., Noguchi, T., Odaka, M., Yohda, M., Inoue, Y., and Endo, I. (2000), 'Arginine 56 mutation in the beta subunit of nitrile hydratase: importance of hydrogen bonding to the non-heme iron center', *J. Inorg. Biochem.,* 80 (3-4), 283-88.

Piotrowski, M. (2008), 'Primary or secondary? Versatile nitrilases in plant metabolism', *Phytochemistry,* 69, 2655-67.

Podar, M., R., E. J., and Richardson, T. H. (2005), 'Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study', *BMC Evolutionary Biology,* 5 (42), 1-13.

Pollard, D. J. and Woodley, J. M. (2007), 'Biocatalysis for pharmaceutical intermediates: the future is now', *Trends in Biotechnology,* 25 (2), 66-73.

Poole, R. K. (2005), 'Nitric oxide and nitrosative stress tolerance in bacteria', *Biochem Soc Trans,* 33 (Pt 1), 176-80.

Pop, M. and Salzberg, S. L. (2008), 'Bioinformatics challenges of new sequencing technology', *Trends Genet.,* 24 (3), 142-49.

Pop, M., Salzberg, S. L., and Shumway, M. (2002), 'Genome Sequence Assembly: Algorithms and Issues', *IEEE Computer,* 35 (7), 47-54.

Pop, M., Phillippy, A., Delcher, A. L., and Salzberg, S. L. (2004), 'Comparative genome assembly', *Brief. Bioinform.,* 5 (3), 237-48.

Popescu, V.-C., Münck, E., Fox, B. G., Sanakis, Y., Cummings, J. G., Turner, I. M. J., and Nelson, M. J. (2001), 'Mössbauer and EPR Studies of the Photoactivation of Nitrile Hydratase', *Biochemistry,* 40, 7984-91.

Prasad, S. and Bhalla, T. C. (2010), 'Nitrile hydratases (NHases): At the interface of academia and industry', *Biotechnology Advances*.

Preiml, M., Hönig, H., and Klempier, N. (2004), 'Biotransformation of β-amino nitriles: the role of the *N*-protecting group', *J. Mol. Biocat. B: Enz,* 29, 115-21.

Prepechalová, I., Martínková, L., Stoltz, A., Ovesná, M., Bezouska, K., Kopecký, J., and Kren, V. (2001), 'Purification and characterization of the enantioselective nitrile hydratase from *Rhodococcus equi* A4', *Appl. Microbiol. Biotechnol.,* 55, 150-56.

Prescott, J. F. (1991), '*Rhodococcus equi*: an animal and human pathogen', *Clin. Microbiol.,* 4, 20-34.

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', *Appl. Microbiol. Biotechnol.,* 65, 168-76.

Pries, F., van der Ploeg, J. R., Dolfing, J., and Janssen, D. B. (1994), 'Degradation of halogenated aliphatic compounds: The role of adaptation', *FEMS Microbiology Reviews,* 15 (2-3), 279-95.

Pucci, M. J. (2006), 'Use of genomics to select antibacterial targets', *Biochemical Pharmacology,* 71 (7), 1066-72.

Quinn, P. J., Carter, M. E., Markey, B., and Carter, G. R. (1994), *Clinical veterinary microbiology* (London: Wolfe Publishing).

Raczynska, J. E., Vorgias, C. E., Antranikian, G., and Rypniewski, W. (2010), 'Crystallographic analysis of a thermoactive nitrilase', *J. Struct. Biol.,* 173 (2), 294-302.

Rahman, A. K., Suginati, N., Hatsu, M., and Takamizawa, K. (2003), 'A role of xylanase, alpha-L-arabinofuranosidase, and xylosidase in xylan degradation', *Can. J. Microbiol.,* 49 (1), 58-64.

Rainey, F. A., Burghardt, J., Kroppenstedt, R. M., Klatte, S., and Stackebrandt, E. (1995), 'Phylogenetic analysis of the genera *Rhodococcus* and *Nocardia* and evidence for the evolutionary origin of the genus *Nocardia* from within the radiation of *Rhodococcus* species', *Microbiology,* 141, 523-28.

Rehdorf, J., Mihovilovic, M. D., and Bornscheuer, U. T. (2010), 'Exploiting the Regioselectivity of Baeyer–Villiger Monooxygenases for the Formation of β-Amino Acids and β-Amino Alcohols', *Angewandte Chemie Int. Ed.,* 49 (26), 4506-08.

Rey, P., Rossi, J.-c., Taillades, J., Gros, G., and Nore, O. (2004), 'Hydrolysis of Nitriles Using an Immobilized Nitrilase: Applications to the Synthesis of Methionine Hydroxy Analogue Derivatives', *Journal of Agricultural and Food Chemistry,* 52 (26), 8155-62.

Richards, G. M. and Mehta, M. P. (2007), 'Motexafin gadolinium in the treatment of brain metastases', *Expert Opinion on Pharmacotherapy,* 8 (3), 351-59.

Roberts, R. J. (2004), 'Identifying protein function - a call for community action', *PLoS Biology,* 2 (3), E42.

Rocap, G. and al, e. (2003), 'Genome divergence in two Prochlorococcus ecotypes refelcts oceanic niche differentiation', *Nature,* 424, 1042-47.

Rodionov, D. A., Dubchak, I. L., Arkin, A. P., Alm, E. J., and Gelfand, M. S. (2005), 'Dissimilatory Metabolism of Nitrogen Oxides in Bacteria: Comparative Reconstruction of Transcriptional Networks', *PLoS Comput Biol,* 1 (5), e55.

Rolf, M. J. and Lim, H. C. (1985), in Moo-Young, M. (ed.), *Comprehensive Biotechnology* (Oxford: Pergamon Press), 165-74.

Rose, M. J., Betterley, N. M., Oliver, A. G., and Mascharak, P. K. (2010), 'Binding of Nitric Oxide to a Synthetic Model of Iron-Containing Nitrile Hydratase (Fe-NHase) and Its Photorelease: Relevance to Photoregulation of Fe-NHase by NO', *Inorganic Chemistry,* 49 (4), 1854-64.

Rosłoniec, K. Z., Wilbrink, M. H., Capyk, J. K., Mohn, W. W., Ostendorf, M., Van Der Geize, R., Dijkhuizen, L., and Eltis, L. D. (2009), 'Cytochrome P450 125 (CYP125) catalyses C26-hydroxylation to initiate sterol side-chain degradation in Rhodococcus jostii RHA1', *Molecular Microbiology,* 74 (5), 1031-43.

Ruimy, R., Boiron, P., Boivin, V., and Christen, R. (1994), 'A phylogeny of the genus *Nocardia* deduced from the analysis of small subunit ribosomal DNA sequences, including transfer of *Nocardia amarae* to the genus *Gordona* as *Gordona amarae* comb. nov.', *FEMS Microbiol. Lett.,* 123, 261-68.

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', *Appl. Microbiol. Biotechnol.,* 85 (5), 1417-25.

Sachpatzidis, A., Dealwis, C., Lubetsky, J. B., Liang, P.-H., Anderson, K. S., and Lolis, E. (1999), 'Crystallographic Studies of Phosphonate-Based α-Reaction Transition-State Analogues Complexed to Tryptophan Synthase', *Biochemistry,* 38 (39), 12665-74.

Sakai, N., Tajika, Y., Yao, M., Watanabe, N., and Tanaka, I. (2004), 'Crystal structure of hypothetical protein PH0642 from *Pyrococcus horikoshii* at 1.6 Å resolution', *Proteins,* 57, 869-73.

Šali, A. and Blundell, T. L. (1993), 'Comparative Protein Modelling by Satisfaction of Spatial Restraints', *Journal of Molecular Biology,* 234 (3), 779-815.

Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., Post, A. F., Hagemann, M., Paulsen, I., and Partensky, F. (2009), 'Ecological genomics of marine picocyanobacteria', *Microbiol. Mol. Biol. Rev.,* 73, 249-99.

Scarrow, R. C., Brennan, B. A., Cummings, J. G., Jin, H., Duong, D. J., Kindt, J. T., and Nelson, M. J. (1996), 'X-ray spectroscopy of nitrile hydratase at pH 7 and 9', *Biochemistry,* 35, 10078-88.

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J., and Schomburg, D. (2011), 'BRENDA, the enzyme information system in 2011', *Nucleic Acids Res.,* 39, D670-D76.

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

Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001), 'Industrial biocatalysis today and tomorrow', *Nature,* 409, 258-68.

Schnoes, A. M., Brown, S. D., Dodevski, I., and Babbitt, P. C. (2009), 'Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies', *PLoS Comput. Biol.,* 5 (12), e1000605.

Scholtz, R., Schmuckle, A., Cook, A. M., and Leisinger, T. (1987), 'Degradation of eighteen 1-monohaloalkanes by *Arthrobacter* sp. strain HA1', *J. Gen. Microbiol.,* 133, 267-74.

Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F., and Schomburg, D. (2002), 'BRENDA: a resource for enzyme data and metabolic information', *Trends in Biochemical Sciences,* 27 (1), 54-56.

Schumacher, J. D. and Fakoussa, R. M. (1999a), 'Degradation of alicyclic molecules by *Rhodococcus ruber* CD4', *Appl. Microbiol. Biotechnol.,* 52 (1), 85-90.

Schumacher, J. D. and Fakoussa, R. M. (1999b), 'Degradation of alicyclic molecules by *Rhodococcus ruber* CD4', *Applied Microbiology and Biotechnology,* 52 (1), 85-90.

Schuppler, M., Mertens, F., Schon, G., and Göbel, U. B. (1995), 'Molecular characterization of nocardioform actinomycetes in activated sludge by 16S rRNA analysis', *Microbiology,* 141, 513-21.

Seto, M., Kimbara, K., Shimura, M., Hatta, T., Fukuda, M., and Yado, K. (1995), 'A novel transformation of polychlorinated biphenyls by *Rhodococcus* sp. strain RHA1', *Appl. Environ. Microbiol.,* 61, 3353-58.

Sewell, B. T., Thuku, R. N., Zhang, X., and Benedik, M. J. (2005), 'Oligomeric structure of nitrilases: effect of mutating interfacial residues on activity', *Ann. N. Y. Acad. Sci.,* 1056, 153-59.

Sewell, B. T., Berman, M. N., Meyers, P. R., Jandhyala, D., and Benedik, M. J. (2003), 'The Cyanide Degrading Nitrilase from *Pseudomonas stutzeri* AK61 Is a Two-Fold Symmetric, 14-Subunit Spiral', *Structure,* 11, 1413-22.

Shaikh, T. R., Gao, H., Baxter, W. T., Asturias, F. J., Boisset, N., Leith, A., and Frank, J. (2008), 'SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs', *Nat Protoc,* 3 (12), 1941-74.

Shao, Z. Q., Seffens, W., Mulbry, W., and Behki, R. M. (1995), 'Cloning and Expression of the *s*-Triazine Hydrolase Gene (*trzA*) from *Rhodococcus corallinus* and Development of *Rhodococcus* Recombinant Strains Capable of Dealkylating and Dechlorinating the Herbicide Atrazine', *J. Bacteriol.,* 177 (20), 5748-55.

Shaw, N. M., Robins, K. T., and Kiener, A. (2003), 'Lonza: 20 Years of Biotransformations', *Adv. Synth. Catal.,* 345 (4), 425-35.

Shaw, N. M., Naughton, A., Robins, K., Tinschert, A., Schmid, E., Hischier, M.-L., Venetz, V., Werlen, J., Zimmerman, T., Brieden, W., de Riedmatten, P., Roduit, J.-P., Zimmerman, B., and Neumüller, R. (2002), 'Selection, Purification, Characterisation, and Cloning of a Novel Heat-Stable Stereo-Specific Amidase from *Klebsiella oxytoca*, and Its Application in the Synthesis of Enantiomerically Pure (*R*)- and (*S*)-3,3,3-Trifluoro-2-hydroxy-2-methylpropionic Acids and (*S*)-3,3,3-Trifluoro-2-hydroxy-2-methylpropionamide', *Organic Process Res. & Dev.,* 6, 497-504.

Shearer, J., Kung, I. Y., Lovell, S., Kaminsky, W., and Kovacs, J. A. (2001), 'Why is there an "inert" metal center in the active site of nitrile hydratase? Reactivity and ligand dissociation from a five-coordinate Co(III) nitrile hydratase model', *J. Am. Chem. Soc.,* 123, 463-68.

Shearer, J., Jackson, H. L., Schweitzer, D., Rittenberg, D. K., Leavy, T. M., Kaminsky, W., Scarrow, R. C., and Kovacs, J. A. (2002), 'The First Example of a Nitrile Hydratase Model Complex that Reversibly Binds Nitriles', *J. Am. Chem. Soc.,* 124, 11417-28.

Singh, R., Sharma, R., Tewari, N., Geetanjali, and Rawat, D. S. (2006), 'Nitrilase and Its Application as a 'Green' Catalyst', *Chemistry & Biodiversity,* 3, 1279-87.

Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I. A., Lesley, S. A., and Godzik, A. (2007), 'XtalPred: a web server for prediction of protein crystallizability', *Bioinformatics Applications Note,* 23 (24), 3403-05.

Smith, C. L., KIco, S. R., and Cantor, C. R. (1988), 'Pulsed-field gel electrophoresis and the technology of large DNA molecules', in Davies, K. E. (ed.), *Genome analysis: a practical approach* (Oxford: IRL Press), 41-72.

Soddell, J. A. and Seviour, R. J. (1995), 'Relationship between temperature and growth of organisms causing *Nocardia* foams in activated sludge plants', *Water Res.,* 29, 1555-58.

Soloshonok, V. A., Fokina, N. A., Rybakova, A. V., Shishkina, I. P., Galushko, S. V., Sorochinsky, A. E., Kukhar, V. P., Savchenko, M. V., and Svedas, V. K. (1995), 'Biocatalytic approach to enantiomerically pure [beta]-amino acids', *Tetrahedron: Asymmetry,* 6 (7), 1601-10.

Solymár, M., Fülöp, F., and Kanerva, L. T. (2002), '*Candida antarctica* lipase A - a powerful catalyst for the resolution of heteroaromatic [beta]-amino esters', *Tetrahedron: Asymmetry,* 13 (21), 2383-88.

Sone, H., Nemoto, T., Ishiwata, H., Ojika, M., and Yamada, K. (1993), 'Isolation, structure, and synthesis of dolastatin D, a cytotoxic cyclic depsipeptide from the sea gare *Dolabella auricularia*', *Tetrahedron Lett.,* 34 (52), 8449-52.

Song, L., Wang, M., Shi, J., Xue, Z., Wang, M.-X., and Qian, S. (2007), 'High resolution X-ray molecular structure of the nitrile hydratase from Rhodococcus erythropolis AJ270 reveals posttranslational oxidation of two cysteines into sulfinic acids and a novel biocatalytic nitrile hydration mechanism', *Biochem. Biophys. Res. Commun.,* 362, 319-24.

Sørensen, H. P. and Mortensen, K. K. (2005), 'Advanced genetic strategies for recombinant protein expression in *Escherichia coli*', *J. Biotechnol.,* 115, 113-28.

Sørensen, H. P., Spreling-Petersen, H. U., and Mortensen, K. K. (2003), 'Production of recombinant thermostable proteins expressed in *Escherichia coli*: completion of protein synthesis is the bottleneck', *J. Chromatogr. B,* 786, 207-14.

Sorkhoh, N. A., Alhasan, R. H., Khanafer, M., and Radwan, S. S. (1995), 'Establishment of oil-degrading bacteria associated with cyanobacteria in oil-polluted soil', *J. Appl. Bacteriol.,* 78, 194-99.

Sorokin, D. Y., van Pelt, S., Tourova, T. P., and Evtushenko, L. I. (2009), '*Nitriliruptor alkaliphilus* gen. Nov., sp. Nov., a deep-lineage haloalkaliphilic actinobacterium from soda lakes capable of growth on aliphatic nitriles, and proposal of *Nitriliruptoraceae* fam. Nov. And *Nitriliruptorales* ord. Nov.', *Int. J. Syst. Evol. Microbiol.,* 59, 248-53.

Sosedov, O., Baum, S., Bürger, S., Matzer, K., Kiziak, C., and Stolz, A. (2010), 'Construction and Application of Variants of the *Pseudomonas fluorescens* EBC191 Arylacetonitrilase for Increased Production of Acids or Amides', *Appl. Environ. Microbiol.,* 76 (11), 3668-74.

Stackebrandt, E. and Goebel, B. M. (1994), 'Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology', *Int. J. Syst. Bacteriol.,* 44, 846-49.

Staden, R. (1977), 'Sequence data handling by computer', *Nucleic Acids Res.,* 4, 4037-51.

Stalker, D. M., Malyj, L. D., and McBride, K. E. (1988), 'Purification and Properties of a Nitrilase Specific for the Herbicide Bromoxynil and Corresponding Nucleotide Sequence Analysis of the *bxn* Gene', *J. Biol. Chem.,* 263 (13), 6310-14.

Stange, R. R., Jeffares, D., Young, C., Scott, D. B., Eason, J. R., and Jameson, P. E. (1996), 'PCR amplification of the *Fas*-1 gene for the detection of virulent strains of *Rhodococcus fascians*', *Plant Pathol.,* 45, 407-17.

Stebbins, C. E., Kaelin Jr., W. G., and Pavletich, N. P. (1999), 'Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function', *Science,* 284, 455-61.

Steenkamp, L. and Brady, D. (2003), 'Screening of commercial enzymes for the enantioselective hydrolysis of R,S-naproxen ester', *Enzyme and Microbial Technology,* 32, 472-77.

Steenkamp, L. and Brady, D. (2008), 'Optimisation of stabilised Carboxylesterase NP for enantioselective hydrolysis of naproxen methyl ester', *Process Biochemistry,* 43, 1419-26.

Steer, D. L., Lew, R. A., Perlmutter, P., Smith, A. I., and Aguilar, M.-I. (2002), 'β-Amino acids: Versatile peptidomimetics', *Curr. Med. Chem.,* 9, 811-22.

Steinhuebel, D., Sun, Y., Matsumura, K., Sayo, N., and Saito, T. (2009), 'Direct Asymmetric Reductive Amination', *J. Am. Chem. Soc.,* 131, 11316-17.

Stenstrom, C. M., Jin, H., Major, L. L., Tate, W. P., and Isaksson, L. A. (2001), 'Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*', *Gene,* 263, 273-84.

Stevenson, D. E., Feng, R., and Storer, A. C. (1990), 'Detection of covalent enzyme-substrate complexes of nitrilase by ion-spray mass spectroscopy', *FEBS Lett.,* 277, 112-14.

Stevenson, D. E., Feng, R., Dumas, F., Groleau, D., Mihoc, A., and Storer, A. C. (1992), 'Mechanistic and structural studies on *Rhodococcus* ATCC 39484 nitrilase', *Biotechnol. Appl. Biochem.,* 15, 283-302.

Stewart, J. D. (1998), 'Cyclohexanone Monooxygenase: A Useful Reagent for Asymmetric Baeyer-Villiger Reactions', *Curr. Org. Chem.,* 2, 211-32.

Stewart, J. D. (2006), 'Genomes as Resources for Biocatalysis', in Laskin, A. I., Sariaslani, S., and Gadd, G. M. (eds.), *Advances in Applied Microbiology* (59: Academic Press), 31-52.

Strukul, G., Varagnolo, A., and Pinna, F. (1997), 'New (old) hydroxo complexes of platinum(II) as catalysts for the Baeyer-Villiger oxidation of ketones with hydrogen peroxide', *Journal of Molecular Catalysis A: Chemical,* 117 (1-3), 413-23.

Sugiura, Y., Kuwahara, J., Nagasawa, T., and Yamada, H. (1987), 'Nitrile hydratase: the first non-heme iron enzyme with a typical low spin Fe(III) active centre', *J. Am. Chem. Soc.,* 109, 5848-50.

Summers, D. (1998), 'Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability', *Mol. Microbiol.,* 29, 1137-45.

Sun, S.-Y., Zhang, X., Zhou, Q., Chen, J.-C., and Chen, G.-Q. (2008), 'Microbial production of cis -1,2-dihydroxy-cyclohexa-3,5-diene-1-carboxylate by genetically modified *Pseudomonas putida*', *Applied Microbiology and Biotechnology,* 80 (6), 977-84.

Takai, S., Fukunaga, N., Ochiai, S., Sakai, T., Sasaki, Y., and Tsubaki, S. (1996), 'Isolation of virulent and intermediately virulent *Rhodococcus equi* from soil and sand on parks and yards in Japan', *J. Vet. Med. Sci.,* 58, 669-72.

Takashima, Y. (1995), 'Process for the production of amide compounds using microorganisms', *European Patent Application*, 10.

Takeuchi, M., Hatano, K., Sedlacek, I., and Pacova, Z. (2002), '*Rhodococcus jostii* sp. nov., isolated from a medieval grave', *Int. J. Syst. Evol. Microbiol.,* 52, 409-13.

Takigawa, H., Kubota, H., Sonohara, H., Okuda, M., Tanaka, S., Fujikura, Y., and Ito, S. (1993), 'Novel allylic oxidation of α-cedrene to sec-cedranol by a *Rhodococcus* strain', *Appl. Environ. Microbiol.,* 59, 1336-41.

Tan, S. (2001), 'A modular polycistronic expression system for overexpressing protein complexes in *Escherichia coli*', *Protein Expres. Purif.,* 21, 224-34.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007), 'EMAN2: An extensible image processing suite for electron microscopy', *Struct. Biol.,* 157, 38-46.

Tang, G., Peng, L., Mann, D., Yang, C., Penczek, P. A., Goodyear, G., Baldwin, P. R., Jiang, W., and Ludtke, S. J. (2006), 'EMAN2: Software for Image Analysis and Single Particle Reconstruction', *Microsc. Microanal.,* 12, 388 CD - 89 CD.

Taniguchi, K., Murata, K., Murakami, Y., Takahashi, S., Nakamura, T., Hashimoto, K., Koshino, K., Dohmae, N., Yohda, M., Hirose, T., Maeda, M., and Odaka, M. (2008), 'Novel catalytic activity of nitrile hydratase from *Rhodococcus* sp. N771', *J. Bioeng. Biosci.,* 106, 174-79.

Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997), 'A genomic perspective on protein families', *Science,* 278, 631-37.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. (2003), 'The COG database: an updated version includes eukaryotes', *BMC Bioinformatics,* 4, 41.

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., and al, e. (2005), 'Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome'', *Proc. Natl. Acad. Sci. USA,* 102, 13950-55.

Theil, F. and Ballschuh, S. (1996), 'Chemoenzymatic synthesis of both enantiomers of cispentacin', *Tetrahedron: Asymmetry,* 7 (12), 3565-72.

Thomas, S. M., DiCosimo, R., and Nagarajan, V. (2002), 'Biocatalysis: applications and potentials for the chemical industry', *Trends in Biotechnology,* 20 (6), 238-42.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994), 'CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice', *Nucl. Acids Res.,* 22, 4673-80.

Thuku, R. N., Weber, B. W., Varsani, A., and Sewell, B. T. (2007), 'Post-translational cleavage of recombinantly expressed nitrilase from *Rhodococcus rhodochrous* J1 yields a stable, active helical form', *FEBS J.,* 274 (8), 2099-108.

Thuku, R. N., Brady, D., Benedik, M. J., and Sewell, B. T. (2009), 'Microbial nitrilases: versatile, spiral forming, industrial enzymes', *J. Appl. Microbiol.,* 106, 703-27.

Tipping, P. J. (1995), 'Foaming in activated-sludge processes - an operators overview', *J. Chart. Inst. Water E.,* 9, 281-89.

Tomioka, N., Uchiyama, H., and Yagi, O. (1994), 'Cesium accumulation and growth characteristics of *Rhodococcus erythropolis* CS98 and *Rhodococcus* sp. strain CS402', *Appl. Environ. Microbiol.,* 60, 2227-31.

Torres Pazmiño, D. E., Dudek, H. M., and Fraaije, M. W. (2010), 'Baeyer-Villiger monooxygenases: recent advances and future challenges', *Current Opinion in Chemical Biology,* 14 (2), 138-44.

Tramper, J., Angelino, S. A., Müller, F., and van der Plas, H. C. (1979), 'Kinetics and stability of immobilized chicken liver xanthine dehydrogenase', *Biotechnol. Bioeng.,* 21 (10), 1767-86.

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', *S. Afr. J. Sci.,* 100, 488-90.

Tsujimura, M., Odaka, M., Nakayama, H., Dohmae, N., Koshino, H., Asami, T., Hoshino, M., Takio, K., Yoshida, S., Maeda, M., and Endo, I. (2003), 'A novel inhibitor for Fe-type nitrile hydratase: 2-cyano-2-propyl hydroperoxide', *J. Am. Chem. Soc.,* 125 (38), 11532-38.

Tucker, T.-A., Crow, S., Jr., and Pierce, G. (2012), 'Effect of growth media on cell envelope composition and nitrile hydratase stability in Rhodococcus rhodochrous strain DAP 96253', *Journal of Industrial Microbiology & Biotechnology,* 39 (11), 1577-85.

Tyler, L. A., Noverson, J. C., Olmstead, M. M., and Mascharak, P. K. (2003), 'Modulation of the pK(a) of metal-bound water via oxidation of thiolato sulfur in model complexes of Co(III) containing nitrile hydratase: insight into possible effect of cysteine oxidation in Co-nitrile hydratase', *Inorg. Chem.,* 42, 5751-61.

Van Beilen, J. B., Smits, T. H. M., Whyte, L. G., Schorcht, S., Röthlisberger, M., Plaggemeier, T., Engesser, K.-H., and Witholt, B. (2002), 'Alkane hydroxylase homologues in Gram-positive strains', *Environmental Microbiology,* 4 (11), 676-82.

van den Berg, B., Ellis, R. J., and Dobson, C. M. (1999), 'Effects of macromolecular crowding on protein folding and aggregation', *EMBO J.,* 18, 6927-833.

Van der Geize, R. and Dijkhuizen, L. (2004), 'Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications', *Curr. Opin. Microbiol.,* 7, 255-61.

Van Domselaar, G. H., Stothard, P., Shrivastava, S., Cruz, J. A., Guo, A., Dong, X., Lu, P., Szafron, D., Greiner, R., and Wishart, D. S. (2005), 'BASys: a web server for automated bacterial genome annotation', *Nuc. Acids Res.,* 33, W455-59.

van Nimwegen, E. (2003), 'Scaling Laws in the functional content of genomes', *Trends Genet.,* 19, 479-84.

van Pée, K.-H. and Unversucht, S. (2003), 'Biological dehalogenation and halogenation reactions', *Chemosphere,* 52 (2), 299-312.

van Vliet, A. H., Stoof, J., Poppelaars, S. W., Bereswill, S., Homuth, G., Kist, M., Kuipers, E. J., and Kusters, J. G. (2003), 'Differential regulation of amidase- and formamidase-mediated ammonia production by the Helicobacter pylori fur repressor', *J Biol Chem,* 278 (11), 9052-7.

Van Wyk, J. C., Tastan Bishop, A. O., Cowan, D. A., Sayed, M. F.-R., Tsekoa, T. L., and Sewell, B. T. (2009), 'The high-resolution structure of a Cobalt containing nitrile hydratase', *PDB code 3HHT,* Protein Data Bank.

Várkonyi, T. and Kempler, P. (2008), 'Diabetic neuropathy: new strategies for treatment', *Diabetes, Obesity and Metabolism,* 10 (2), 99-108.

Vejvoda, V., Sveda, O., Kaplan, O., Prikrylova, V., Elisakova, V., Himl, M., Kubac, D., Pelantova, H., Kuzma, M., Kren, V., and Martinkova, L. (2007), 'Biotransformation of heterocyclic

dinitriles by Rhodococcus erythropolis and fungal nitrilases', *Biotechnol Lett,* 29 (7), 1119-24.

Vejvoda, V., Kaplan, O., Bezouška, K., Pompach, P., Šulc, M., Cantarella, M., Benada, O., Uhnáková, B., Rinágelová, A., Lutz-Wahl, S., Fischer, L., Křen, V., and Martínková, L. (2008), 'Purification and characterization of a nitrilase from *Fusarium solani* O1', *J. Mol. Cat. B: Enzymatic,* 50 (2–4), 99-106.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D. Y., Paulsen, I., Nelson, K. E., Nelson, W., and al, e. (2004), 'Environmental genome shotgun sequencing of the Sargasso Sea', *Science,* 304, 66-74.

Vereecke, D., Messens, E., Klarskov, DeBruyn, A., VanMontagu, M., and Goethals, K. (1997), 'Patterns of phenolic compounds in leafy galls of tobacco', *Planta,* 201, 342-48.

Veselá, A. B., Franc, M., Pelantová, H., Kubáč, D., Vejvoda, V., Šulc, M., Bhalla, T. C., Macková, M., Lovecká, P., Janů, P., Demnerová, K., and Martínková, L. (2010), 'Hydrolysis of benzonitrile herbicides by soil actinobacteria and metabolite toxocity', *Biodegradation,* 21, 761-70.

Villaverde, A. and Carrio, M. M. (2003), 'Protein aggregation in recombinant bacteria: biological role of inclusion bodies', *Biotechnol. Lett.,* 25, 1385-95.

Wackett, L., Sadowski, M., Martinez, B., and Shapir, N. (2002), 'Biodegradation of atrazine and related s-triazine compounds: from enzymes to field studies', *Appl. Microbiol. Biotechnol.,* 58 (1), 39-45.

Walmsley, A. R., Barrett, M. P., Bringaud, F., and Gould, G. W. (1998), 'Sugar transporters from bacteria, parasites and mammals: structure-activity relationships', *Trends Biochem Sci,* 23 (12), 476-81.

Wang, P. and Krawiec, S. (1994), 'Desulfurization of dibenzothiopene of 2-hydroxybiphenyl by some newly isolated bacterial strains', *Arch. Microbiol.,* 161, 266-71.

Wang, W.-C., Hsu, W.-H., Chien, F.-T., and Chen, C.-Y. (2001), 'Crystal structure and site-directed mutagenesis studies of N-carbamoyl-D-amino acid amidohydrolase from *Agrobacterium radiobacter* reveals a homotetramer and insight into a catalytic cleft', *J. Mol. Biol.,* 306, 251-61.

Warhurst, A. W., Clarke, K. F., Hill, R. A., Holt, R. A., and Fewson, C. A. (1994), 'Production of catechols and muconic acids from various aromatics by the styrene-degrader *Rhodococcus rhodochrous* NCIMB 13259', *Biotechnol. Lett.,* 16, 513-16.

Warren, A. S., Archuleta, J., Feng, W.-C., and Setubal, J. C. (2010), 'Missing genes in the annotation of prokaryotic genomes', *BMC Bioinformatics,* 11, 131.

Wayne, L. G., Brenner, D. J., Colwell, R. R., and al., e. (1987), 'Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics', *Int. J. Syst. Bacteriol.,* 37, 463-64.

Whiteford, N., Haslam, N., Weber, G., Prügel-Bennet, A., Essex, J. W., Roach, P. L., Bradley, M., and Neylon, C. (2005), 'An analysis of the feasibility of short read sequencing', *Nucleic Acids Res.,* 33 (19), e171.

Whyte, L. G., Smits, T. H., Labbé, D., Witholt, B., Greer, C. W., and van Beilen, J. B. (2002), 'Gene cloning and characterization of multiple alkane hydroxylase systems in *Rhodococcus* strains Q15 and NRRL B-16531', *Appl. Environ. Microbiol.,* 68 (12), 5933-42.

Widersten, M., Gurell, A., and Lindberg, D. (2010), 'Structure-function relationships of epoxide hydrolases and their potential use in biocatalysis', *Biochimica et Biophysica Acta (BBA) - General Subjects,* 1800 (3), 316-26.

Wieser, M. and Nagasawa, T. (2000), 'Stereoselective Nitrile-Converting Enzymes', in Patel, R. N. (ed.), *Stereoselective Biocatalysis*, 467-86.

Wieser, M., Takeuchi, K., Wada, Y., Yamada, H., and Nagasawa, T. (1998), 'Low-molecular-mass nitrile hydratase from *Rhodococcus rhodocrous* J1: purification, substrate specificity and comparison with the analogous high-molecular-mass enzyme', *FEMS Microbiol. Lett.,* 169, 17-22.

Williamson, D. S., Dent, K. C., Weber, B. W., Varsani, A., Frederick, J., Thuku, R. N., Cameron, R. A., van Heerden, J. H., Cowan, D. A., and Sewell, B. T. (2010), 'Structural and biochemical characterization of a nitrilase from the thermophilic bacterium, *Geobacillus pallidus* RAPc8', *Appl. Microbiol. Biotechnol.,* 88, 143-53.

Winter, J. M., Behnken, S., and Hertweck, C. (2011), 'Genomics-inspired discovery of natural products', *Current Opinion in Chemical Biology,* 15 (1), 22-31.

Woods, N. R. and Murrell, J. C. (1990), 'Epoxidation of gaseous alkenes by a *Rhodococcus* sp.', *Biotechnol. Lett.,* 12, 409-14.

Woodward, J. D., Weber, B. W., Scheffer, M. P., Benedik, M. J., Hoenger, A., and Sewell, B. T. (2008), 'Helical structure of unidirectionally shadowed metal replicas of cyanide hydratase from *Gloeocercospora sorghi*', *Journal of Structural Biology,* 161 (2), 111-19.

Wu and Li (2003), 'Biocatalytic asymmetric hydrolysis of (±)-β-hydroxy nitriles by *Rhodococcus* sp. CGMCC 0497', *J. Mol. Cat. B: Enzymatic,* 22, 105-12.

Wu, B., Szymanski, W., Heberling, M. M., Feringa, B. L., and Janssen, D. B. (2011), 'Aminomutases: mechanistic diversity, biotechnological applications and future perspectives', *Trends in Biotechnology,* 29 (7), 352-62.

Wu, S., Fallon, R. D., and Payne, M. S. (1997), 'Over-production of stereoselective nitrile hydratase from *Pseudomonas putida* 5B in *Escherichia coli*: activity requires a novel downstream protein', *Appl. Microbiol. Biotechnol.,* 48, 704-08.

Wu, X., Jornvall, H., Berndt, K. D., and Oppermann, U. (2004), 'Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance', *Biochem. Biophys. Res. Comm.,* 313, 89-96.

Wünsche, K., Schwaneberg, U., Bornscheuer, U. T., and Meyer, H. H. (1996), 'Chemoenzymatic route to β-blocker via 3-hydroxy esters', *Tetrahedron: Asymmetry,* 7, 2017-22.

Wyatt, J. M. and Linton, E. A. (eds.) (1988), *Cyanide Compounds in Biology*, eds. Evered, D. and Harnett, S. (Ciba Foundation Symposium 140, Chichester: Wiley) 32-48.

Xing, F., Hiley, S. L., Hughes, T. R., and Phizicky, E. M. (2004), 'The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs', *J. Biol. Chem.,* 279, 17850-60.

Xiong, F., Shuai, J.-J., Peng, R.-H., Tian, Y.-S., Zhao, W., Yao, Q.-H., and Xiong, A.-S. (2011), 'Expression, purification and functional characterization of a recombinant 2,3-dihydroxybiphenyl-1,2-dioxygenase from *Rhodococcus rhodochrous*', *Molecular Biology Reports,* 38 (7), 4303-08.

Yam, K., van der Geize, R., and Eltis, L. (2010), 'Catabolism of Aromatic Compounds and Steroids by *Rhodococcus*', in Alvarez, H. M. (ed.), *Biology of Rhodococcus* (Microbiology Monographs, 16: Springer Berlin / Heidelberg), 133-69.

Yam, K. C., Okamoto, S., Roberts, J. N., and Eltis, L. D. (2011), 'Adventures in Rhodococcus - from steroids to explosives', *Can. J. Microbiol.,* 57 (3), 155-68.

Yamada, H. and Kobayashi, M. (1996), 'Nitrile hydratase and its application to industrial production of acrylamide', *Biosci. Biotech. Biochem.,* 60, 1391-400.

Yamaki, T., Oikawa, T., Ito, K., and Nakamura, T. (1997), 'Cloning and sequencing of a nitrile hydratase gene from *Pseudonocardia thermophilia* JCM3095', *J. Ferment. Bioeng.,* 85, 474-77.

Yamamoto, Oishi, K., Fujimatsu, I., and Komatsu, K.-I. (1991), 'Production of *R*-(-)mandelic acid from mandelonitrile by *Alcaligenes faecalis* ATCC 8750', *Appl. Environ. Microbiol.,* 57, 3028-32.

Yamamoto, K., Ueno, Y., Otsubo, K., Kawakami, K., and Komatsu, K.-I. (1990), 'Production of S-(+)-ibuprofen from a nitrile compound by *Acinetobacter* sp. strain AK226', *Appl. Environ. Microbiol.,* 56, 3125-29.

Yamamoto, K., Nishimura, M., Kato, D.-i., Takeo, M., and Negoro, S. (2011), 'Identification and characterization of another 4-nitrophenol degradation gene cluster, nps, in *Rhodococcus* sp. strain PN1', *Journal of Bioscience and Bioengineering,* 111 (6), 687-94.

Yang, Y., Gao, P., Liu, Y., Ji, X., Gan, M., Guan, Y., Hao, X., Li, Z., and Xiao, C. (2011), 'A discovery of novel *Mycobacterium tuberculosis* pantothenate synthetase inhibitors based on the molecular mechanism of actinomycin D inhibition', *Bioorganic & Medicinal Chemistry Letters,* 21 (13), 3943-46.

Yu, H.-Y., Shi, Y., Luo, H., Tian, Z.-L., Zhua, Y.-Q., and Shen, Z.-Y. (2006), 'An over expression and high efficient mutation system of a cobalt-containing nitrile hydratase', *J. Mol. Biocat. B: Enz,* 43, 80-85.

Zaitsev, G. M., Uotila, J. S., Tsitko, I. V., Lobanok, A. G., and Salkinoja-Salonen, M. S. (1995), 'Utilization of halogenated benzenes, phenols, and benzoates by Rhodococcus opacus GM-14', *Appl Environ Microbiol,* 61, 4191-201.

Zerbino, D. R. and Birney, E. (2008), 'Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs', *Genome Res.,* 18, 821-29.

Zhang, J.-D., Li, A.-T., Yang, Y., and Xu, J.-H. (2010a), 'Sequence analysis and heterologous expression of a new cytochrome P450 monooxygenase from *Rhodococcus* sp. for asymmetric sulfoxidation', *Appl. Microbiol. Biotechnol.,* 85 (3), 615-24.

Zhang, Z.-J., Xu, J.-H., He, Y.-C., Ouyang, L.-M., and Liu, Y.-Y. (2010b), 'Cloning and biochemical properties of a highly thermostable and enantioselective nitrilase from *Alcaligenes* sp. ECU0401 and its potential for (*R*)-(-)-mandelic acid production', *Bioprocess Biosyst. Eng.,* online edition.

Zheng, G.-W. and Xu, J.-H. (2011), 'New opportunities for biocatalysis: driving the synthesis of chiral chemicals', *Current Opinion in Biotechnology,* 22 (6), 784-92.

Zheng, R.-C., Zheng, Y.-G., and Shen, Y.-C. (2010), 'Acrylamide, Microbial Production by Nitrile Hydratase', *Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation and Cell Technology*.

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), 9638-48.

Zhu, D., Rios, B. E., Rozzell, J. D., and Hua, L. (2005), 'Evaluation of substituent effects on activity and enantioselectivity in the enzymatic reduction of aryl ketones', *Tetrahedron Asymmetry,* 16, 1541-46.

Zumft, W. (1993), 'The biological role of nitric oxide in bacteria', *Archives of Microbiology,* 160 (4), 253-64.

Appendices

7. Appendix

7.1. Chapter 1

7.1.1. Recent patent searches for *Rhodococcus* related biotechnology

A search for "*Rhodococcus* biotechnology" in Google patent search reveals about 740 results, while a search for "*Rhodococcus rhodochrous* biotechnology" produces about 240 relevant results (personal search, June 2011). Including "nitrilase" or "nitrile hydratase" in the "*Rhodococcus rhodochrous* biotechnology" search term produces 98 and 115 results, respectively. Sorting the results by date of patent filing and application shows that most patents involving nitrile hydratase and nitrile from *R. rhodochrous* were filed between 2006 and 2009, with only one patent filed in 2010 for nitrile hydratase.

7.2. Chapter 2

7.2.1. Genome sequencing projects involving South African research groups

Some South African genome projects (not including any human genome sequencing) are highlighted in

Table 7.1. To the best of our knowledge, the sequencing of the full genome of *R. rhodochrous* ATCC BAA-870 is the first microbial genome sequencing project undertaken in South Africa with the aim of having the sequence information for applied biotechnology use.

Table 7.1: Genome sequencing projects involving South African research groups

Genome sequence	Date sequenced	Relevance	Responsible Group
Pantoea ananatis - Proteobacterium - 4.7 Mb - 4349 genes	Sequenced: May 2007 The first plant pathogen genome sequenced in Africa Completed: March 2010	Virulent strain; causal agent of bacterial blight and die-back of young eucalypts in South Africa. It also causes a stalk disease of local maize, and a variety of other symptoms on onions, melons, pineapple, rice and sudan grass in other parts of the world. Also reported to cause infections in humans.	Prof. Teresa Coutinho, Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria and the Scottish Crops Research Institute, UK
Eucalyptus grandis Eucalyptus tree genome 640 Mbp 40,000 genes	Sequenced: October 2010 To be published: early 2012	Important commercial crop that could be used as a biofuel	Prof. Zander Myburg, Department of Genetics, Forestry and Agricultural Biotechnology Institute, University of Pretoria. Collaboration with 130 researchers in 18 nations, including the US Department of Energy's Joint Genome Institute
Mycobacterium tuberculosis KZN 605 and KZN R506 Actinobacteria ~4.4 Mbp each ~4000 genes each	KZN 605 Completed: December 2007 KZN R506 Sequenced: April 2010 (currently a draft sequence)	Extensively-drug resistant tuberculosis ; highly virulent strain endemic to the KwaZulu-Natal region of South Africa	KwaZulu-Natal Research Institute for Tuberculosis and HIV, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal and the Broad Institute, Cambridge
Human immunodeficiency virus 1 (HIV-1) subtype D strain 8916 bp RNA	December 2007	The second full-length non-syncytium-inducing (NSI) subtype D strain described; In South Africa, HIV-1 subtype B and D viruses were responsible for the initial epidemic during the 1980s	Department of Medical Virology, National health Laboratory Service (NHLS) and University of Stellenbosch, Tygerberg Academic Hospital

7.2.2. Cost motivation

In attempting the cloning of nitrile hydratase and amidase from *R. rhodochrous* ATCC BAA-870, the option of sequencing the genome as an alternative for *ad hoc* gene cloning approach had been considered a few times. The cost to benefit ratio had been steadily

decreasing, and by 2009, was beginning to outweigh the cost of designing and ordering of cloning primers, and sequencing of PCR and cloned products. Improvements in short read sequence (SRS) technology has made sequencing considerably cheaper. Four years ago, we were quoted ~ R 390 000 (~47 500 USD) to sequence an estimated 5.5 Mbp genome using Sanger sequencing. The total cost of sequencing the genome during 2010 has amounted to R16 000 (~1 900 USD), including additional mate-paired sequencing to fill in unknown genome sequence gaps. The cost of cloning these genes without prior knowledge of their exact sequences has far outweighed the benefits of obtaining them as functional expressing proteins experimentally. Instead, directed cloning of other enzyme targets and streamlining of the process to obtain these target proteins would be possible.

7.2.3. SRS assembly: Using the Edena programme to assemble short sequence reads

Edena processes the assembly of millions of short reads generated by the Illumina Genome Analyzer, and is based on a traditional overlap layout assembly framework (Pop *et al.* 2002). There are four major steps in the programme:

1. Removal of redundant reads so that the dataset size is more manageable
2. Overlap detection and overlap graph construction
3. Graph cleaning: simplification and ambiguity resolution
4. Contig production.

Edena separates the overlapping and the assembling operations into two distinct running modes. In the overlap step, all exact overlaps between any pair of reads are computed and structured in a graph. The reads are indexed in a prefix array and overlaps are revealed by dichotomic search in the arrays. In the layout step, the graph is analysed to remove transitive and spurious edges. Contigs that can be assembled following unambiguous path in the graph are given as output. Edena allows the production of contigs of several kbp with a near full coverage of the bacterial genome being sequenced. The overlapping mode is invoked by providing the program with a FASTA or FASTQ short reads dataset (-r option). In

case of a FASTQ file, the quality information is not used by Edena. The overlapping mode calculates all the overlaps displaying a minimum length of 20 by default. It builds a transitively reduced overlap graph and saves the result in a binary file suffixed with ".ovl".

The assembling mode is invoked by providing the program with a .ovl file (-e option). This mode performs the assembly and outputs three files. The one suffixed with ".fasta" contains the contigs, the one suffixed with ".cov" provides the depth coverage of each base in the contigs, and finally the one suffixed with ".info" provides a summary of the parameters that were used and some information regarding the assembly run.

Parameters that can be changed when using Edena to assemble contigs are:

- Contig minimum output size
- Overlap length and minimum overlap size
- Sensitivity vs. specificity trade-off
    - Small value: Higher frequency of chance overlaps → causes path branching in graph (sensitivity favoured)
    - Large value: Creates more dead-end (DE) paths, i.e. reads not extended by overlapping reads on one side (specificity favoured).

If Contig Production is run in strict mode, Edena starts generating contig sequences. In non-strict mode, one more cleaning step is performed. Longer overlaps more reliable than shorter ones. Non-strict mode tends to produce longer contigs at the expense of additional misassemblies.

The assembly of reads using Edena was compared in strict and non-strict mode while changing the minimum overlap size (contig minimum output size (100 bp) and overlap length (M) were kept constant). In ‘non-strict’ mode, an additional cleaning up step is implemented, based on the fact that longer overlaps are more reliable than shorter overlaps. Non-strict mode allows for bigger overlap of bp’s, generating fewer contigs but at the expense of misassemblies. In strict assembly mode, the minimum contig overlap size (m) is 22. The minimum overlap length is 20 bp by default, while the programme default for minimum overlap size is 22. In overlap mode, transitive edges and self-overlaps are removed, and the

number of nodes (the number of reads in the non-redundant set) and edges (the number of overlaps) were calculated (Table 7.2). With the minimum overlap length at 20bp, the data obtained in overlap mode is the same regardless of the minimum overlap length (Table 7.3).

Table 7.2: Overall assembly statistics for reads from *R. rhodochrous* BAA-870 using the s_8_sequence.ovl overlay file in overlay mode

Fasta entries	5651222 bp
Sequence discarded	7255 bp
Sequence ok	5643967 bp
Unique sequence	4273289 bp
Reads length	36 bp
No. of nodes	4273289
No. of edges	4167859
Minimum overlap size	20 bp

The second phase of the programme, assembly mode, generates contig size and number data, as well as the N50 value which describes the assembly quality (the N50 contig size is the value X such that at least half of the genome is contained in contigs of size X or larger).

Table 7.3: Comparison of assembly results obtained using Edena in strict and non-strict mode using a 22-bp minimum contig size

	Strict Mode	Non-strict Mode
No. of contigs	4740	3238
Total no. of bases	5743018	5741376
No. of reads assembled	72.4 %	73.3 %
N50	2425 (2.43 kbp)	3485 (3.49 kbp)
Mean contig size (bp)	1211.0	1773.0
Maximum size (bp)	30378	43601
Minimum size (bp)	100	100

An assembly comparison of strict mode and minimum contig size is given in Table 7.4, while Table 7.5 shows assembly results using the non-strict mode of Edena.

Table 7.4: Comparison of assembly results obtained using Edena in strict mode changing the minimum contig size (m)

	Minimum contig size							
	20	21	22	23	24	26	28	30
No. of contigs	6780	5063	4740	5118	6014	9230	13909	16710
Total no. of bases	5841413	5782234	5743018	5703488	5647805	5452515	4988800	3824226
No. of reads assembled	71.9 %	72.3 %	72.4 %	72.3 %	71.9 %	70.2 %	65.7 %	53.4 %
N50	1494	2178	2425	2170	1683	919	473	253
Mean contig size (bp)	861	1142	1211	1114	939	590	358	228
Maximum size (bp)	12001	14423	30378	20604	20604	18389	9304	5996
Minimum size (bp)	100	100	100	100	100	100	100	100

Overlay file s_8_sequence.ovl was generated in overlay mode, and used for data generation in assembly mode.

In strict mode, the minimum contig size that gave the highest N50 (2425) and percentage coverage (72.4 %), and the lowest number of contigs (4740), was 22 bp. The allowed minimum contig size was tested to compare the minimum number of base pairs that gave the optimum in non-strict mode.

In the case where the minimum overlap size in the overlap phase was changed to 16 bp, we now have the ideal data for the sequences with a minimum contig size of 19 for assembly (Table 7.6). The smallest number of contigs obtained in this scenario was 2082, with an N50 value of 5540. In this case the percentage of reads assembled was 73.7 %.

Table 7.5: Comparison of assembly results obtained using Edena in non-strict mode changing the minimum contig size (m)

	Minimum contig size				
	18	20	21	22	23
No. of contigs	2206	2206	2637	3238	4083
Total no. of bases	5787574	5787574	5766569	5741376	5706510
No. of reads assembled	73.6 %	73.6 %	73.5 %	73.3 %	73.0 %
N50	5230	5230	4241	3485	2693
Mean contig size (bp)	2623	2623	2186	1773	1397
Maximum size (bp)	60890	60890	56988	43601	27876
Minimum size (bp)	100	100	100	100	100

(Overlay file s_8_sequence_unstrict.ovl was generated and applied to assembly mode while changing m)

Table 7.6: Comparison of assembly results obtained using Edena in non-strict mode changing the minimum contig size (m) with minimum overlap size in overlap mode (M) constant at 16 bp

	Minimum contig size			
	16	18	19	20
No. of contigs	9416	2554	2082	2206
Total no. of bases	6000123	5837819	5808397	5787574
No. of reads assembled	73.4 %	73.8 %	73.7 %	73.6 %
N50	951	4368	5540	5230
Mean contig size (bp)	637	2285	2789	2623
Maximum size (bp)	6340	18505	30554	60890
Minimum size (bp)	100	100	100	100

Overlay file s_8_sequence_unstrict_M16.ovl was used to generate data in assembly mode.

7.2.1. The nucleotide composition of the *R. rhodochrous* genome

Various graphical statistics show the sliding nucleotide composition of the BAA-870 genome as a function of GC and AT content (Figure 7.1), as well as the direct and reverse strand composition as a fraction of the genome (Figure 7.2 and Figure 7.3).

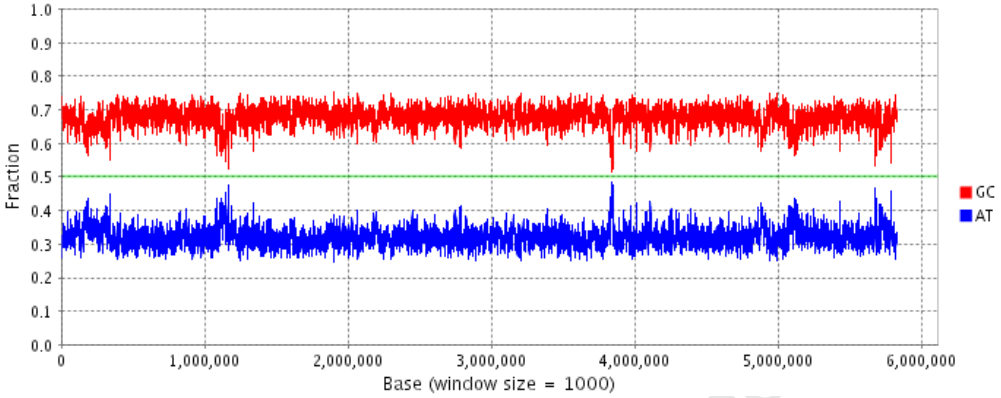


Figure 7.1: Sliding nucleotide composition for *R. rhodochrous* ATCC BAA-870.

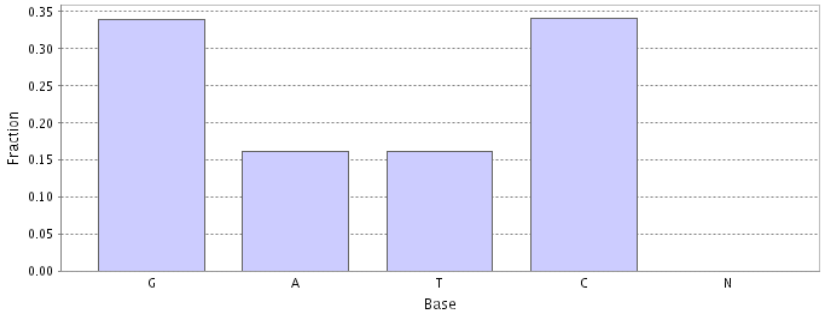


Figure 7.2: Direct strand nucleotide composition for *R. rhodochrous* ATCC BAA-870.

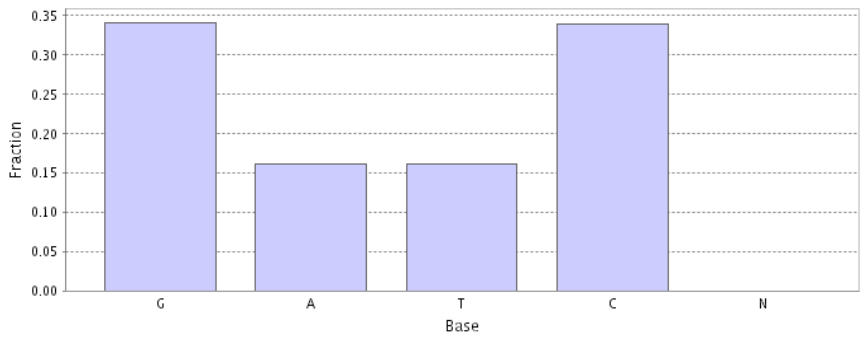


Figure 7.3: Reverse strand nucleotide composition for *R. rhodochrous* ATCC BAA-870.

7.2.2. Periplasmic and secreted protein fate in the *R. rhodochrous* genome

The periplasm contains proteins distinct from those in the cytoplasm and have various functions in cellular processes, including transport, degradation, and motility. The percentage and fate of periplasmic proteins in BAA-870 are shown in Figure 7.4 and Figure 7.5 respectively. Periplasmic proteins would mostly include hydrolytic enzymes such as protease and nucleases, proteins involved in binding of ion, vitamin and sugar molecules, and those involved in chemotaxic responses. Detoxifying proteins such as penicillin binding proteins are presumed to be located mostly in the periplasm.

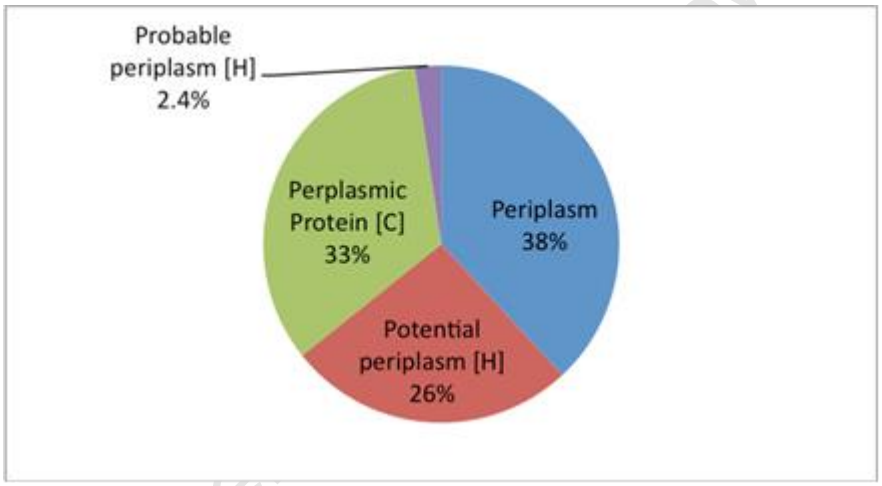


Figure 7.4: Periplasm protein percentage in *R. rhodochrous* ATCC BAA-870.

[H]: Homology to a SwissProt Entry and [C]: homology to a CyberCell entry, as inferred by BASys annotation.

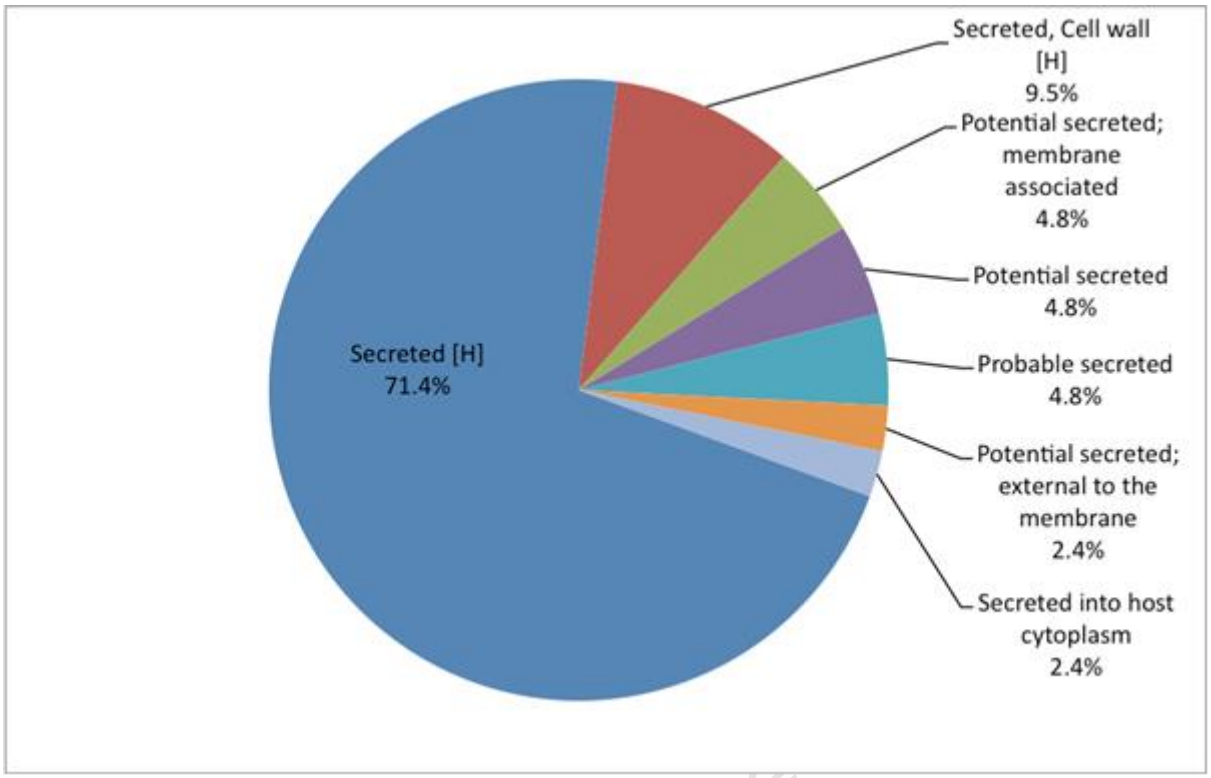


Figure 7.5: The fate of secreted proteins in *R. rhodochrous* ATCC BAA-870.

[H]: Homology to a SwissProt Entry, as inferred by BASys annotation.

7.3. Cloning and expression vectors

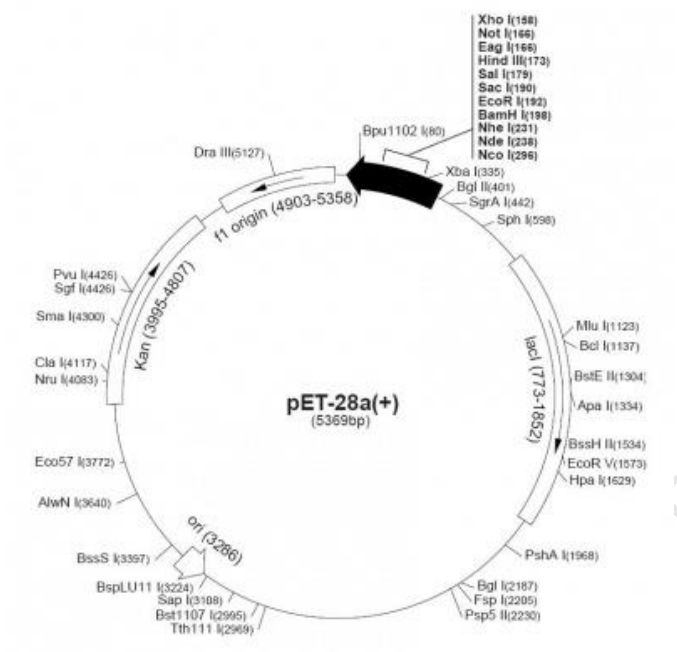


Figure 7.6: Directional cloning vector map, pET-28a(+), showing the *Nde*I and *Hin*dIII restriction sites.

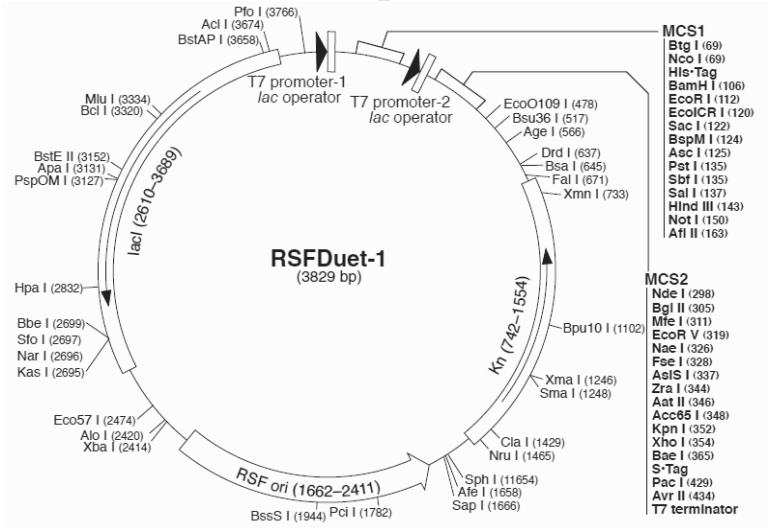


Figure 7.7: Directional cloning vector map, pRSFDuet-1.

The restriction sites *Nde*I and *Pac*I in multiple cloning site2 of the vector were used for the cloning of the alpha-chaperone construct.

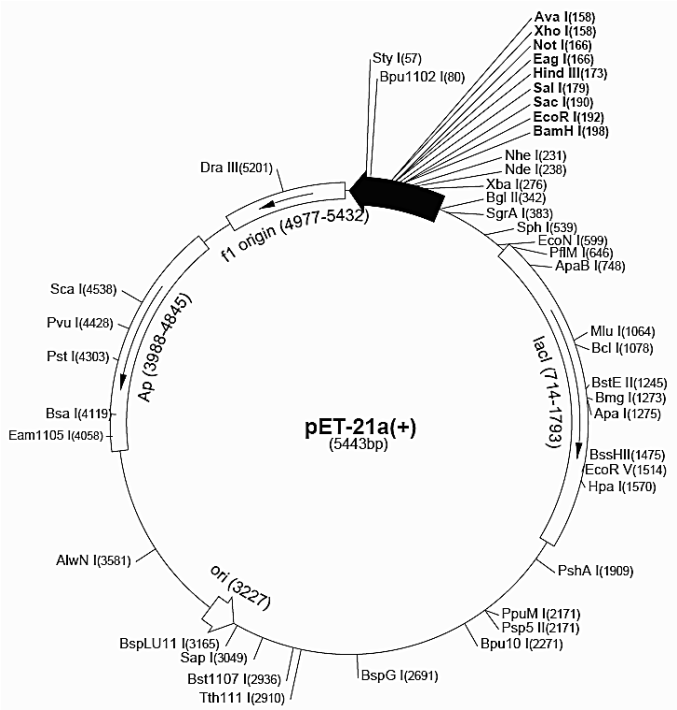


Figure 7.8: Directional cloning vector map, pET-21a(+).

The restriction sites *NdeI* and *XhoI* in the multiple cloning site of the vector were used for the cloning of the beta subunit of nitrile hydratase.

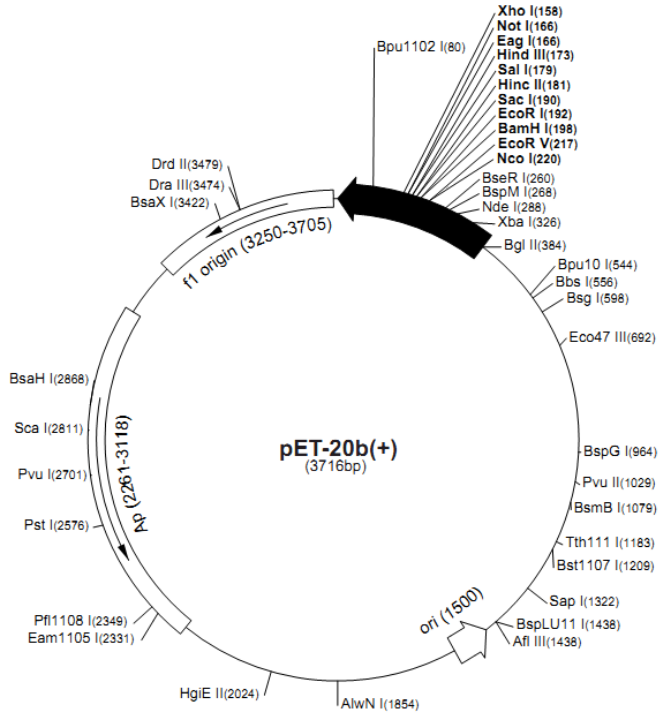


Figure 7.9: Directional cloning vector map, pET-20b(+).

7.4. Chapters 3 and 4

7.4.1.1. Spectrophotometric assay for benzamide formation

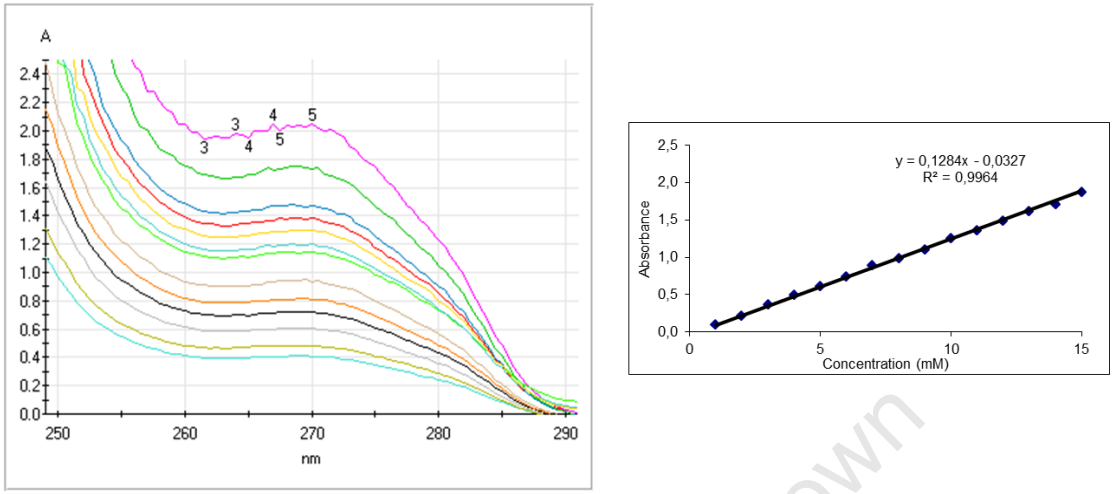


Figure 7.10. Developing the spectrophotometric assay for NHase.

Concentrations of benzonitrile and benzamide were altered in the same sample to find an optimal wavelength to distinguish the two compounds. The absorbance at 252 nm was chosen as the wavelength at which to measure benzamide formation. Standard curves were constructed by measuring increasing concentrations of benzamide and benzonitrile at 252 nm to ensure that absorbance of each remained linear with concentration.

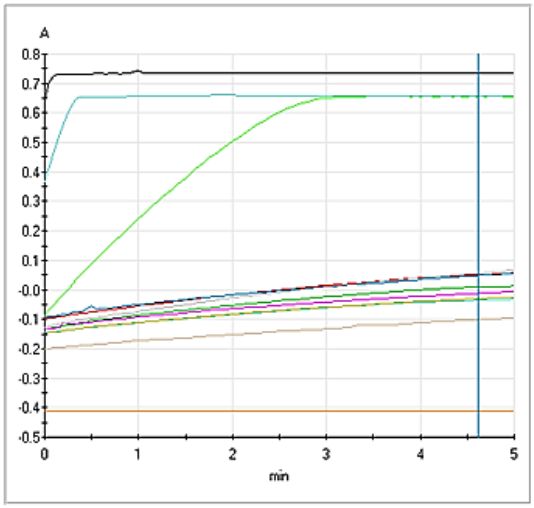


Figure 7.11: Example of NHase activity measurement in Q Sepharose ion exchange fractions using the absorbance assay for benzamide formation.

Fractions containing NHase, eluting between 100 and 400 mM KCl, were tested for benzonitrile conversion by measuring the formation of benzamide spectrophotometrically (diluted samples) at 252 nm. The rate of benzamide formation was recorded over 5 minutes from addition of fraction containing NHase. If the rate of benzamide formation was not linear, samples were diluted appropriately. The orange line with zero gradient is control benzonitrile.

7.4.1.2. Standard curves: Typical examples

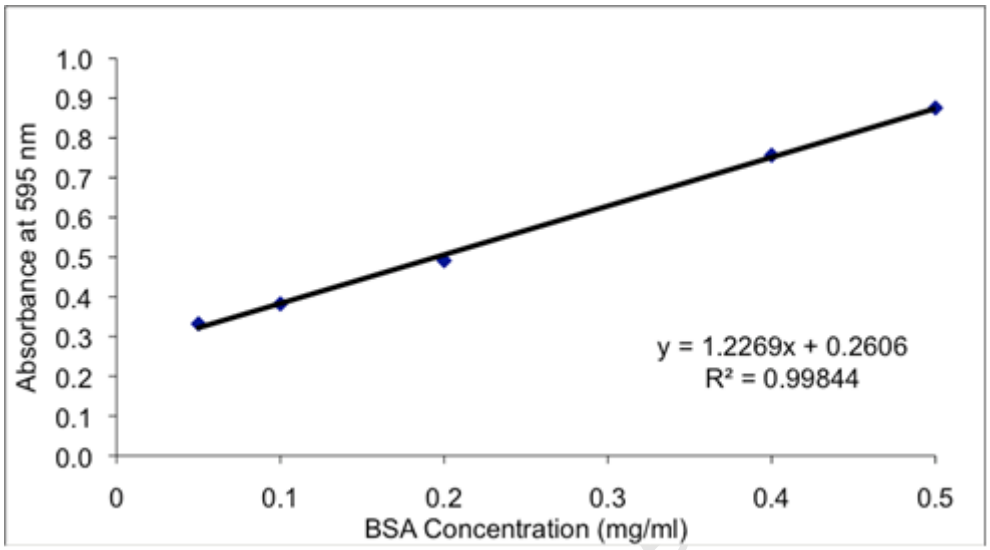


Figure 7.12: BioRad protein concentration standard curve.

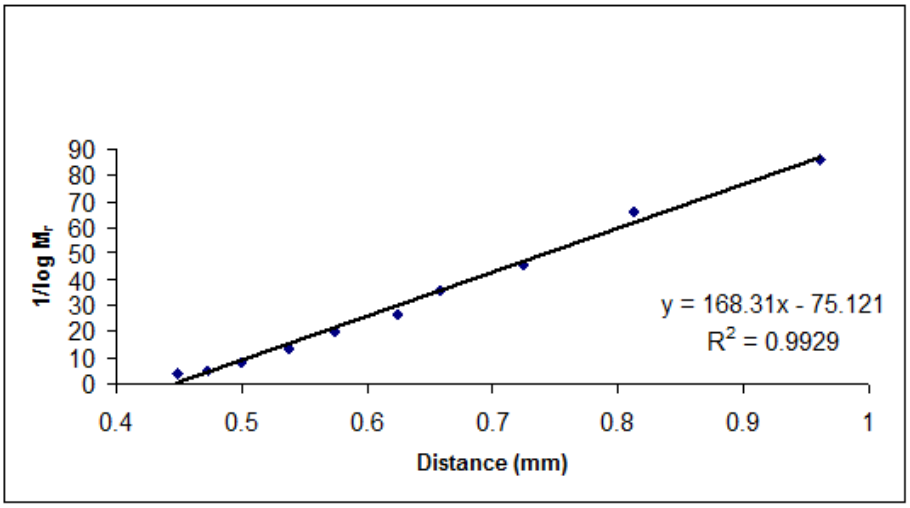


Figure 7.13: Standard curve of markers run on a 12% SDS-PAGE gel.

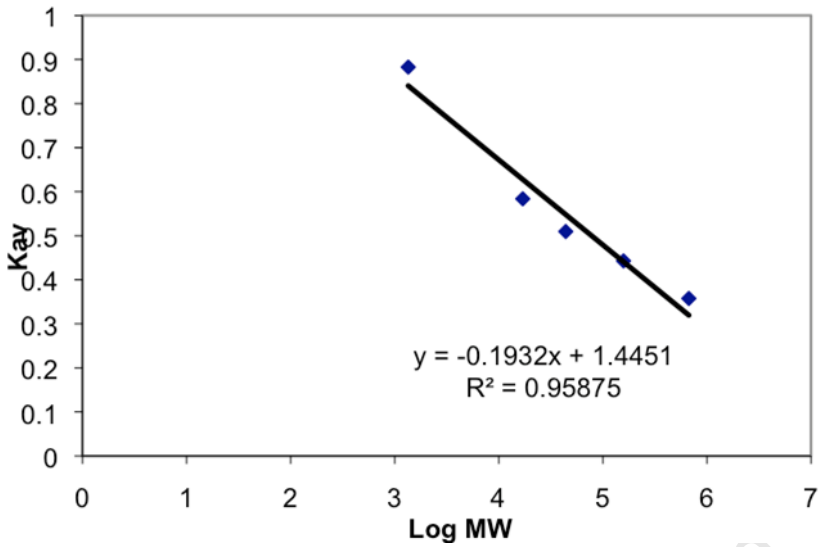


Figure 7.14: Selectivity curve for PWXL 4000G FPLC size exclusion chromatography column.

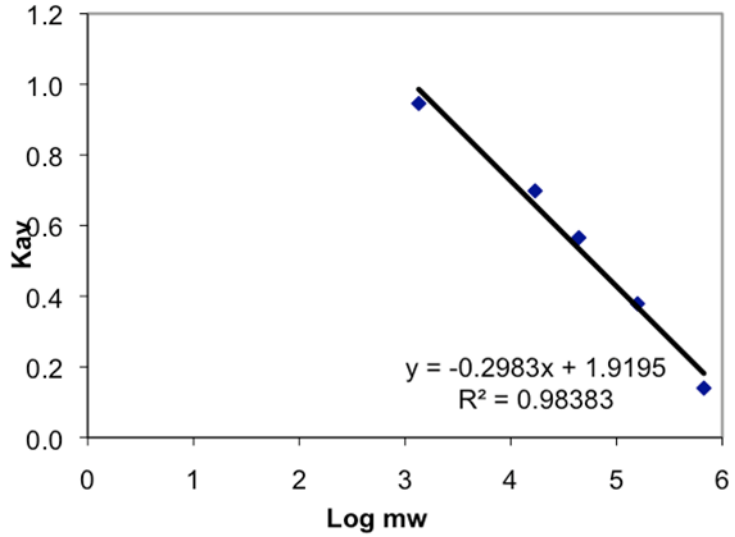


Figure 7.15: Standard curve for S200 size exclusion chromatography column.