The Relationship between Structure and Specificity in the Plant Nitrilases

Jeremy David Woodward

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

In the Department of Molecular and Cell Biology

UNIVERSITY OF CAPE TOWN

August 2011

Supervisor: Prof. B.T. Sewell

For Bettina

I know the meaning of plagiarism and declare that all of the work in this thesis, save for that which is properly acknowledged, is my own:

Signed…….(Jeremy David Woodward)

Acknowledgements

I would like to thank my supervisor Trevor Sewell for his longstanding support and encouragement of my work. Also Markus Piotrowski for suggesting many of the experiments performed here, ideas and helpful discussions, providing access to the majority of the material used and for hosting me for a year in his laboratory in the *Lehrstuhl Für Pflanzenphysiologie* at *Ruhr Universität*, Bochum.

Without the encouragement and support of my parents none of this would have been possible.

Thanks also to my lab-mates who have contributed to this work through helpful discussions: Jason Van Rooyen, Ndoriah Thuku, Tim Frouws, Jean Watermeyer and Serah Kimani. To Brandon Weber for teaching me protein expression and purification (from scratch) and many invaluable troubleshooting sessions. For Mohammed Sayed for helping me not to break too much expensive equipment. For her invaluable contribution to this work, which includes a yearlong collaboration and teaching me molecular biology: Inga Trompetter.

This thesis was supported in part by funds from the National Research Foundation (South Africa) and the Deutsche Forschungsgemeinschaft (Germany).

The Carnegie Corporation of New York is gratefully acknowledged for funding the infrastructure for structural biology in South Africa.

Abstract

Nitrilases (EC 3.5.5.1) catalyse the enantioselective hydrolysis of a variety of organic nitriles to the corresponding amide and/ or carboxylic acid. These reactions are important in the manufacture of fine chemicals, pharmaceutical intermediates, plastics and paints. Industrial uses of nitrilases are limited however, as wild-type enzymes suffer from limitations such as: a lack of control of the acid:amide ratio of products, and limited availability of substrate specificities.

The location of the substrate-binding pocket within the nitrilase helix was determined by three-dimensional helical reconstruction, homology modelling and docking. A strong correlation was found between the relative orientation of nitrilase monomers in the helix (the helical twist) and the overall size of the R-group of the preferred substrate. This effect was confirmed by identifying a single amino acid exchange that almost completely exchanged both the substrate preference of two enzymes as well as their helical twists. While the helical twist accounts for differences in the overall size of the substrate, differences in the chemistry of the substrate are due to amino acid residues lining the substrate pocket. Homology modelling, site-directed mutagenesis and substrate screening led to the identification of three substrate-specifying motifs that altered the substrate specificity of nitrilases when exchanged.

Two closely related plant nitrilases with substantially different acid:amide product ratios were analysed by homology modelling. A tyrosine lying close to the enzyme active site was shifted by a single amino acid insertion in the mainly amide-producing enzyme. Repositioning this residue by deleting the insertion changed the proportion of acid from 17% to 97% of the total. Two isoforms of nitrilase 4 with distinct acid: amide ratios have also been identified; increased amide production is correlated with an arginine to lysine exchange in the substrate-binding pocket. This effect is proposed to be due to a shift in the position of the substrate. A detailed reaction mechanism accounting for these effects is proposed on the basis of a structural analysis of nitrilase homologues. It is suggested that the acid:amide ratio depends on factors that influence the protonation of the amino group of the tetrahedral intermediate. In the first case this occurs by disrupting the hydrogen-bonding network around the amino group and in the second, by shifting the amino group away from the optimal location for protonation relative to the active site residues.

Abbreviations

2PAN	2-Phenylacetonitrile
4HPAN	4-Hydroxyphenylacetonitrile
Ala(CN)	ß-Cyanoalanine
At	*Arabidopsis thaliana* (L.) Heynh
Cr	*Capsella rubella* Reut.
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
E. coli	*Escherichia coli*
His_6	Hexahistidine
HPLC	High performance liquid chromatography
IAN	Indol-3-acetonitrile
kat	Specific activity (mol/s)
LB	Lysogeny broth
Nitrilase	Nitrilase superfamily branch 1 (EC 3.5.5.1)
NIT	Nitrilase
NIT1	Homologues of the nitrilase 1, 2 and 3 from *Arabidopsis thaliana*
NIT4	Homologues of the nitrilase 4 from *Arabidopsis thaliana*
Ps	*Pseudomonas stutzeri*
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions per minute
SD	Standard deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sal	*Sinapis alba* L.
Sar	*Sinapis arvensis* L.
Tris	Tris(hydroxymethyl)aminomethane
3BUT	3-butenenitrile

Table of Contents

0.1 Introduction

0.1.1 Nitrilases

Nitrilases (EC 3.5.5.1) are members of the nitrilase superfamily (Pace and Brenner, 2001) that catalyse the hydrolysis of organic nitriles (R-C≡N) to the corresponding carboxylic acid with the release of ammonia. A varying proportion of amide product is also formed (nitrile hydratase activity) and in some cases this accounts for the majority of total enzyme activity (Fernandes *et al*., 2006). In a few cases, a very small proportion of amidase activity is also measurable (Kobayashi *et al*., 1998a; Piotrowski *et al*., 2001) (Fig. 1).

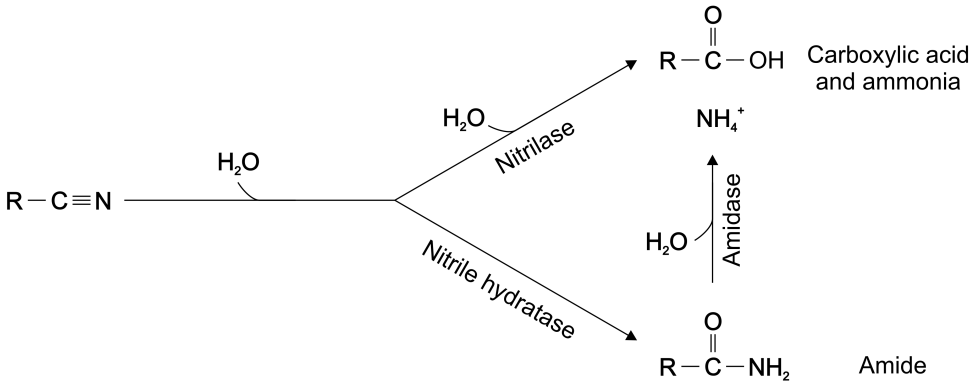


Figure 1. Nitrilase superfamily (Pace and Brenner, 2001) branch 1 (EC 3.5.5.1) catalysed reactions. The nitrilase reaction involves hydrolysis of nitriles to carboxylic acid and ammonia (Pace and Brenner, 2001). The nitrile hydratase reaction leads to the hydrolysis of nitriles with the formation of amide (Stevenson *et al*., 1992). A very low level of amidase activity (hydrolysis of amide to carboxylic acid and ammonia) has also been measured in nitrilases (Kobayashi *et al*., 1998a; Piotrowski *et al*., 2001).

0.1.2 Industrial considerations

Nitrilases are interesting biocatalysts for synthetic chemistry because of their ability to convert easily synthesized nitriles to higher-value amides and acids (Banerjee *et al*., 2006). The majority of industrial nitrile hydrolysis occurs under either highly acidic or basic conditions, consumes large amounts of energy and produces a high proportion of toxic waste (Kukushkin and Pombeiro, 2005). Nitrilases perform nitrile hydrolysis under close to physiological conditions of temperature, pH and ionic strength. Furthermore, nitrilases can provide selective nitrile hydrolysis in the presence of labile functional groups (Singh *et al*., 2006) or even selectively hydrolyse dinitriles (Kobayashi *et al*., 1988; Gagavan *et al*., 1999; Effenberger and Osswald, 2001; Zhu *et al*., 2007). The majority of nitrilases are enantioselective (Martínková

and Kren, 2002). Nitrilases are currently marketed by DuPont, Lonza, BASF and Mitsubishi Rayon (Brady *et al*., 2004). They are used in the production of high-value organic molecules such as: nicotinic acid (an essential human vitamin) (Bhalla *et al*., 1992), pyrazinoic acid (antimycobacterial) (Kobayashi *et al*., 1990), Ibuprofen (non-steroidal anti-inflammatory drug) (Yamamoto *et al*., 1990), (R)-mandelic acid (Yamamoto *et al*., 1991), (S)-phenyl-lactic acid and (R)-3-hydroxy-4-cyano-butric acid, an intermediate in the synthesis of Lipitor® (Diversa Inc.).

Nitrilase reaction parameters need to be improved before nitrilase-catalysed nitrile hydrolysis finds more widespread industrial use (Brady *et al*., 2004). The Verenium Corporation (San Diego, CA) has recently developed nitrilases with increased activity and stability at increased pH and temperature. Four US patent applications to this effect (20100273659, 20100216214, 20100086986, 20100009426) were issued in 2010. Recent studies have also attempted to exert control over the amide:acid ratio to reduce unwanted reaction product and identify determinants of acid – or amide product formation (Sosedov *et al*., 2010; Kiziak and Stoltz, 2009). Activity-based screening (Goddard and Reymond, 2004) and genome mining (DeSantis *et al*., 2003), have enabled the identification of nitrilases with improved specificity spectra and enantiospecificity.

0.1.3 Structural biology of nitrilases

While no atomic-resolution structural data from a nitrilase is available, several members of the nitrilase superfamily (Pace and Brenner, 2001) have been crystallized. These include the N-carbamyl-D-amino acid amidohydrolase from *Agrobacterium sp.* (1erz) Nakai *et al*. (2000) and (1uf4, 1uf5, 1uf7, 1uf8) Hashimoto *et al*. (unpublished). Amidases from *geobacillus pallidus RAPc8* (2plq) (Argarkar *et al*., 2006), *Nesterenkonia sp*. (3hkx) (Nel *et al*., 2011), *Pseudomonas aeruginosa* (2uxy) (Andrade *et al*., 2007) and *Helicobacter pylori* (2dyu, 2dyv, 2e2k, 2e2l) (Hung *et al*., 2007). A putative CN hydrolase from *Saccharomyces cerevisiae* (1f89) (Kumaran *et al*., 2003); a hypothetical protein from *Pyrococcus horikoshii* (1j31) (Sakai *et al.*, 2004); the beta-alanine synthase from *Drosophila melanogaster* (2vhh, 2vhi) (Lundgren *et al*., 2008) and an enzyme that showed very low nitrilase activity, but with no other known activity (Mueller *et al*., 2006) from *Pyrococcus abyssi* (3ivz, 3ki8, 3klc, 3iw3) (Raczynska *et al*., 2011) with 87% identity and 96% similarity to 1j31.

All crystallised superfamily members share a αββα-fold and have an average root mean square deviation (RMSD) of ~1.5 Å (Thuku *et al*., 2009). The arrangement of the active-site cys-glu_1-lys-glu_2 tetrad is highly conserved (Kimani *et al*., 2007), as is the general arrangement of the active-site pocket and some conserved residues. 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, the "A-surface" (Sewell *et al*., 2003) forming a αββα-αββα fold with a characteristic extended shape (Fig. 2). All nitrilase superfamily members investigated so far with significant nitrilase activity form helical – or helix-like homo-oligomers (Sewell *et al*., 2003; Thuku *et al*., 2007; Vejvoda *et al*., 2008; Woodward *et al*., 2008; Dent *et al*., 2009; Williamson *et al*., 2010; Kaplan *et al*., 2011).

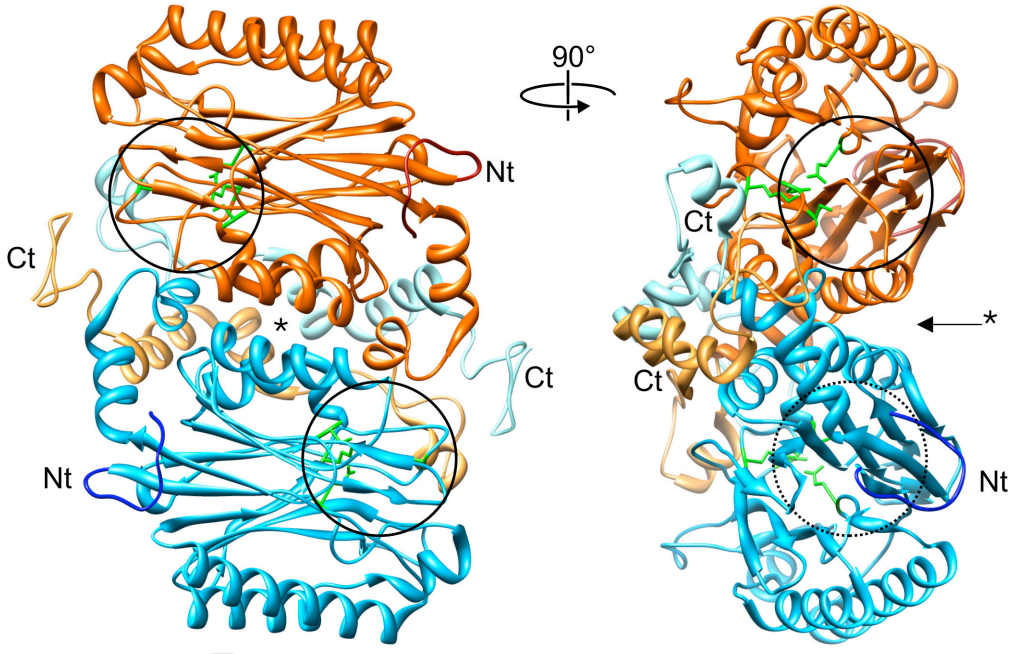


Figure 2. Nitrilase superfamily dimer crystal structure (amidase from *G. pallidus* RAPc8 PDB: 2plq (Argarkar *et al*., 2006)) showing the characteristic extended αββα-αββα fold. Individual monomers making up the dimer (orange, blue) interact across the two-fold (180°) symmetric (*) A-interface. The variable N – and C-termini (Nt, Ct) are coloured darker and lighter respectively. The conserved catalytic residues are shown in green surrounded by the active site pockets (circles).

Three-dimensional maps of entire homo-oligomeric nitrilase complexes have been obtained at low resolution. These maps have been combined with homology models in order to identify those amino acids potentially involved in interface formation (reviewed by Thuku *et al*., 2009). The fundamental unit of oligomerization is the conserved nitrilase superfamily dimer (Sewell *et al*., 2003), which interacts with

adjacent dimers across the "C-interface". After making one complete turn of a left-handed (Woodward *et al*., 2008) spiral/ helix, a third interaction occurs across the helix, this is termed the "D-interface" (Sewell *et al*., 2003) or the "F-interface" Woodward *et al*. (2008) depending on its location. A fourth interaction has been observed in nitrilases that form terminating spirals (*P. stutzeri* cyanide dihydratase (cynD) and *B. pumilus* C1 cynD above pH 6) termed the "E-interface" (Sewell *et al*., 2003) (Fig. 3). A toplogy diagram of a nitrilase superfamily monomer is shown in (Fig. 4) this α-helix and β-sheet numbering is used in the remainder of the thesis.

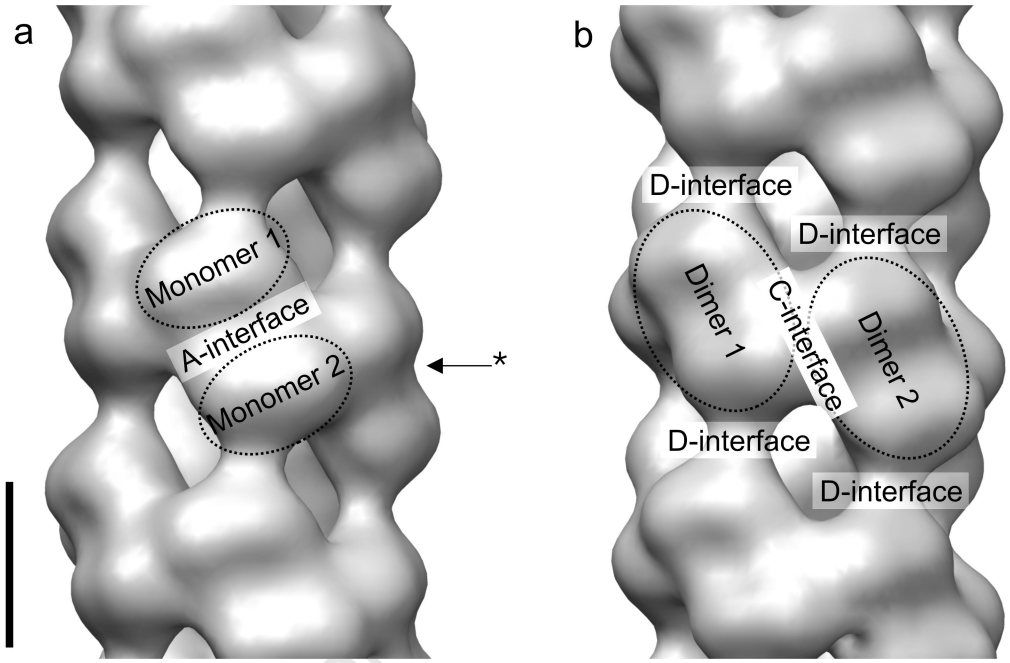


Figure 3. Nitrilase helical oligomerization (nitrilase from *R. rodochrous* J1 (Thuku *et al*., 2007)) (**a**) Monomers (e.g. monomer 1, 2) interact across the two-fold symmetric A-interface to form a dimer, a second A-interface is shown (*) to illustrate the characteristic nitrilase superfamily dimer shape (see Fig. 2). (**b**) The helix is built up of dimers interacting across the C-interface (e.g. dimer 1, 2) to make up a left-handed helix, spiral or turn. After one complete turn, the D-interface makes an interaction; in this case dimer n will interact with dimer n+5. Scalebar: 5 nm.

0.1.4 Physiological role of nitrilases

The physiological role and natural substrates have not been identified for the majority of nitrilases (Thuku *et al*., 2009) although in the majority of bacteria and fungi nitrilase expression is inducible (Banerjee *et al*., 2002) which may indicate a role in nitrile detoxification or utilization (Thuku *et al*., 2009). In plants, nitrilases are involved in cyanide detoxification and nitrogen recycling and the catabolism of cyanogenic glycosides (Piotrowski 2008). Recent results also suggest they may also play a role in glucosinolate catabolism (Janowitz *et al*., 2009).

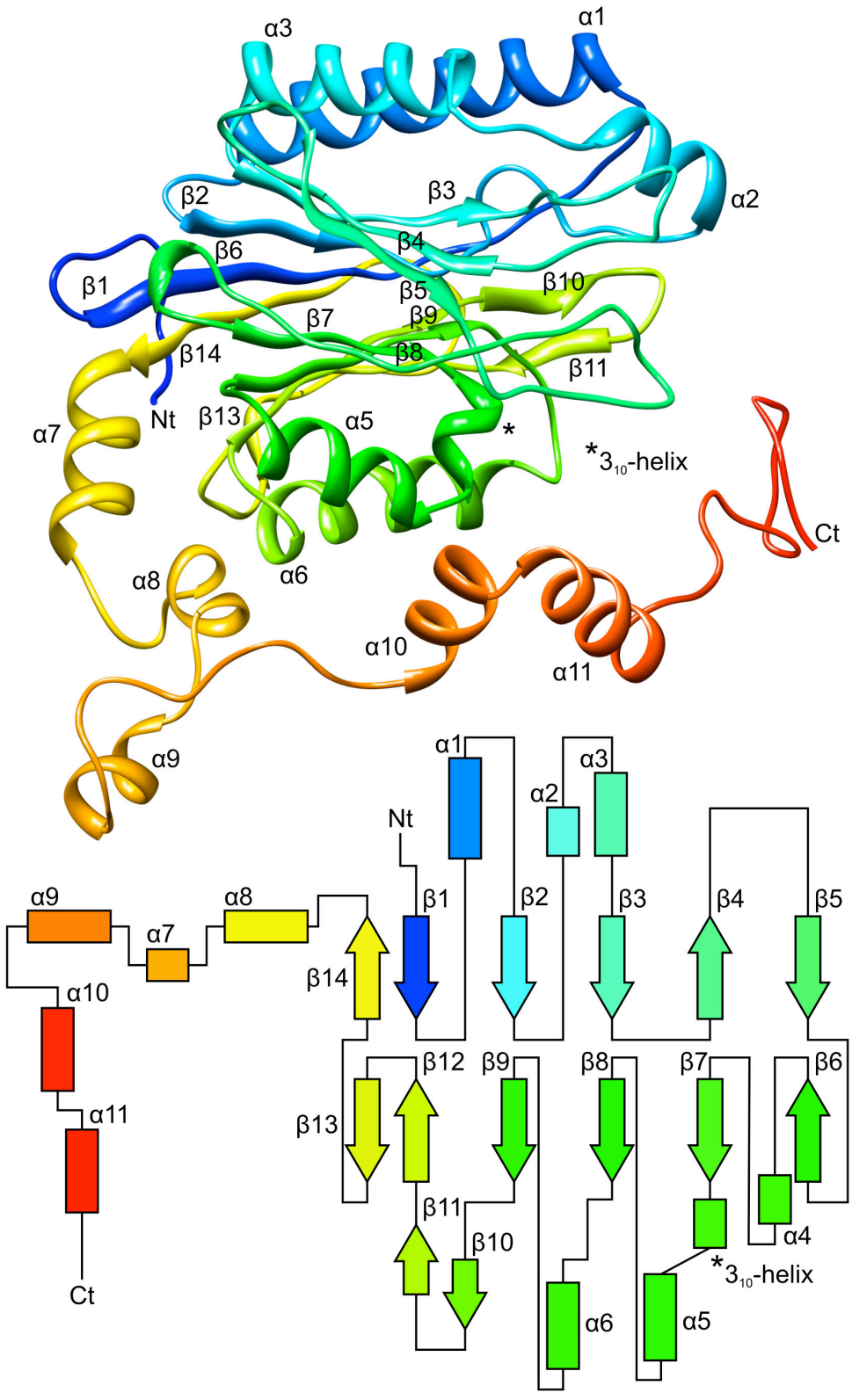


Figure 4. Nitrilase superfamily monomer fold: *G. pallidus* RAPc8 amidase (after Kimani *et al*., 2007). Cartoon representation and diagram indicating α-helix and β-sheet topology. The N-terminus (Nt) and C-terminus (Ct) are indicated. The active-site cysteine resides on the 3_{10}-helix (*).

All plants possess at least one nitrilase 4 (NIT4) homologue (Piotrowski, 2008). NIT4 enzymes catalyse the hydrolysis of β-cyanoalanine (β(CN)), an intermediate breakdown product of cyanide with high specificity to form the amino acids asparagine (amide) and aspartic acid (acid) as well as ammonia (Piotrowski *et al*., 2001). The nitrilase 1 (NIT1) family of enzymes, which includes nitrilase 1, 2 and 3, most likely arose from a gene duplication event as the result of a polyploidization event early in the phylogeny of the *Brassicaceae* (Piotrowski, 2008).

NIT1 enzymes display activity for a broad range of substrates, but some NIT1 enzymes are specialised for hydrolysing nitriles that are identical or nearly identical to the breakdown product of the major glucosinolates found in particular plant species. NIT1 enzymes are therefore believed to be involved in the endogenous catabolism of glucosinolates. Interestingly, NIT1 enzymes display no activity for β(CN) the physiological substrate of NIT4 (Piotrowski, 2008). More than 120 glucosinolates have been described thus far, making NIT1 enzymes a very interesting source for nitrilases with novel substrate specificities. Because NIT1 enzymes are monophyletic and several have only diversified in recent evolutionary history (Janowitz *et al*., 2009), they represent a powerful tool for determining the factors responsible for differences in substrate specificity and amide:acid ratio.

0.1.5 Aims

A large collection of plant nitrilases, including both nitrilase 4 and nitrilase 1 enzymes have been cloned, expressed and characterised with respect to substrate specificity and amide:acid ratio in the laboratory of Markus Piotrowski. This collection includes nitrilases with very high sequence conservation (up to 90% homology) with defined functional differences of substrate specificity or amide:acid ratio of products. This unique collection, as well as data from fungal and bacterial nitrilases in the laboratory of Trevor Sewell forms the basis of the experimental material described in this thesis. This material is analysed here by homology modelling using crystallised nitrilase superfamily enzymes as templates. Electron microscopy and image processing are used to determine the low-resolution (~25 Å) quaternary structure of nitrilase complexes. Site-directed mutagenesis and activity screening are used to test hypotheses concerning specific residues.

0.2 Materials and methods

0.2.1 Homology modelling

The target protein sequences were submitted for structural sequence alignment to the GenTHREADER (Jones, 1999; McGuffin and Jones, 2003) web server. Template alignments (pdb codes: 1f89, 1j31, 1uf5, 2plq, 2vhh, 2dyu) were manually merged using secondary structure predictions from PSIPRED (McGuffin *et al*., 2000). Nitrilase dimer homology models were built using MODELLER (Sali and Blundell, 1993; Fiser and Sali, 2003). No attempt was made to model loop insertions relative to the templates. Side-chain conformations were adjusted using SCWRL4 (Krivov *et al*., 2009). The resulting models were aligned with the template structures in three dimensions using UCSF Chimera (Pettersen *et al*., 2004) and the initial structural sequence-alignment was adjusted on a per-residue basis, taking conservation of the template structures, the predicted secondary structure of the target sequences, target and template sequence similarities and the three-dimensional position of each structural element into consideration. The corrected alignment was used as a new MODELLER input and the process was iterated until an optimal correspondence was obtained. This process ensured that (at least in highly structurally conserved regions) there was a one-to-one correspondence in three dimensions between all template – and target residues.

0.2.2 Molecular biology

0.2.2.1 Overlap extension PCR

Overlap extension PCR (Ge & Rudolph, 1996) was performed using KAPA HiFi DNA polymerase (PEQLAB Biotechnology GmbH, Karlsruhe, Germany) to construct a chimeric enzyme comprised of *At*NIT4 and the C-interface loop lying between β10 and β11 (see Fig. 4) of *At*NIT1 according to the method of Wurch *et al*. (1998) (Fig.5).

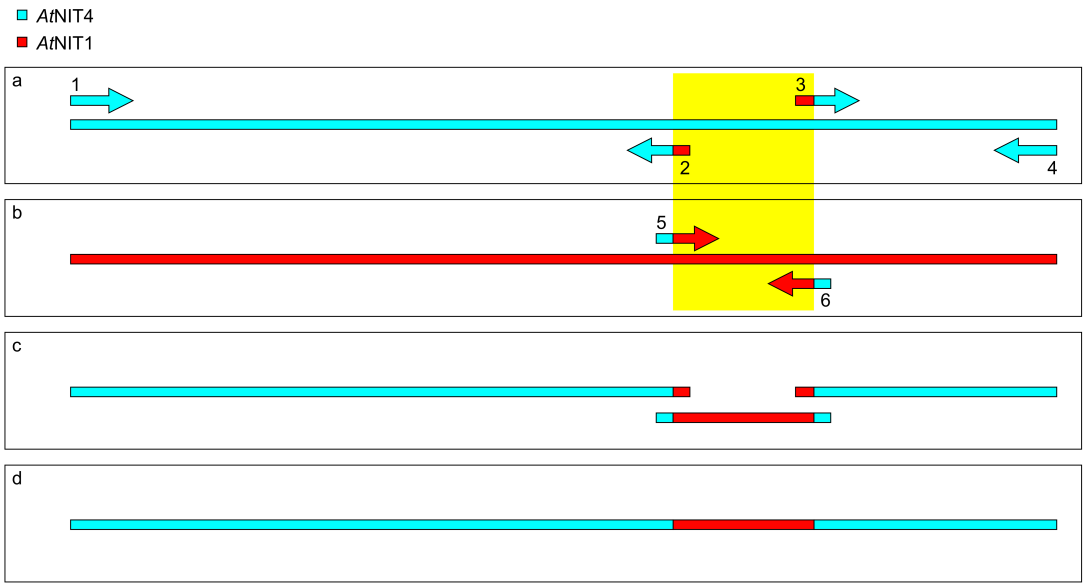


Figure 5. Construction of an "*At*NIT4 *At*NIT1 loop" chimera by overlap-extension PCR. (**a**) The region shown in yellow is homologous to a C-interface loop in the crystal structure of beta-alanine synthase from *D. melanogaster* (2vhi: Lundgren *et al*., 2008). In the first PCR step, the two flanking *At*NIT4 fragments are amplified. Primers 1-6 are listed in the table below, the 5' end of the first fragment and the 3' end of the second fragment are partially derived from *At*NIT1. (**b**) Amplification of the loop from *At*NIT1, in this case, the primers are partially complementary to *At*NIT4 (listed in the table below). (**c**) The third step, the fragments are mixed in equimolar amounts and a PCR is run with no additional primers. The fragments anneal across the overlapping region (as shown) and extend to product the PCR product shown in (d). (**d**) The full-length construct is amplified using primers (1) and (4) and the final product is ligated into the cloning vector.

*At*NIT4 with *At*NIT1 "C-surface" loop chimera:

Primer	Sequence
(1) *At*NIT4_F*	CATATGTCCATGCAACAAGAAACGTCTC
(2) *At*NIT4_loop_R	GACGAAACATCCACCCTCAAGTGCAATATG
(3) *At*NIT4_loop_F	ATTTCACCTTTGGGAATTGTTTTAGC
(4) *At*NIT4_His_R*	GTCGACGACGGATTCATCTTCCATTACCTTTG
(5) *At*NIT1_loop_F	TTGAGGGTGGATGTTTCGTCTTGTCGGC
(6) *At*NIT1_loop_R	AATTCCCAAAGGTGAAATAATGACACTTC

*Piotrowski *et al*. (2001)

0.2.2.2 Site-directed mutagenesis

Site-directed mutagenesis based on the QuickChange™ protocol was performed with partially overlapping primers (Zeng *et al*., 2004) using KAPAHiFi DNA polymerase (PEQLAB Biotechnology GmbH, Karlsruhe, Germany). Primers were designed by attempting to fulfil the following criteria: a minimum of 8 non-overlapping 3' bases; G/C at both ends; Tm > 64ºC; mutation >4 bases from 5' end and 6-8 bases from the 3' end (Fig. 6).

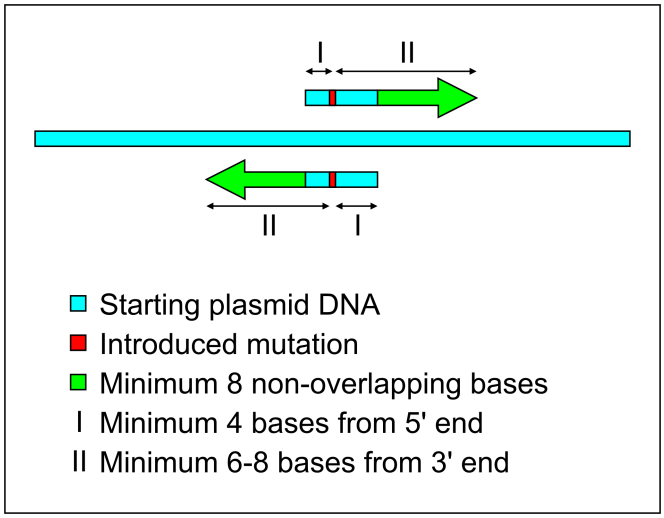


Figure 6. Site-directed mutagenesis according to the protocol of Zeng *et al*. (2004). The starting plasmid DNA is shown by a cyan line, arrows indicate the forward (right-hand facing) and reverse (left-hand facing) primers. Primer overhangs are used to reduce the affinity of primer-dimers relative to plasmid binding, a minimum of eight non-overlapping bases are used (green). The introduced mutation is indicated in red, the minimum number of bases between the mutation and the 3' and 5' ends is indicated by roman numerals.

*Cr*NIT1a substrate specificity exchange with *Sal*NIT1c:

Primer	Sequence	T_m (°C)
*Cr*NIT1a_T153A_L154M_F	TGCCCACAGCCATGGAACGTTGCATATGGG	75.00
*Cr*NIT1a_T153A_L154M_R	CGTTCCATGGCTGTGGGCATGAGCTTACG	74.78
*Cr*NIT1a_G209W_F	ACTGCTGATTGGTCGAAAGAATGGCAATCGTCC	74.38
*Cr*NIT1a_G209W_R	CCATTCTTTCGACCAATCAGCAGTAGGTGCAC	74.16
*Cr*NIT1a_F80Q_F	CTGTGGGAGTTCAAAACGAAGAAGGCCGTGAC	72.28
*Cr*NIT1a_F80Q_R	CTTCGTTTTGAACTCCCACAGCCAAACCAAACC	71.35

*Sal*NIT1c substrate specificity exchange with *Cr*NIT1a:

Primer	Sequence	T_m (°C)
*Sal*NIT1c_A153T_M154L_F	CATGCCAACAACTCTGGAACGTGTCATATGGG	74.16
*Sal*NIT1c_A153T_M154L_R	CACGTTCCAGAGTTGTTGGCATGACTTTACGG	74.16
*Sal*NIT1c_W209G_F	TGCTGATGGGTCCAAGGAATGGCAATCG	75.25
*Sal*NIT1c_W209G_R	TCCTTGGACCCATCAGCAGTAGGTGC	74.77
*Sal*NIT1c_Q80F_F	CGGTTGGGGTTTTTAACGAAGAAGGTCGTGACC	72.56
*Sal*NIT1c_Q80F_R	CTTCTTCGTTAAAAACCCCAACCGCCATGCCG	72.28

*At*NIT4 substrate specificity exchange with *Sal*NIT1c:

Primer	Sequence	T_m (°C)
*At*NIT4_WSKE_F	ACTGCTGATTGGAGCAAAGAATGGCTAGCATCAATGACTC	65.6
*At*NIT4_WSKE_R	ACGATCGGTAAGAAACGAGGTTAGTCGTCATCCACGTG	65.8
*At*NIT4_L154M_F	TACAGCTATGGAACGTTGCATTTGGGGATTTG	72.1
*At*NIT4_L154M_R	TTTACGTTGCAAGGTATCGACATCCGTACTC	71.8

*Sal*NIT1c substrate specificity exchange with *At*NIT4:

Primer	Sequence	T_m (°C)
*Sal*NIT1c_SRET_F	ACTGCTGATTCGCGCGAGACATGGCAATCGTCGGTGATGC	73.2
*Sal*NIT1c_SRET_R	CTGCTAACGGTACAGAGCGCGCTTAGTCGTCATCCACGTG	73.2
*Sal*NIT1c_M154L_F	AACAGCTTTGGAACGTGTCATATGGGG	72.5
*Sal*NIT1c_M154L_R	TGTGCAAGGTTTCGACAACCGTACTG	72.5

*At*NIT4 amide: acid ratio mutants:

Primer	Sequence	T_m (°C)
*At*NIT4_I196V_F	CGGTGCTGCTGTTTGTTGGGAAAATAGG	72.9
*At*NIT4_I196V_R	CCAACAAACAGCAGCACCGATTTTCCC	72.5
*At*NIT4_IV_RK_F	CTGCTGTTTGTTGGGAAAATAAGATGCCTTCTTTGAG	74.5
*At*NIT4_IV_RK_R	CATCTTATTTTCCCAACAAACAGCAGCACCGATTTTC	74.5

Reactions were assembled in 25 µl volumes, comprised of:

5 µl 5× KAPA HiFi Fidelity buffer*
0.75 µl dNTP mix*
0.75 µl forward primer (final concentration: 0.15 µM)
0.75 µl reverse primer (final concentration: 0.15 µM)
1 µl ~0.5 ng/µl template DNA
0.5 µl (0.5 U) polymerase*
16.25 µl Sterile distilled H2O (to 25 µl)
*PEQLAB Biotechnologie GmbH, Karlsruhe

Reaction conditions:

1	95°C	4 min	Goto 2
2	98°C	20 sec	Goto 3
3	60°C – 72°C*	45 sec*	Goto 4
4	72°C	30 s/kbp*	Goto 2 × 20
5	72°C	4 min	Goto 6
6	4°C	Hold	

*Reaction conditions were varied if the reaction was initially unsuccessful.

0.2.2.3 Transformation and sequencing

Initial PCR product transformation into *E. coli* XL1-blue (*rec* A1, end A1, *gyr* A96, *thi*1, *hsd* R17, *sup* E44, *rel* A1, *lac* [F', *pro*AB, *lacI*^{Q}ZΔM15, Tn10 (*Tet*^{r})]) (Bullock *et al*., 1987) was performed by electroporation (Gene Pulser II, BioRad, München,

Germany) at 2.5 kV, 25 μF and 200 Ω. The cells were regenerated in autoclaved 2YT medium without antibiotics at 37ºC for 30min. A sample (200 μl) was plated onto autoclaved 1.5% (w/v) bacto-agar LB solid-medium plates (Ausubel *et al*., 1995) containing ampicillin (100 μg/ml) and incubated overnight at 37°C. In the case of blue-white selection, the solid medium contained: 0.8 mg X-Gal and 8 mg IPTG per plate in addition to the above. Colonies were selected and used to inoculate three YT2 cultures of 4 ml each (ampicillin (100 μg/ml) and chloramphenicol (30 μg/ml)), and grown at 37°C with constant agitation overnight (~16 hours).

Plasmid DNA was purified using a miniprep kit (QIAprep® Spin Miniprep kit, Qiagen, Hilden) a sample was sent for sequencing and the remainder stored at -20°C. Plasmids from positive clones were heat-shock transformed into Ca^{2+} competent *E-coli* BL21-Codon plus (DE3)-RIL (*E. coli* B, F-, *omp*T, *hsbS* (rb-mb-), *dcm*+, Tetr, *gal*λ, (DE3), *end*A, Hte [*arg*U, *ile*Y, *leu*W, Camr]) (Stratagene, Amsterdam, Netherlands). The cells were regenerated in 2YT medium for 30 min and 50 μl was plated out onto autoclaved 1.5% (w/v) bacto-agar LB solid-medium plates (Ausubel *et al*., 1995) containing ampicillin (100 μg/ml). A single colony was selected and used to inoculate 4 ml of 2YT-medium containing ampicillin (100 μg/ml) and chloramphenicol (30 μg/ml), grown at 37°C (220 rpm) overnight (~16 hours). The cells were mixed with 60% (v/v) glycerol (1:1) to a final concentration of 30% glycerol and placed in long-term storage at -80°C.

0.2.3 Protein expression and purification for activity assays

The cells were streaked from frozen glycerol stocks onto autoclaved 1.5% (w/v) bacto-agar LB solid-medium plates (Ausubel *et al*., 1995) containing ampicillin (100 μg/ml) and incubated overnight at 37°C. Single colonies were selected and used to inoculate 20 ml 2YT (Ausubel *et al*., 1995) liquid starter cultures containing ampicillin (100 μg/ml) and chloramphenicol (30 μg/ml) and grown for 16 hours at 37°C with aeration. Large Y2T cultures (600 ml) containing ampicillin (100 μg/ml) and chloramphenicol (30 μg/ml) were inoculated with the entire volume of starter culture and grown for ~10 hours at 37°C with constant agitation (220 rpm). The cells

were pelleted at 5000×G for 20 min at 4°C, vortexed until they formed a loose slurry, aliquoted, flash frozen in liquid nitrogen and stored at -80°C.

The frozen cells were resuspended in 60 ml lysis buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl) with 5 mM β-Mercaptoethanol and 1 mg/ml lysozyme and placed in a bath sonicator (Sonorex RK 510 S. Bandelin, Berlin) at 0°C for 15 min. The cells were disrupted by sonication (B-17 sonifier. Branson, Banbury, USA) using a tapered micro tip in 30 s intervals for a total of 4 min and clarified by centrifugation at 13 000 ×G for 30 min. The target protein was precipitated from solution by the gradual addition of ammonium sulphate to a final saturation of 40% with stirring on ice for 30 min. The protein was pelleted at 13 000 ×G for 20 min, resuspended in 12 ml of lysis buffer and centrifuged (5000×G, 5 min).

The protein solution was loaded onto an equilibrated (10 column volumes of lysis buffer containing 10 mM imidazol) 2ml Ni^{2+}-NTA-agarose column (GE healthcare, USA) and washed with 5 column volumes of wash buffer (lysis buffer containing 40 mM imidazol). The protein was eluted with 1.25 column volumes of lysis buffer containing 250 mM imidazol. The eluent (2.5 ml) was loaded onto a PD-10 size exclusion column and eluted with 3.5 ml of KPi buffer (50 mM potassium phosphate pH 8.0, 0.1 mM DTT). The protein solution was divided into 500 µl aliquotes and flash frozen in liquid nitrogen in 1.5 ml Eppendorf tubes and stored at -80°C.

0.2.4 Protein expression and purification for electron microscopy

Ten ml starter cultures (LB, ampicillin (100 µg/ml) and chloramphenicol (30 µg/ml)) were inoculated with single colonies and grown overnight at 37°C with aeration. Four ml were extracted and used to inoculate 200 ml growth cultures (LB, ampicillin (100 µg/ml) and chloramphenicol (30 µg/ml)) and grown at 37°C with aeration for approximately 3 hours until an OD_{600} of between 0.4 and 0.7 was reached. IPTG (0.3 mM) was added to induce expression of the recombinant protein for a further 3 hours: the cells were pelleted at 5000×G for 30 min at 4°C, the supernatant discarded and the pellet softened by vortexing before being frozen at -20°C.

The frozen pellets were resuspended in gel-filtration buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl) with complete protease inhibitor cocktail (Roche, Germany) and disrupted by sonication (Misonix Inc. Farmingdale, USA) (9 watts in 15 s intervals for a total of 4 min) at 0°C. The cell debris was pelleted at 13 000×G for 30 minutes, contaminating protein was precipitated at an ammonium sulphate saturation of 20% at 0°C for 30 minutes, pelleted at 10 000×G for 10 min and discarded. The heterologously expressed protein precipitated from solution at an ammonium sulphate saturation of 40%, and was pelleted at 10 000×G for 20 min. The pellet was resuspended in gel-filtration buffer with 50 mM imidazol, filtered through a 0.22 μm filter and applied to an equilibrated Ni^{2+}-NTA-agarose column (GE Healthcare, USA), washed with 10 column volumes of gel-filtration buffer with 50 mM imidazol and eluted with gel-filtration buffer containing a linear gradient of imidazol from 50 mM to 0.5 M. The peak containing the protein of interest was identified by SDS PAGE and concentrated to a volume of 500 μl by ultrafiltration with a 30 kDa Nanosep Concentrator (Pall Corporation, East Hills, USA). The concentrated protein solution was separated by gel-filtration on a Sephacryl S-300 HR gel-filtration column (GE-Healthcare, USA). The protein eluted as a range of sizes, the leading edge of the high molecular weight fraction was selected and allowed to polymerise at 4°C for a few days.

0.2.5 SDS-PAGE

Protein purity was estimated by SDS polyacrylamide gel-electrophoresis (Fig. 7) (SDS-PAGE) (Laemmli, 1970) using a Mini Protean 3 cell (Bio-Rad) stained with 0.25% Coomassie Brilliant Blue R250 in destain solution (45% (v/v) methanol and 10% (v/v) acetic acid in distilled water). A 5% stacking – and 12% separating gel were used to resolve the nitrilase subunit MW of ~36 KDa. Molecular weight estimates were made with pre-stained molecular weight markers (Fermentas Life Sciences).

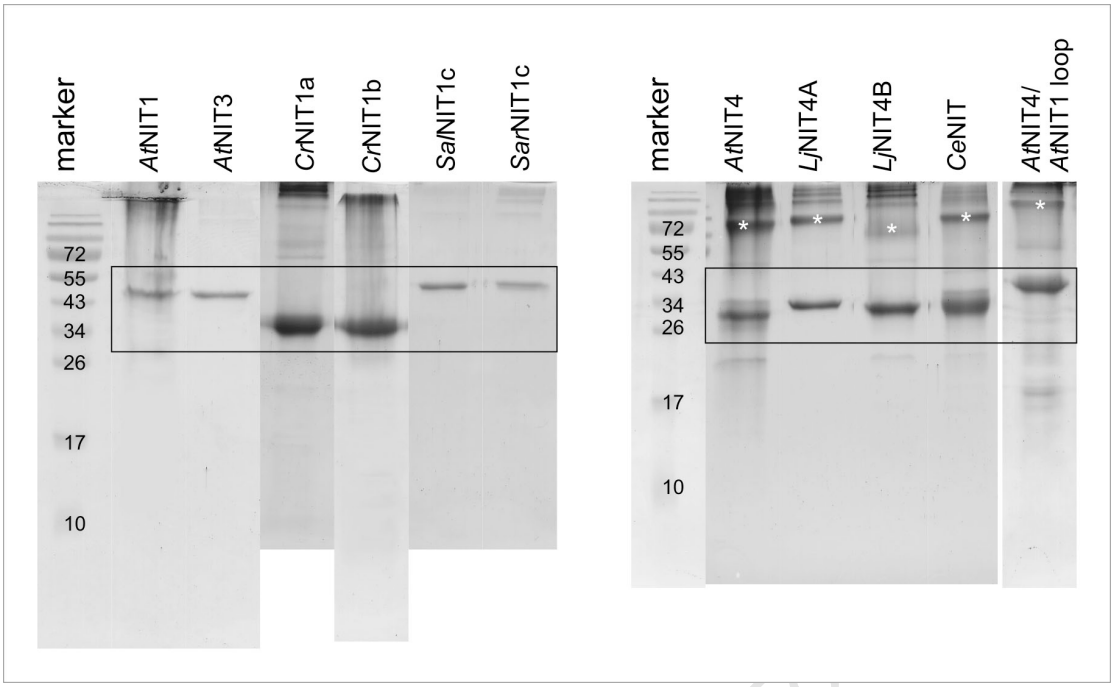


Figure 7. Protein purity assessed by SDS polyacrylamide gel-electrophoresis. A rectangular box indicates the nitrilase band. In the case of nitrilase 4 (right), a band that runs at a molecular mass of just above twice that of the monomer is visible (*) sequencing this band by mass spectroscopy has demonstrated that this band consists of nitrilase enzyme (Markus Piotrowski, personal communication).

0.2.6 Bradford assay

Protein concentration was estimated using the Bradford assay (Bradford, 1976). The OD_{620} of Bradford reagent (0.01% (w/v) serva blue G 250, 4.7% (v/v) ethanol, 8.5% (v/v) phosphoric acid) was measured with a dilution series of protein to ensure that the measurement is performed in the linear protein concentration range. A standard curve was constructed using bovine serum albumin (BSA). Photometric readings were performed on a micro-titre plate using a plate reader (Titertek Multiscan plus MKII).

0.2.7 Specific activity measurement

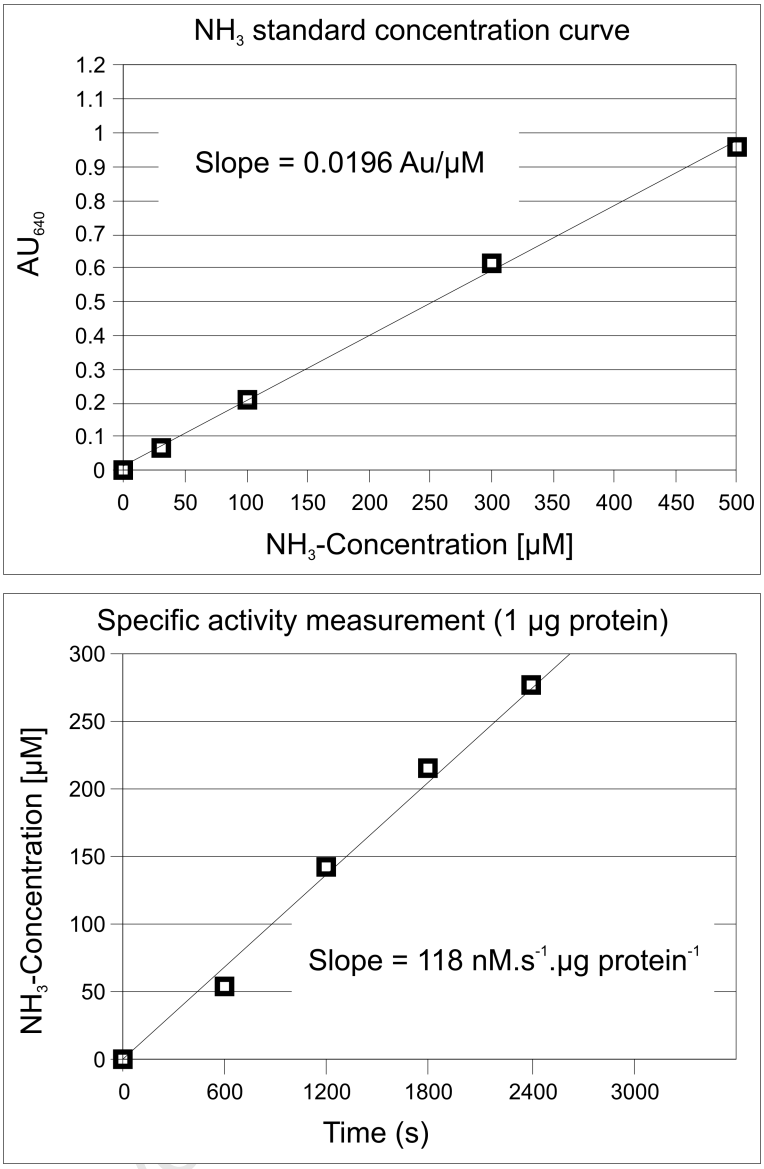


Figure 7. Specific nitrilase activity assay. Ammonia standard concentration curve: 4 concentrations of NH_4Cl were assayed to produce the standard curve shown. A slope of ~0.02 Au/µM and correlation coefficient (R^2) of ~0.998 was typical. Specific activity measurement: the NH_3 concentration was calculated (using the ammonia standard curve) at time intervals (in this case, 10 min each). The NH_3 concentration was plotted as a function of time to ensure that at least two of the measurements were still in the linear range. Dividing the slope of the linear portion by the protein concentration provides the specific activity of the sample in nKat ($mol.s^{-1}$) per mg protein. This procedure was performed in triplicate and the mean specific activity and standard deviation (SD) values were calculated.

Depending on the level of activity, an enzyme concentration of between 0.1-10 µg/ml was used in specific activity measurements. Specific activity assays were performed in triplicate. The reaction tubes contained: enzyme, 50 mM potassium phosphate buffer pH 8.0, 1 mM DTT with 2.5 mM substrate. A negative control, consisting of heat-denatured enzyme (incubated at 100°C for 10 min), was tested in parallel to every experimental condition (DTT and substrate were added just before incubation).

Samples were incubated at 37°C; a time-series of 4 samples were taken at intervals of between 5-30 min (depending on the activity). Substrates tested: 3-butenenitrile (3BUT) (Lancaster Synthesis GmbH, Mühlheim, Germany), 4-hydroxyphenylacetonitrile (4HPAN) (Sigma-Aldrich, GmbH, Steinheim, Germany) and β-cyanoalanine (β(CN)) (Sigma-Aldrich, GmbH, Steinheim, Germany). Substrates were made up as 250 mM stock solutions in 1 ml quantities and stored at -20°C. After observing the inhibitory effect of methanol on 4HPAN activity, 4HPAN was dissolved in 100% methanol and subsequently diluted to 10% methanol by the addition of distilled water. β(CN) was dissolved in H_2O and 3BUT was dissolved in methanol.

0.2.8 Photometric detection of free ammonia

The concentration of free ammonia in solution was determined using the Berthelot-reaction (Hiller and Van Slyke, 1933). Glass tubes (10 ml) were prepared with 100 µl sodium phenolate, at each time-point 100 µl of sample was pipetted into a glass tube, halting the nitrilase reaction. Sodium hypochloride (100 µl × 20 mM) and sodium pentacyanonitrosylferrate (II) (100 µl × 0.01% (v/v)) were added to the glass tube, which was placed in boiling water for 2 min. After allowing the solution to cool to room temperature, 600 µl of distilled water was added to the solution and the extinction coefficient at 640 nm was recorded. A standard concentration curve was constructed with NH_4Cl and used to determine the ammonia concentration in each sample. After ensuring that the reaction was still in the linear range, the specific activity was calculated for each enzyme/ substrate combination (Fig. 8).

0.2.9 Negative-stain electron microscopy

The purified protein solution was pipetted onto glow discharged carbon-coated 4 mm copper grids. The protein was allowed to adhere for 30 s, blotted, washed three-times with distilled water and stained with uranyl acetate, blotted again and allowed to dry at room temperature. The prepared grid was loaded into a FEI/Tecnai F20 FEGTEM equipped with a 4k×4k CCD camera (GATAN US4000 Ultrascan, California, USA)

and imaged at 200 kV with a defocus (under focus) of 300-500 nm under standard low dose conditions. Focus and astigmatism were corrected using the real-time power spectrum.

0.2.10 Helical image processing

Particles were selected using BOXER (Ludke *et al*., 1999); nitrilase filaments were windowed along the helical axis with 90% overlap and vertically aligned. The resulting helical segments were exported into Spider format (Frank *et al*., 1996) using em2em (Image Science, Berlin Germany). Images containing more than one helix or debris were deleted, as well as those containing curved or damaged filaments. Image processing was performed using Spider V17 (Frank *et al*., 1996) for Mac OSX. High intensity pixels (>99th percentile) were assigned the pixel density of the image average. The images were band-pass Fourier filtered to a resolution of between 270 Å and 16 Å and normalised to a mean of 0 and standard deviation (SD) of 1. The segments were interpolated down, with summation of adjacent pixels, by a factor of 2, yielding 64×64 pixel images.

0.2.11 Helical reconstruction

The filtered and normalised segments were padded to 512×512 pixels and multiplied by a Gaussian mask to produce a smooth intensity fall-off. Power spectra were calculated from the padded images, averaged together and windowed to 128×128 pixels. The pitches of each of the three visible layer-lines, at the available resolution (~25 Å), were determined directly from the averaged power spectra. The helical handedness was taken from Woodward et al. (2008) where unidirectional shadowing was used to unambiguously identify left-handed one-start helical striations. Indexing was accomplished by comparing the ratios of the predicted principal maxima of different Bessel orders to the corresponding experimentally determined ratios measured from the power spectra.

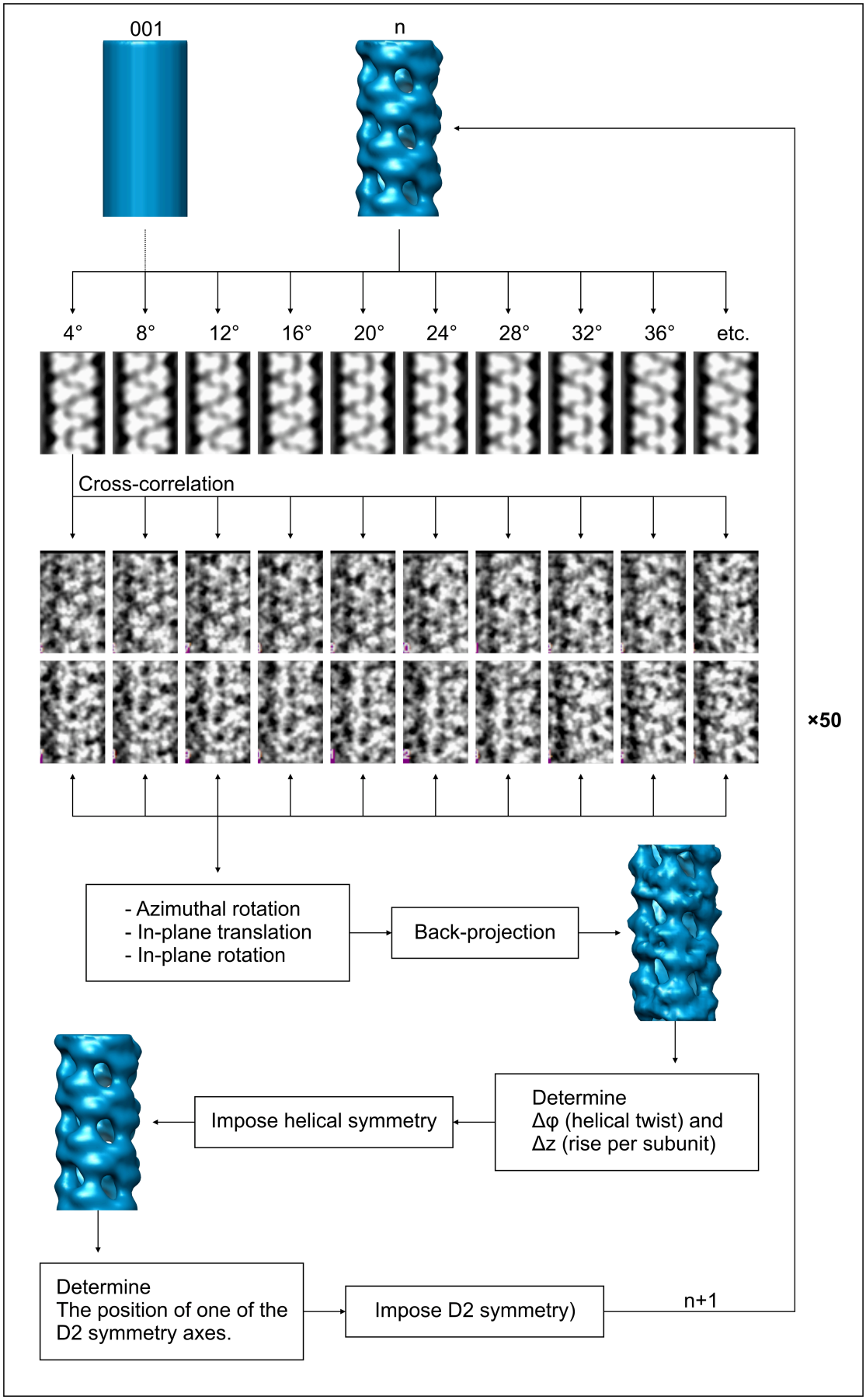

001
n
4°
8°
12°
16°
20°
24°
28°
32°
36°
etc.
Cross-correlation
×50
- Azimuthal rotation
- In-plane translation
- In-plane rotation
Back-projection
Determine
Δφ (helical twist) and
Δz (rise per subunit)
Impose helical symmetry
Determine
The position of one of the
D2 symmetry axes.
Impose D2 symmetry)
n+1

Figure 8 (overleaf). Overview of the reconstruction algorithm. Helical volumes were reconstructed using the iterative helical real space reconstruction (IHRSR) algorithm (Egelman, 2000) with the additional constraint of two-fold symmetry search and imposition. Short helical segments were windowed from the original micrographs with 90% overlap using 64×64 pixel sized boxes (Egelman, 2000). The algorithm was initiated using a featureless cylinder and the helical symmetry values estimated by indexing. The model was re-projected at 4° intervals about the helical axis to produce reference images that were cross-correlated with the input images in order to refine the three-dimensional alignment parameters of each input image. The new alignment parameters were used to produce a new unsymmetrised volume, the helical parameters were measured from the new volume and imposed. The additional search and imposition of the two-fold nitrilase symmetry axis further constrains the reconstruction process.

The iterative helical real-space reconstruction (IHRSR) algorithm described by Egelman, (2000) was used to reconstruct the filaments in three dimensions. Alignment parameters of the helical segments (in-plane rotation, in-plane translation and angle about the helical axis) were refined iteratively by projection matching, initiated with a featureless cylinder. Helical symmetry was imposed during the first iteration and refined by determining the symmetry of the resulting volume from the previous iteration, searching for the two-fold symmetry axis and applying these (Fig. 9). Multiple helical symmetries were consistent with the data at the available resolution, because while layer lines 2 and 4 could be assigned to Bessel orders -1 and -2 respectively, layer line 3 could be assigned to Bessel orders 3,4 or 5 (Woodward *et al*., 2008). The correct helical symmetry was determined by evaluating whether or not the algorithm converged on a stable solution. The IHRSR algorithm is well established for the determination of helical twist (Egelman, 2000). The method relies on a procedure whereby the helical twist and rise are measured directly from the electron density in an iterative way. Failure to converge on the correct helical twist results in a failure of the electron density map to converge on a stable solution. In order to test the reproducibility of the procedure, independent reconstructions were produced from different starting symmetries; the resulting models had helical twists that were identical to the third or fourth decimal place. Convergence, measured by the change in helical symmetry, occurred after approximately 30 iterations of the IHRSR algorithm.

0.2.12 *Docking*

Docking PDB models into electron density maps can provide information about the nature of the helical contacts occurring between protein monomers. This procedure has been applied extensively to nitrilase enzymes in the past (e.g. Sewell et al., 2003;

Thuku et al., 2007; Vejvoda et al., 2008; Woodward et al., 2008; Dent et al., 2009; Williamson et al., 2010; Kaplan et al., 2011). The insights gained from these past studies were used in this thesis, but the results from this procedure were not included again. This was because the region found to be responsible, in the plant nitrilases, for determining helical twist lacks a crystallized homologue and therefore could not be accurately modeled. Presumably, this region forms a helical contact, but in the absence of evidence, the location of this contact would be purely speculative. This meant that no new insights were obtained from the docking result and it was excluded.

0.2.13 *Number of short helical segments used in the reconstruction*

*At*NIT4	1015
*Ce*NIT	969
*Lj*NIT4A	1662
*At*NIT3	2736
*At*NIT1	1081
*Cr*NIT1a	592
*Cr*NIT1a F80H	1188
*Cr*NIT1b	1079
*Cr*NIT1b H80F	1824
*Sal*NIt1c	1158

0.2.14 Single-particle image processing

Single particles were selected manually and windowed using boxer to 200×200 pixels. The images were normalised to a mean of 0 and standard deviation of 1 and low-pass filtered to 16 Å. These and all subsequent image-processing steps were performed using Spider V17 (Frank *et al*., 1996).

0.2.15 Single-particle reconstruction

The selected particles were classified into 32 classes by rotationally invariant k-means classification (Penczek *et al*., 1996). The most populous 16 classes were refined by several cycles of multi-reference alignment (Joyeux and Penczek, 2002) and averaging with the exclusion of images with low correlation scores. A starting model was created using the Spider V17 (Frank *et al*., 1996) implementation of the SIRT (simultaneous iterative reconstruction technique) algorithm (Penczek *et al*., 1992) using the relative orientations of these averages estimated by the method of common lines (Penczek *et al*., 1996). This initial volume was refined by fifty cycles of

projection matching: the model was reprojected at quasi-equally spaced intervals of 15° to create a set of reference images that were used to align the input images by multireference alignment. The input images were averaged to create new classes with the improved alignment data and used to create a new volume, which was fed back into the algorithm again.

0.2.16 Resolution estimation

The resolution of the reconstructed volumes was estimated using spectral signal to noise ratio (SSNR) (Unser *et al*., 1987, Penczek, 2002). The Fourier Shell Correlation was calculated according to: FSC = SSNR/(SSNR+1) where FSC = 0.5.

0.2.17 Data visualisation

All molecular visualisation and high quality image rendering was performed using UCSF Chimera (Pettersen *et al*., 2004). Threshold values were calculated from the predicted volume enclosing the molecular weight of the protein complex using an average protein density value of 0.73 Da.Å^{-3} (Quillin *et al*., 2000).

0.2.18 *Mutant list*

Helical twist exchanges	**Active-site pocket exchanges cont.**
*Cr*NIT1a to *Cr*NIT1b helical twist mutants	*At*NIT4 to *Sal*NIT1c mutants
*Cr*NIT1a F80H*	*At*NIT4 L169M
*Cr*NIT1b H80F*	*At*NIT4 S224W R225S E226K T227E
*Sal*NIT1c W209G A153T M154L Q80F	*At*NIT4 S224W R225S E226K T227E L169M
*Cr*NIT1a G209W T153A L154M F80Q	*Sal*NIT1c to *At*NIT4 mutants
Active-site pocket exchanges	*Sal*NIT1c M154L
*Sal*NIT1c to *Cr*NIT1a mutants	*Sal*NIT1c W209S S210R K211E E212T
*Sal*NIT1c W209G	*Sal*NIT1c W209S S210R K211E E212T M154L
*Sal*NIT1c A153T M154L	Amide:acid mutants
*Sal*NIT1c W209G A153T M154L	*Sar*NIT1c ΔG66*
*Cr*NIT1a to *Sal*NIT1c mutants	*Sal*NIT1c +G66*
*Cr*NIT1a G209W	*Sar*NIT1c ΔG66 Y68A*
*Cr*NIT1a T153A L154M	*Sar*NIT1c ΔG66 Y68F*
*Cr*NIT1a G209W T153A L154M	*Sal*NIT1c Y68A*
Interface disruption mutants	*Sal*NIT1c Y68F*
*At*NIT4 237(AtNIT4) 278(AtNIT4)	*At*NIT4 Y82F*
P. stutzeri CynD Δ296	*At*NIT4 R201K*

(*Trompetter, 2010)

SECTION 1: Substrate specificity

1.1 Introduction

1.1.1 Interface formation

The conservation of the C-interface among all active nitrilases points to oligomerisation as being important for nitrilase activity (Sewell *et al*., 2005). Furthermore, it has been generally observed that while nitrilases exist as a mixture of helical or helix-like oligomeric states in solution, the dimer fraction, when isolated by gel-filtration, shows no activity (e.g. Nagasawa *et al*., 2000; Thuku *et al*., 2007). In *R. rhodochrous* J1 nitrilase the activity of the dimer fraction can be restored by addition of the substrate benzonitrile, which results in the formation of decamers (Nagasawa *et al*., 2000) and in the case of NIT4b from *Lotus japonicus* by a freeze-thaw cycle (M. Piotrowski, personal communication). The cynD from *B. pumilus* makes the transition between short spirals and extended helices at pH 5.5 and simultaneously shows a slight increase in activity (Jandhyala *et al*., 2005). Mutation of the putative interface residues generally leads to a loss of activity, which has been explained as oligomer disruption (Sewell *et al*., 2005).

Grasses (*Poaceae*) express two nitrilase 4 isoforms that illustrate the complex relationship between subunit association and activity. These enzymes have a dual catalytic function that appears to be associated with complex formation (Jenrich *et al*., 2007; Kriechbaumer *et al*., 2007). For example, the nitrilase NIT4A and NIT4B2 isoforms from *Sorghum bicolor* are inactive when isolated from one another. On mixing, the two isoforms associate to form a NIT4A/ NIT4B heterocomplex that hydrolyses β-cyanoalanine as well as other substrates, the best being 4-hydroxyphenolacetonitrile. Inactivation of the catalytic site by site-directed mutagenesis has demonstrated that β-cyanoalanine catalytic activity is restricted to the NIT4A isoform, while the NIT4B2 isoform accounts for the rest of the catalytic activity (Jenrich *et al*., 2007).

Kimani *et al*. (2007) suggested a possible explanation for the role of oligomerisation on activity: the C-interface positions a fourth catalytic residue (Glu_2) (2plq: Glu_{142}) within the active site by providing stability to an otherwise flexible loop. Two lines of evidence support this hypothesis; the first comes from the crystal structure of the β-

alanine synthase from *Drosophila melanogaster* (Lundgren *et al*., 2008), which was crystallized in a helical turn comprised of four dimers (2vhi). The inner pair of dimers make helical contacts at the interface which holds a short alpha helix (α4) in position, maintaining Glu_2 in the conserved position in the active site. The dimers at the outer ends of the turn make no helical contacts and both Glu_2 (2vhi: Glu_{208}) and α4 are disordered. The second line of evidence comes from inactivation of the *G. pallidus* amidase by site-directed mutagenesis of Glu_2 (E142L) (Brandon Weber, personal communication).

1.1.2 Substrate specificity

Nitrilases are often selective for a specific substrate or group of closely related substrates. For instance, the nitrilase 4 (NIT4) from plants is highly specific for β-cyanoalanine (β(CN)) and shows negligible nitrilase or nitrile hydratase activity against other nitriles (Piotrowski *et al*., 2001). While the majority of nitrilase 1 enzymes from plants show broad substrate specificity, two mustard NIT1's are specific for 2-phenylacetonitrile (2PAN) and 4-hydroxyphenylacetonitrile (4HPAN) respectively (Trompetter, 2010). The cyanide dihydratases, which mainly occur in bacteria, (Meyers *et al*., 1993) exclusively catalyse the hydrolysis of cyanide to formic acid and ammonia (nitrilase reaction). The cyanide hydratases from filamentous fungi catalyse the hydrolysis of cyanide to formamide (nitrile hydratase reaction) (Dumestre *et al*., 1997), but at least in the case of *Fusarium lateritium* (Nolan *et al*., 2003) also show significant nitrilase activity against a range of larger substrates. The activity and substrate specificity of the cyanide hydratases has further been called into question by the discovery of a nitrilase from *Aspergillus niger* K10 (Kaplan *et al*., 2011) that shows a high degree of similarity to a number of cyanide hydratases (Basile *et al*., 2008) but catalyses a range of nitriles to the corresponding acid. Until further work in this direction is done, the cyanide hydratases in general cannot be considered to be specific for cyanide or exclusive hydratases.

The plant nitrilases from group 1 (NIT1, NIT2, NIT3) in general show broad substrate specificity for a variety of aromatic as well as aliphatic substrates. However, certain exceptions exist; *Cr*NIT1a has peak saturated aliphatic nitrile activity against butanitrile, but much more significant activity against the unsaturated 3-butenitrile. *Cr*NIT1b is 85% identical with *Cr*NIT1A but shows a very different specificity

profile; it has broader substrate specificity and a preference for larger substrates with peak activity against heptanitrile/ 6-heptenitrile. *At*NIT1 has peak activities measured for larger substrates still, with a peak aliphatic activity against octanenitrile. (Osswald *et al*., 2002). While the physiological substrate for *At*NIT3 is unknown, aliphatic substrate screening has revealed a preference for very long substrates with a peak saturated aliphatic activity against nonanitrile and even significant activity against dodecanitrile (Fig. 10).

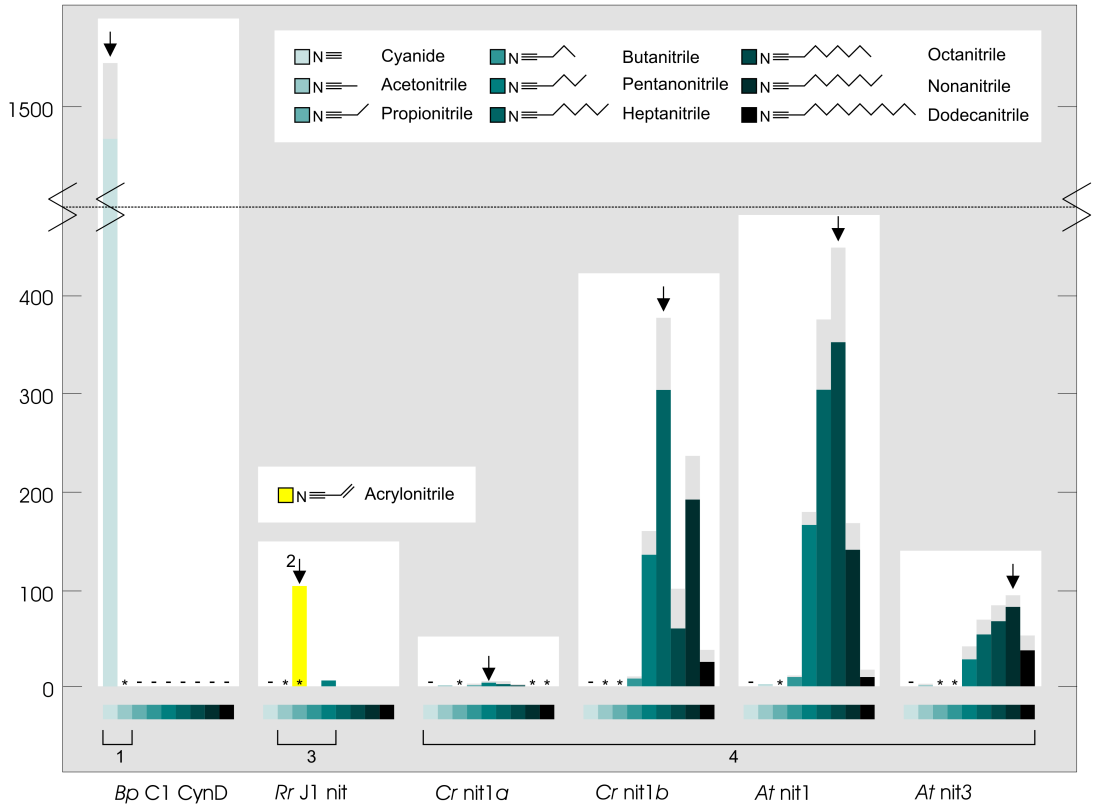


Figure 10. Substrate specificities of a sample of nitrilases that catalyse the conversion of aliphatic nitriles of different length (activities in nKat±SD). An arrow indicates peak activity "*" indicate that no activity above the control was measured, "-" indicates that the substrate wasn't tested. The cyanide dihydratase from *B. pumillus* C1 (1: Meyers *et al*., 1993) is highly specific for cyanide. The nitrilase from *R. rhodochrous* J1 shows negligible activity against saturated aliphatic nitriles (3: Kobayashi *et al*., 1989) except pentanonitrile (~6 nKat/mg protein). This result was explained by 2: Nagasawa *et al*. (2000) by demonstrating that aliphatic nitriles do not promote subunit association and therefore show low specific activity. By preincubating the enzyme with benzonitrile, a higher specific activity for acrylonitrile was measured (105 nKat). A sample of plant nitrilases: *C. rubella* NIT1a, *C. rubella* NIT1b, *A. thaliana* NIT1 and *A. thaliana* NIT3 have been systematically tested for preferred substrate length (4: Trompetter, 2010). A trend can be discerned with *Cr*NIT1a preferring shorter – and *At*NIT3 preferring longer substrates.

There is a strong correlation between the number of carbon atoms of the R-group of the preferred aliphatic substrate and that of the preferred aromatic substrate amongst

those nitrilases that are active against both aliphatic and aromatic substrates. Among the nitrilases under study, this includes the nitrilase from *R. rhodochrous* J1; *A. thaliana* NIT1 and NIT3. *S. alba* NIT1c and *S. arvensis* NIT1c are specific for aromatic nitriles of a specified length (Fig. 11). Nitrilases are sometimes classified according to their substrate specificity into three groups: aromatic, aliphatic and arylacetonitriles (O'Reilly and Turner, 2003; Banerjee *et al*., 2002). However, this classification scheme is of limited utility because of the large overlap in specificities observed among the nitrilases studied here and is therefore not used in this work.

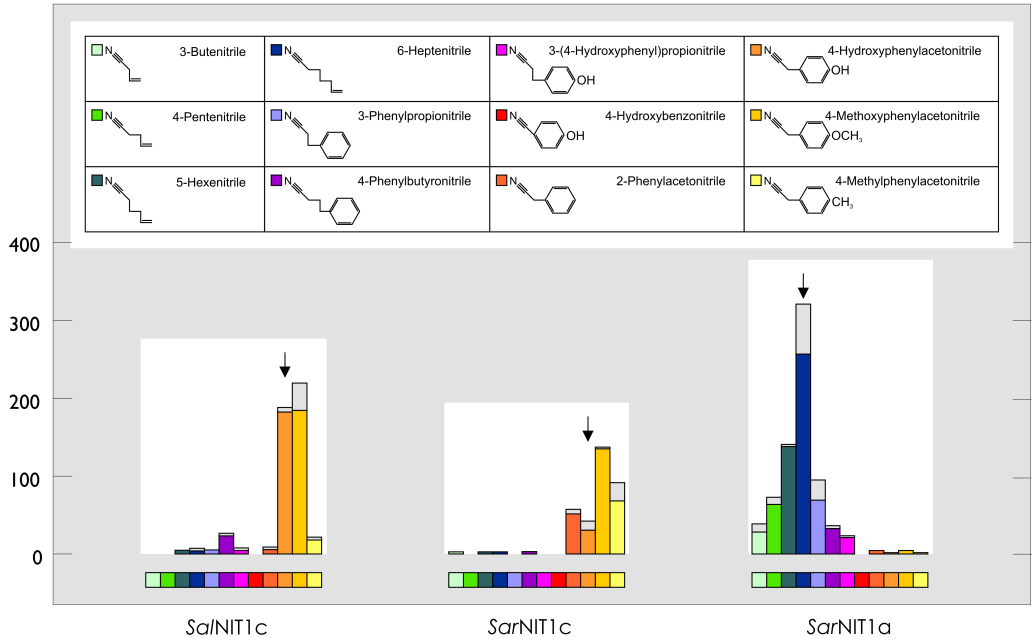


Figure 11. Comparison of the nitrilase activity of mustard nitrilases. *Sal*NIT1c only shows significant activity against 4-hydroxyphenylacetonitrile and 4-methoxyphenylacetonitrile. *Sar*NIT1c is active against aromatic substrates of the correct length. *Sar*NIT1a shows typical broad NIT1 substrate specificity. (Activities in nKat per mg protein±SD). Data courtesy of Inga Trompetter, Ruhr University, Bochum.

1.1.3 Aims

In this section, the following aims will be addressed:

- Determine the three dimensional helical structure of a sample of plant nitrilases at ~2 nm resolution by electron microscopy and image processing.
- Use homology modelling and docking to produce pseudo atomic-resolution models of the plant nitrilases.

- Analyse the amino acid residue composition of the active-site pockets from nitrilases with differing substrate preference to identify determinants of specificity.

In the remainder of this section I will:

- Exchange the amino acids identified from the procedure above by site directed mutagenesis and test the resulting mutant enzymes for changes in substrate specificity.

1.1.4 Brief summary of the findings described in this section

The helical reconstructions produced here have now more than doubled the number of known nitrilase quaternary structures and demonstrated the extent to which nitrilase enzymes vary in their helical structure. The helical twist, defining the angle between adjacent nitrilase dimers across the C-interface, in particular shows a greater degree of variability than previously observed. Interestingly, for a given enzyme, this helical parameter was found to correlate closely with the size of the preferred substrate. This relationship was further explored by site-directed mutagenesis; a single amino-acid exchange was found to not only change the size of the preferred substrate, but also the helical twist.

The helical twist is not the only substrate determinant: in some cases, two nitrilases display the same helical twist and yet, the preferred substrates are different. This effect was explored by a combination of homology modelling and site-directed mutagenesis. I found that active-site pocket residue exchanges between pairs of nitrilases with different preferred substrates decrease the wild-type specific-activity while increasing the specific activity towards the preferred substrate of the other enzyme, but only in cases where the two enzymes display sufficiently similar helical twist. A model for nitrilase 4 binding was proposed on the basis of the C171A/V236A Mutant of N-carbamyl-D-amino acid amidohydrolase complexed with N-carbamyl-D-methionine (1UF5: Hashimoto *et al*., unpublished), this model predicted that β(CN) binds to an arginine in the active site of *Arabidopsis thaliana* nitrilase 4 (*At*NIT4) via its carboxyl group. On the basis of this prediction a novel substrate has been found for *At*NIT4.

1.2 Results

Some discussion has been included in this section in order to improve readability.

1.2.1 Interface disruption

Three interface mutants were constructed by in-vitro mutagenesis: *P. stutzeri* CynD loop mutant (Δ220-234) (Sewell *et al*., 2005), *P. stutzeri* CynD C-terminus truncation (Δ296) (Sewell *et al*., 2005) and *A. thaliana* NIT4/ NIT1 loop exchange (this work) (Fig. 12). The first two mutations (*P. stutzeri* CynD Δ220-234 and *P. stutzeri* CynD Δ296) have been shown to result in enzyme inactivation (Sewell *et al*., 2005). However, it has not been demonstrated that this inactivation is actually the result of oligomer disruption.

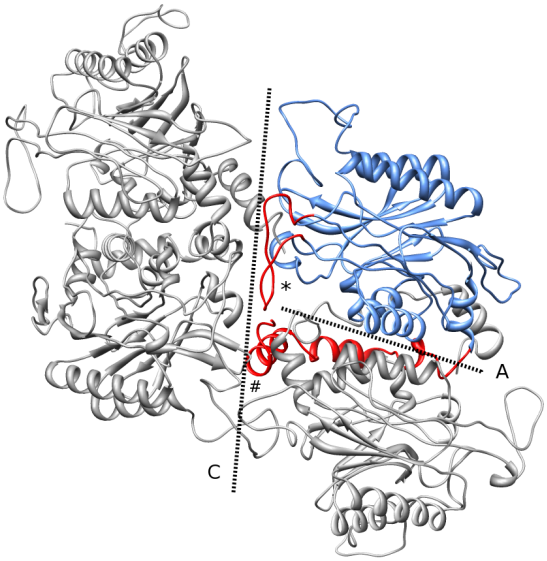


Figure 12. Nitrilase interface disruption mutants. Four nitrilase monomers are shown, pairs of monomers interact across the A-interface (A) to form the conserved nitrilase superfamily dimer. Pairs of dimers interact in turn across the C-interface (C) leading to extension of the helix. A single monomer is shown in blue, the loop (*P. stutzeri* CynD (residues 220-234) indicated in red (*) forms a two-fold symmetric interaction across the C-interface. The C-terminus (*P. stutzeri* CynD Δ296) indicated in red (#) appears to constitute the major stabilising interaction across the A-interface.

These regions were identified as candidates for this work on the basis of homology modelling. The C-interface of *P. stutzeri* CynD and *A. thaliana* NIT4/ *A. thaliana* NIT1 were modelled on the basis of the crystal structure of β-alanine synthase from *D. melanogaster* (2vhi), which makes a homooctomeric helix-like assembly (Lundgren *et al*., 2008). In this structure, a loop lying between β10 and β11 (see Fig. 4) forms a two-fold symmetric interaction across the C-interface between adjacent dimers stabilized by two salt bridges (E298, K306). This interaction appears to be the

major determinant for stabilising the helical association across the C-interface. *P. stutzeri* CynD Δ220-234 expressed in *E-coli* as inclusion bodies, which could be purified, but extensive attempts to refold the protein failed. A *At*NIT4/ *At*NIT1 loop chimera was generated by overlap extension PCR in an attempt to disrupt the interaction across the C-interface, while possibly still allowing the protein to fold correctly. This region shows significantly lower sequence homology (~20%) between *At*NIT1 and *At*NIT4 than the average between the two enzymes (~63%). Both *P. stutzeri* CynD Δ296 and the *At*NIT4 loop exchange mutant expressed solubly in *E-coli* BL21 and showed no evidence of activity. Gel-filtration chromatography indicated that both enzymes mainly form dimers; this was confirmed by electron microscopy and image processing in the case of the *At*NIT4 loop exchange mutant (Fig. 13).

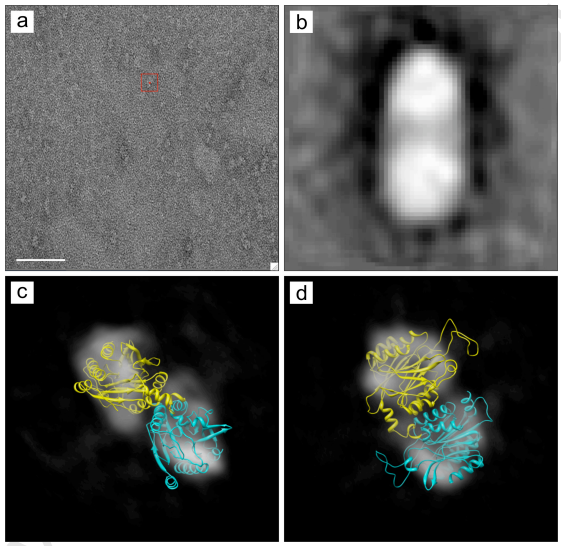


Figure 13. *At*NIT4 with the C-interface loop from *At*NIT1. (**a**) Unprocessed low-dose negative stain micrograph at 50 000× magnification showing individual dimers, one of which is indicated by a red square. Scalebar 50 nm. Very little aggregation and no helical association can be seen in contrast to wild-type *At*NIT4 (Fig. 16). (**b**) Reference-free alignment average of boxed particles from (a), two distinct lobes are visible, side of box: 20 nm. (**c**, **d**) Phantom views of a single particle reconstruction of the dimer from the front and side. A homology model of the nitrilase has been docked into the reconstructed volume. The size and spacing of the two lobes correlates well with the nitrilase A-interface dimer.

1.2.2 The position of the substrate within the binding pocket.

The major determinant of the position of the substrate within the active site is the position of the active-site cysteine. The nitrile forms a covalent thioimidate intermediate as the result of a nucleophilic attack by the cysteine on the nitrile carbon. (see section 2). This reduces the possible orientations that the R-group of the substrate

can take within the binding pocket, to a single degree of freedom: rotation about the carbon-sulphur (C-nitrile, S-cysteine) bond. Steric hinderance further constrains the possible orientations of the R-group. Substrates, especially large substrates (e.g. dodecanitrile), require substantial space within the binding pocket to prevent steric clashes. This further limits the possible orientations of the R-group to angles about the carbon-sulphur bond with sufficient space for a large R-group to access and bind. Further evidence comes from crystal structures of homologous enzymes that have been co-crystallized with trapped substrates, reaction intermediates or inhibitors. Two crystal structures in particular: a C171A/ V236V mutant of N-carbamyl-D-amino acid amidohydrolase complexed with N-carbamyl-D-methionine (1UF5, Hashimoto *et al*., unpublished) and the amidase from *Pseudomonas aeruginosa* with a trapped acyl-transfer reaction intermediate (2UXY, Andrade *et al*., 2007) provide the templates required to model substrate-enzyme interactions.

1.2.3 The role of quaternary structure

It has been previously demonstrated several times and has been demonstrated again here (section 1.2.1) that nitrilases require interface formation for activity. The outer border of the substrate binding pocket is bounded by, in some cases a loop, and in others a short helix, which contains "the second catalytic glutamic acid". This region is connected to the enzyme by two flexible loops, and it is difficult to see how the position of these loops is maintained. The closest homologue to this for which atomic resolution structural information is available is the β-alanine synthase from *Drosophila melanogaster* (2vhi) where these loops are held in place by interacting across the C-interface. At either end of a short helix, the loop is disordered and cannot be seen in the crystal structure. Furthermore, one of the loops interacting across the C-interface appears to act as a brace, limiting the size of the binding pocket. An increase in helical twist would therefore act in the same way as the tightening of a spring – squeezing adjacent subunits against one another and restricting the size of possible substrates. Differences in the preferred size of substrates between different nitrilases might therefore be attributable to rearrangements at the C-interface. Electron microscopy was therefore used to measure the relative orientations of subunits within the nitrilase fibre.

1.2.4 Nitrilase helical reconstruction

Eight nitrilases derived from plants and one from *Caenorhabditis elegans* (but most likely originally derived from plants by horizontal gene transfer) (Griemert, 2010) were subjected to electron microscopy for the first time. All of the nitrilases investigated formed helical assemblies (Fig. 14a, 14b, 14c, 14d; 15a; 16a, 16b, 16c) except for *Sar*NIT1c, which formed "ring-like" structures (Fig. 15d). Negative-stain electron microscopy indicated that the fibres were relatively short (<100 nm) and rather curved, especially *At*NIT1 and *At*NIT3 (Fig. 14c, 14d). The fibres were reconstructed in three-dimensions by the iterative helical real-space reconstruction (IHRSR) algorithm (Egelman, 2000). The resulting helical symmetry of the enzymes varied but mainly in the angle between dimers (Δφ) and ranged from -75° for *Ce*NIT to -66° for *At*NIT3 (Fig. 14d, 15c).

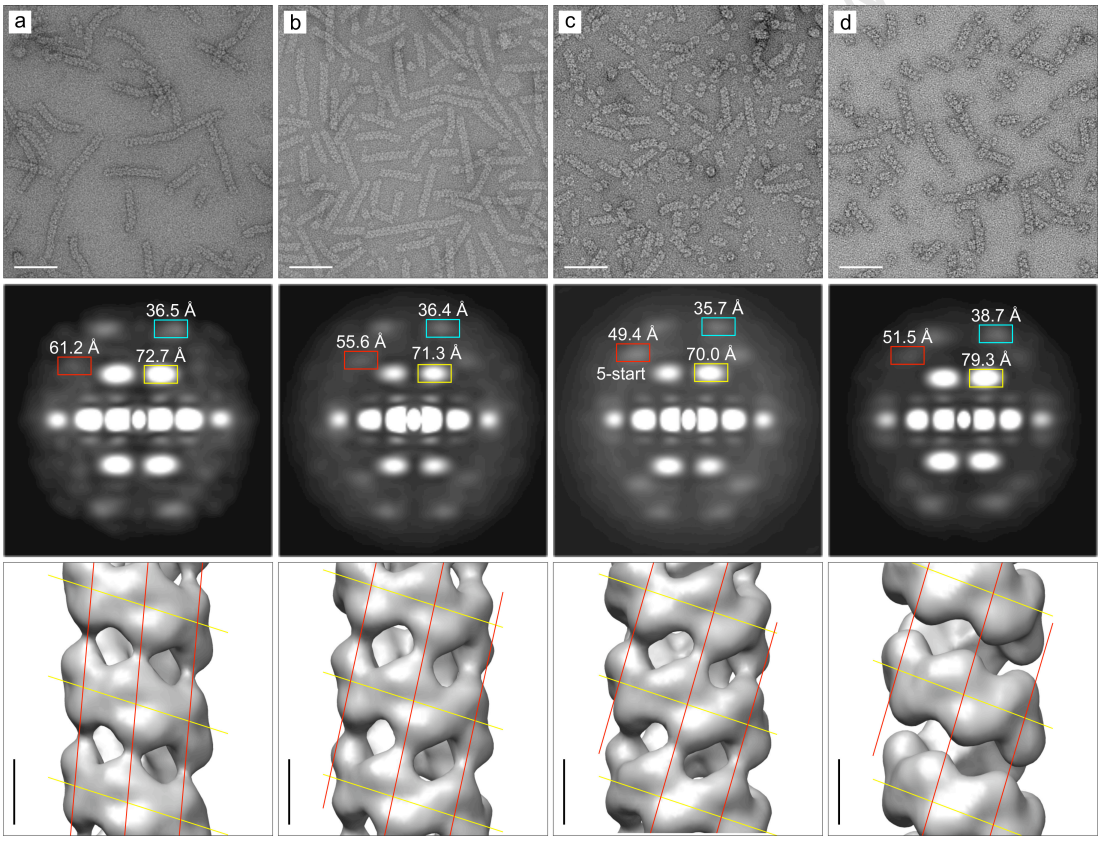


Figure 14. Reconstructions of nitrilase 1 enzymes. For each enzyme, a negative-stain transmission electron microscopy image taken at 50 000× magnification (scalebar: 50 nm) is shown in the upper row. The second row shows the averaged two-dimensional power spectrum calculated from vertically aligned short helical segments. The pitches of the one-start (yellow), five-start (red) and six-start (blue) helices are shown. The third row shows the final reconstructed volume (scalebar: 5 nm). (**a**) *Capsella rubella* nitrilase 1a (*Cr*NIT1a) has a helical rise (Δz) of 14.3 Å and a twist per subunit (Δφ) of -69.9º. (**b**) *Capsella rubella* nitrilase 1b (*Cr*NIT1b) has a helical rise (Δz) of 13.8 Å and a twist per subunit (Δφ) of -68.5º. (**c**) Arabidopsis thaliana nitrilase 1 (*At*NIT1) has a helical rise (Δz) of 13.4 Å and a twist per subunit (Δφ) of -67.1º. (**d**) Arabidopsis thaliana nitrilase 3 (*At*NIT3) has a helical rise (Δz) of 14.7 Å and a twist per subunit (Δφ) of -66.2º.

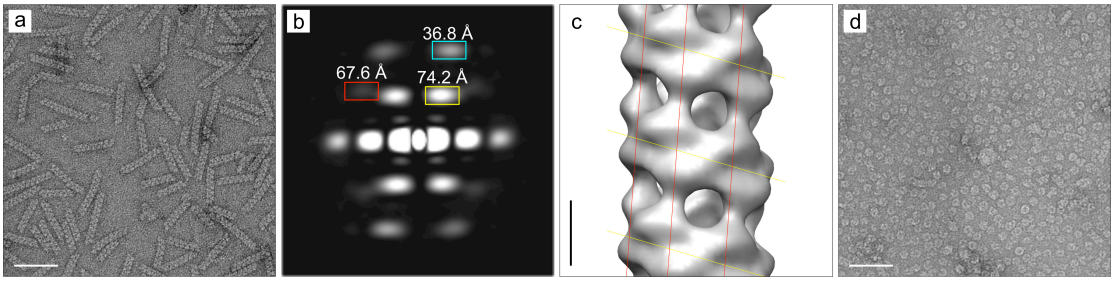


Figure 15. Mustard nitrilase 1 enzymes. (**a**) Negative-stain transmission electron microscopy image of *Sinapis alba* nitrilase 1c (*Sal*NIT1c) at 50 000× magnification. The enzyme forms long helices. Scalebar: 50 nm. (**b**) Averaged two-dimensional power-spectrum calculated from short helical segments. The pitches of the one-start (yellow), five-start (red) and six-start (cyan) layer lines are indicated. (**c**) Reconstructed helical volume of *Sal*NIT1c, the fibre has a rise per subunit (Δz) of 14.8 Å and a twist per subunit (Δφ) of -70.6º, the one-start (yellow) and five-start (red) helices are indicated. Scalebar: 5 nm. (**d**) Negative stain electron microscopy image of *Sinapis arvensis* nitrilase 1c (*Sar*NIT1c) at 50 000× magnification. This enzyme forms rings with a molecular mass of ~300 kDa. This type of nitrilase ring-structure has been investigated previously (Williamson *et al*., 2010) Scalebar: 50 nm.

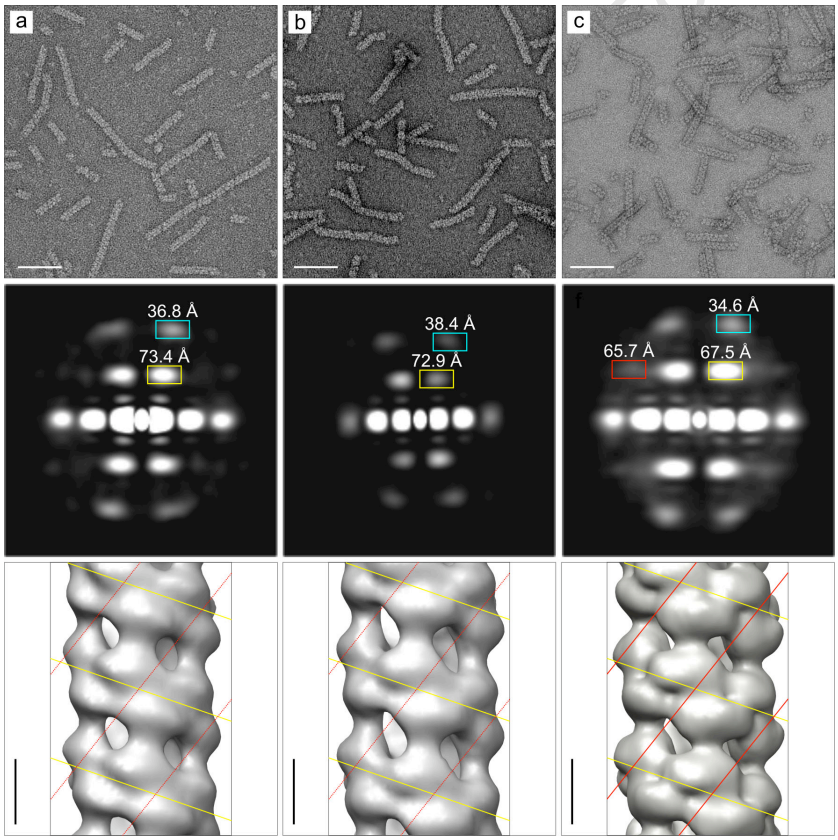


Figure 16. Reconstructions of nitrilase 4 enzymes. First row: negative stain micrographs. Scalebar: 50 nm. Second row: powerspectra calculated from short helical segments, one-start (yellow), five-start (red) and six-start (blue) layer-lines are shown. Last row: reconstructed helical volumes. Scalebar: 5 nm. (**a**) *Caenorhabditis elegans* nitrilase (*Ce*NIT): helical rise (Δz) of 14.3 Å and a twist per subunit (Δφ) of -74.0º. (**b**) *Arabidopsis thaliana* nitrilase 4 (*At*NIT4): helical rise (Δz) of 15.3 Å and a twist per subunit (Δφ) of -73.5º. (**c**) *Lotus japonicus* nitrilase 4A (*Lj*NIT4A): helical rise (Δz) of 14.9 Å and a twist per subunit (Δφ) of -75.1º.

1.2.5 Substrate size vs. helical twist

In order of decreasing twist, (Fig. 14a, b, c, d) the helix opens up, the diameter of the fibre increases and the location of the D-interface shifts. The density across the interface shifts position from above the dimer in the case of *Cr*NIT1a to the right-hand edge in the case of *At*NIT1 to making no contact in the case of *At*NIT3 (Fig. 14a, c, d). Interestingly, if we order the nitrilases according to preferred substrate size (Fig. 10) from smallest to largest preferred substrate we obtain the same order: *Cr*NIT1a (3-butenitrile), *Cr*NIT1b (6-heptenitrile), *At*NIT1 (octanitrile) and *At*NIT3 (nonanitrile). In fact, for all nitrilases tested a clear correlation is observed between Δφ and the length of the R-group of the preferred substrate (Fig. 17). Furthermore, it was observed that the greater the twist of the enzyme (and smaller the substrate) the more specific the enzyme tends to be. On the two extremes, *Bp*CynD is highly specific for cyanide while *At*NIT3 has broad substrate specificity (Fig. 10).

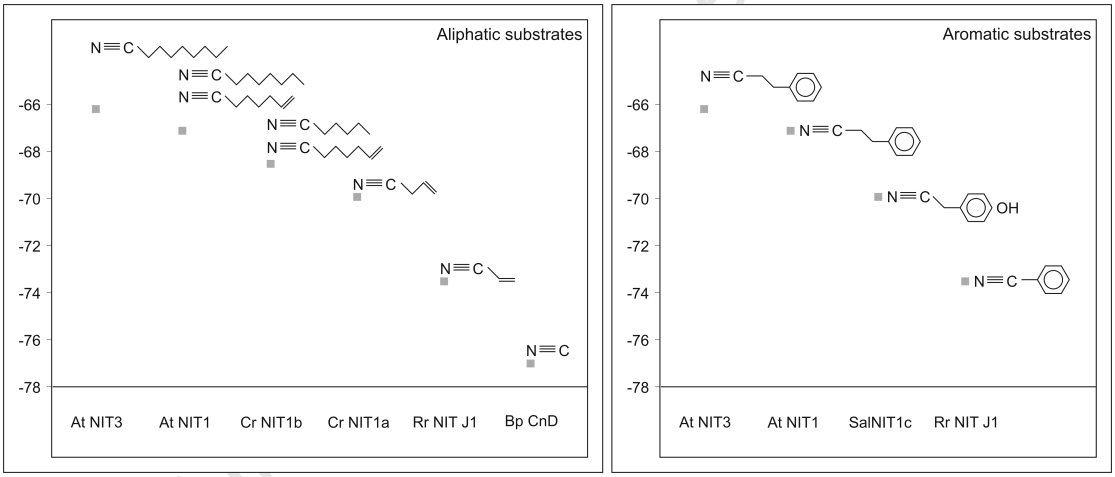


Figure 17. Substrate size against helical twist for all nitrilase enzymes that produce an acid product. Nitrilases that favour larger substrates have a strong tendency to have a smaller twist. *At*NIT3 has a generally low activity against a wide range of different substrates but has a tendency to prefer larger substrates (Fig. 10). *At*NIT1 has peak activity against the aliphatic substrates octanitrile and the aromatic substrate 2-phenylacetonitrile (Trompetter, 2010). *Cr*NIT1b has a tendency to prefer slightly shorter substrates, with highest activity being measured against 6-heptenitrile, whereas *Cr*NIT1a (which is 88% identical) prefers 3-butenitrile. *Sal*NIT1c is specific for 4-hydroxyphenylacetonitrile. The precise activity spectrum of *Rr* (J1) nitrilase has not been measured, but Nagasawa *et al*. (2000) found high activity against acrylonitrile. The cyanide dihydratase *Bp*CnD has the smallest substrate (cyanide) and the largest twist.

1.2.6 Exchanging the helical twist

These results led to the formulation of the hypothesis that the angle between adjacent subunits results in a change in active-site pocket size, which in turn results in a change

in the size of the preferred substrate. If this were actually the case, we would expect that any mutation that results in a change in the helical twist of a plant nitrilase would also be associated with a corresponding change in substrate specificity. *Cr*NIT1a has a preference for 3-butenitrile; while *Cr*NIT1b prefers 6-heptenitrile (Fig. 10); they have very high sequence identity (88%) and no amino acid residue differences within 10 Å of the active site (Fig.19, 20) *Cr*NIT1a has a Δφ of -69.9º, while *Cr*NIT1b has a Δφ of -68.5º (Fig. 14a, b). In order investigate whether or not any specific amino acid changes were responsible for specifying Δφ, chimeric enzymes were generated between *Cr*NIT1a and *Cr*NIT1b by *in vitro* mutagenesis and screened for activity with 3-butenitrile and 6-heptenitrile (Fig. 18).

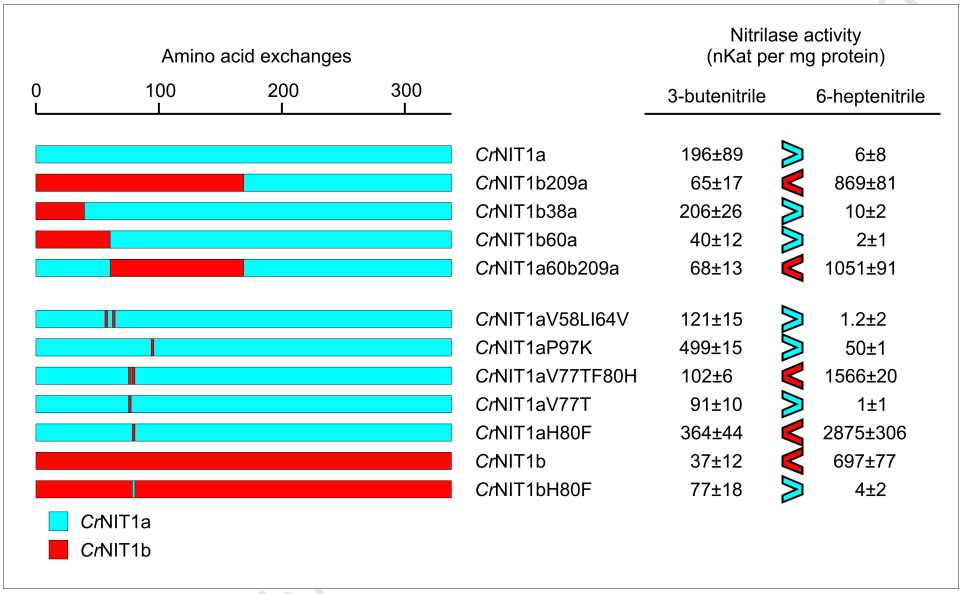
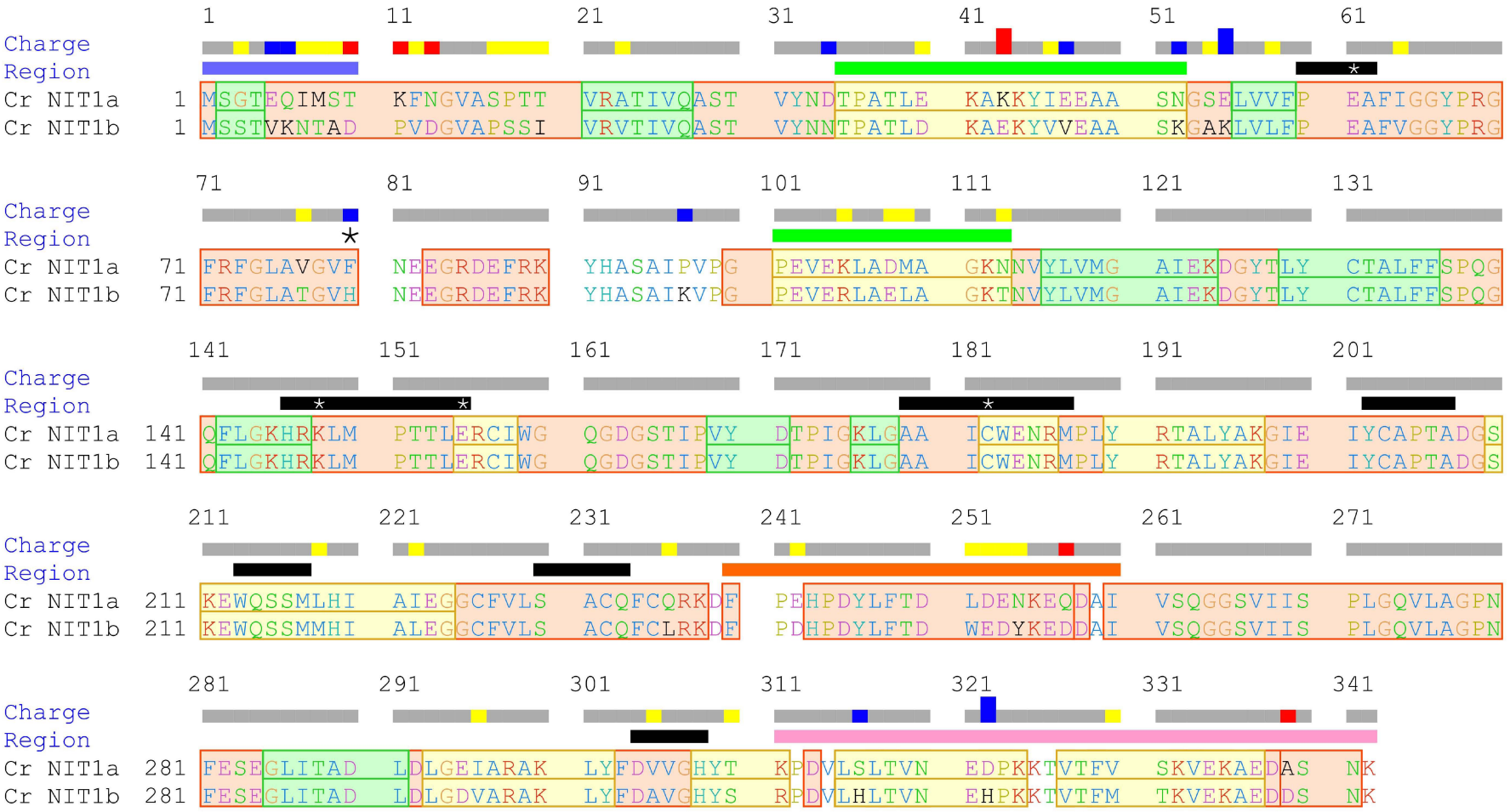


Figure 18. (after Trompetter, 2010). A series of *Cr*NIT1a/ *Cr*NIT1b chimeras were produced by overlap extension PCR and site directed mutagenesis. On the left-hand-side is a list of the mutants. Red (*Cr*NIT1b) and cyan (*Cr*NIT1a) indicate how the chimeras were constructed. Each mutant was tested for specific activity against 3-butenitrile and 6-heptenitrile. The activities are given in nKat per mg protein (mean±SD). Preference for 3-butenitrile or 6-heptenitrile is indicated by > or <. A single amino acid residue, the "substrate specificity residue" appears to be responsible for determining the preferred substrate. Identification of this residue was performed in collaboration with Inga Trompetter from Ruhr University in Germany who performed all of the molecular biology, protein purification and substrate screening for the identification of this residue.

Figure 19. (**overleaf**). Comparison of the primary amino acid structures of *Cr*NIT1a and *Cr*NIT1b. Differences are indicated by yellow (neutral exchange), red (increased acidity) and blue (increased basicity). Different regions are coloured according to their position in the nitrilase homology model, green: D-interface, black: active-site, orange: C-interface loop and pink: C-terminus. The residue identified is indicated (*), a histidine in this position produces an enzyme with a preference for 6-heptenitrile, while a phenylalanine results in a preference for 3-butenitrile, regardless of the other exchanges.


Charge
Region
Cr NIT1a
Cr NIT1b

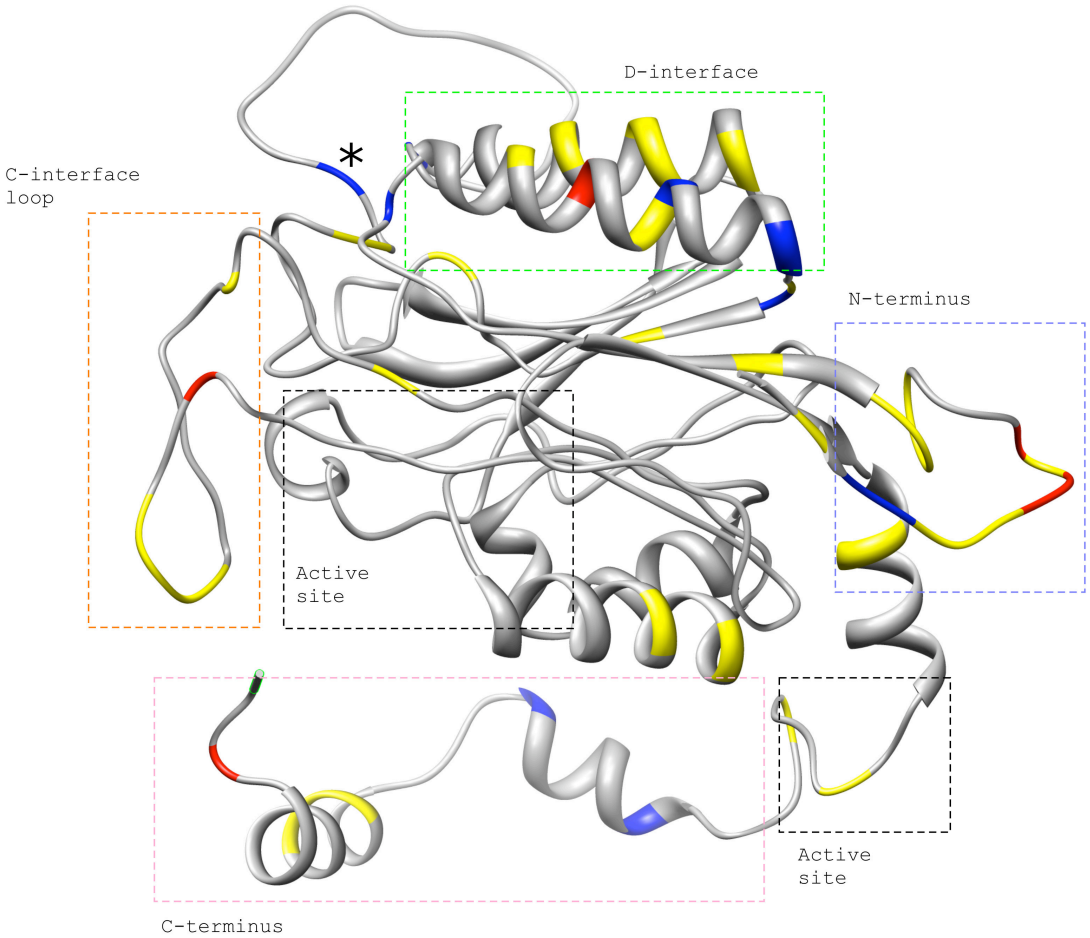


Figure 20. Homology model of *Cr*NIT1a showing the sites of exchange between *Cr*NIT1a and *Cr*NIT1b. Interestingly, most of the differences occur at the periphery of the enzyme, mostly in locations that contribute to interface formation. No amino acid exchanges occur within the active site pocket. The "substrate specifying residue" indicated by (*) is found on an insertion that occurs in nitrilases but none of the other nitrilase superfamily members.

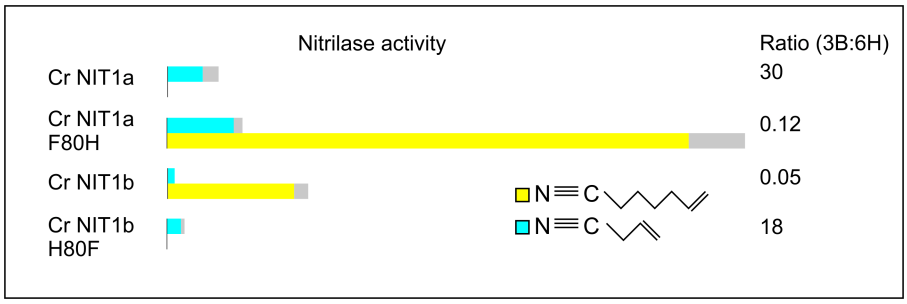


Figure 21. Nitrilase activity (nKat) measured for 6-heptenitrile (yellow) and 3-butenitrile (cyan). *Cr*NIT1a has a distinct preference for shorter substrates with a high 3-butenitrile to 6-heptenitrile activity ratio (3BUT:6H). *Cr*NIT1b on the other hand prefers longer substrates and has a low 3BUT:6H. The mutants however, are exchanged: *Cr*NIT1a F80H has a substrate preference that resembles *Cr*NIT1b while *Cr*NIT1bH80F resembles *Cr*NIT1a.

Wild type *Cr*NIT1a has a Δφ of -69.9º and a preference for shorter substrates (3BUT: 196±89, 6H: 6.4±8.4), while wild type *Cr*NIT1b has a Δφ of -68.5º and a preference for longer substrates (3BUT: 37.1±12, 6H: 697±77) (Fig. 17). But the mutants, *Cr*NIT1a F80H (3BUT: 364±44, 6H: 2870±300) and *Cr*NIT1b H80F (3BUT: 77±19, 6H: 4.2±1.5) have an exchanged substrate preference compared to the wild type enzymes (specific activity measured in nKat per mg protein±SD) (Fig. 21). Negative stain electron microscopy and three-dimensional reconstruction were therefore performed to determine whether or not these mutants also display an exchange in Δφ. The mutants still associate to form helices (Fig. 22), the IHRSR algorithm was initiated with the wild-type helical parameters and iterated until the algorithm converged on a stable helical symmetry.

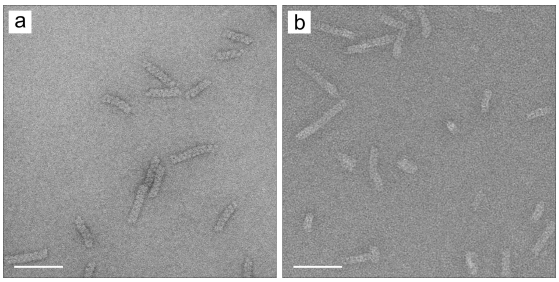


Figure 22. Low-dose negative stain electron microscopy image at 50 000× magnification, scale bar: 50 nm. (**a**) *Cr*NIT1a F80H. (**b**) *Cr*NIT1b H80F.

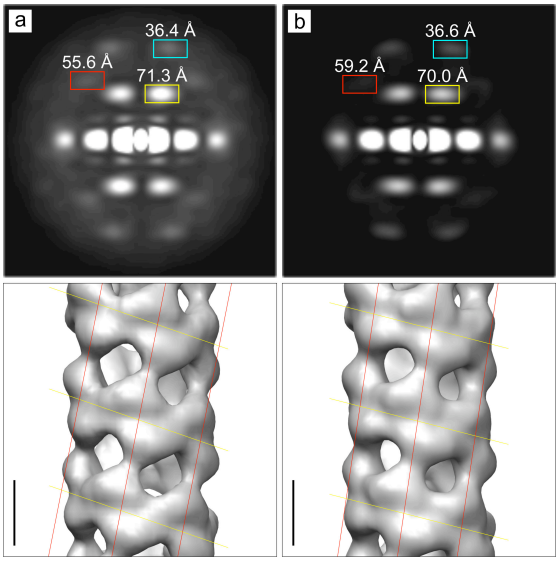


Figure 23. First row: averaged power spectra calculated from short vertically aligned helical segments. The one-start (yellow), five-start (red) and six-start (blue) layer-lines are shown. Second row: Three dimensional helical reconstructions, the one-start (yellow) and five-start (red) helices are shown scale bar: 5 nm. (**a**) *Cr*NIT1b (Δz = 13.8 Å, Δφ = -68.5º) (**b**) *Cr*NIT1b H80F (Δz = 13.9 Å, Δφ = -69.5º).

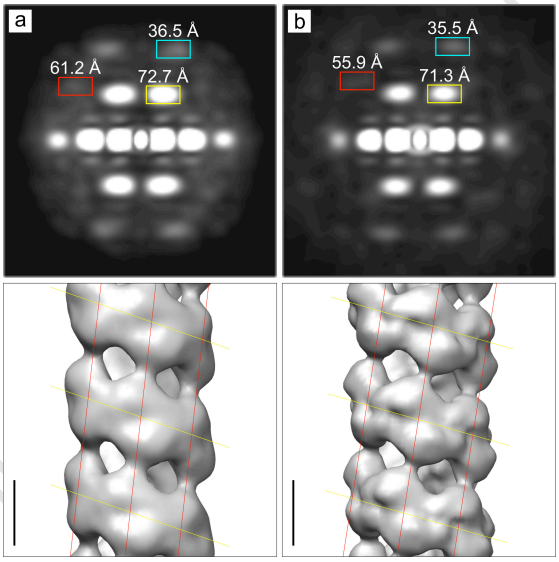


Figure 24. First row: averaged power spectra calculated from short vertically aligned helical segments. The one-start (yellow), five-start (red) and six-start (blue) layer-lines are shown. Second row: Three dimensional helical reconstructions, the one-start (yellow) and five-start (red) helices are shown scale bar: 5 nm. (a) *Cr*NIT1a (Δz = 14.3 Å, Δφ = -69.9º) (b) *Cr*NIT1aF80H (Δz = 13.7 Å, Δφ = -69.0º).

Mutants of *Cr*NIT1a and *Cr*NIT1b show changes in Δφ in the direction expected according to the proposed relationship between substrate size and helical twist (Fig. 17). The mutant of *Cr*NIT1b increased in helical twist by ~1° (*Cr*NIT1b Δφ = -68.5º; *Cr*NIT1b H80F Δφ = -69.5º) and the mutant of *Cr*NIT1a decreased in helical twist by ~1° (*Cr*NIT1a Δφ = -69.9º; *Cr*NIT1aF80H Δφ = -69.0º). This is illustrated in (Fig. 25).

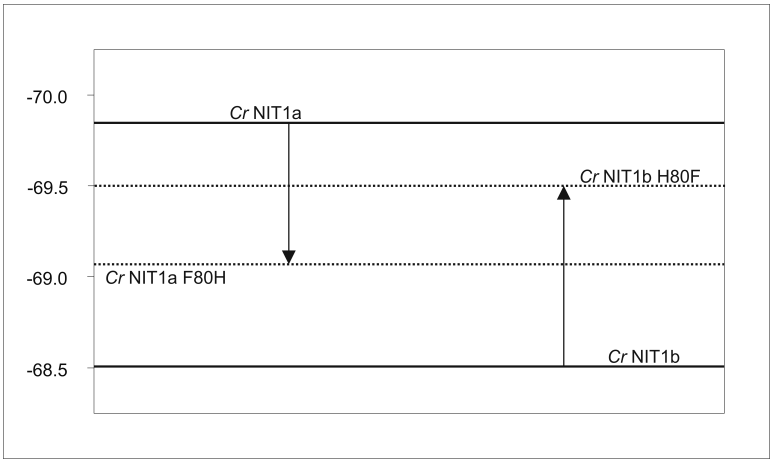


Figure 25. Change in helical twist of *Capsella rubella* nitrilase 1 mutants. *Cr*NIT1a (Δφ = -69.9º) decreases by ~1° in *Cr*NIT1a F80H (Δφ = -69.0º). *Cr*NIT1b (Δφ = -68.5º) on the other hand increases in twist by ~1° in *Cr*NIT1bH80F (Δφ = -69.5º).

Note that while the helical symmetry has not been completely exchanged, neither has the substrate preference. This effect is illustrated in Fig. 26.

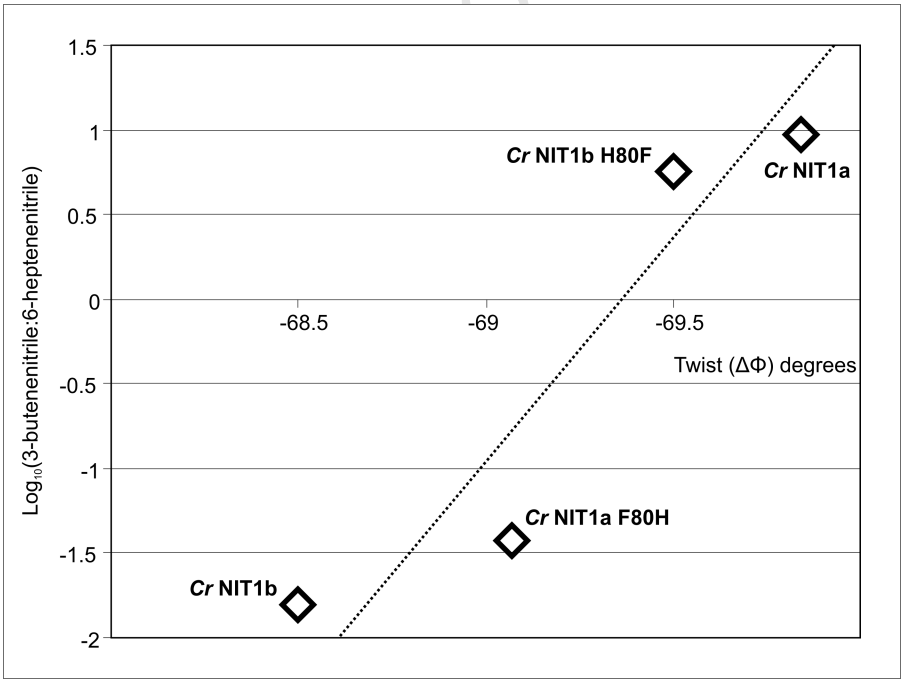


Figure 26. Log-linear plot of the ratio of 3-butenitrile activity to 6-heptenitrile activity (3BUT:6B) as a function of helical twist (Δφ). While not exhibiting a strictly linear relationship, *Cr*NIT1a and *Cr*NIT1b H80F cluster together around a high ratio of 3BUT:6H and large twist. *Cr*NIT1b and *Cr*NIT1a F80H both cluster around a low 3BUT:6H ratio and small twist.

1.2.7 Interface formation and specificity

The overall size of the preferred substrate is correlated with the angle between adjacent dimers across the C-interface. This may be due to changes in compression of the enzyme as helical twist varies (Fig. 27). This may have the effect of changing the volume of the substrate-binding pocket, as substrates bind in the region of the interface between adjacent subunits (Fig. 28).

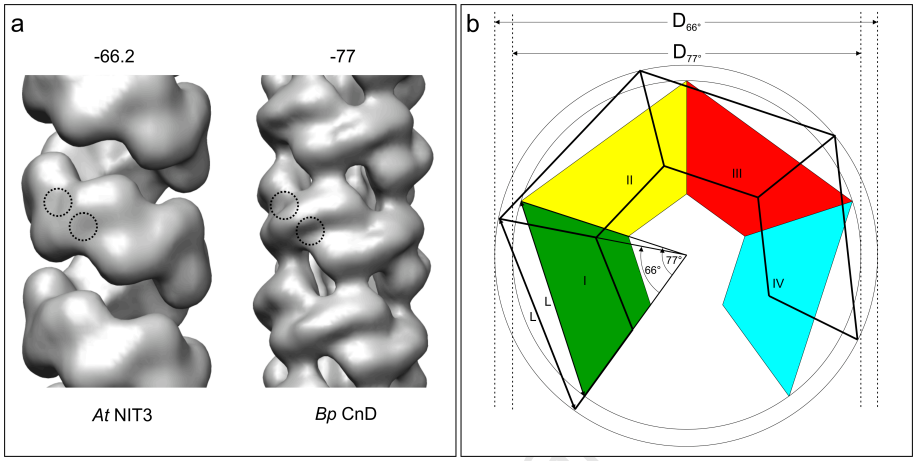


Figure 27. The effect of increasing helical twist on nitrilase enzymes. (**a**) Comparison of the helical structure of *At*NIT3 with a preference for large substrates and *Bp*CnD with a preference for short substrates. The position of the active site, lying on either side of the two-fold symmetric C-interface is shown (dashed circles). (**b**) Simplified model of the nitrilase helix, viewed from above. Two nitrilase fibres are superimposed, showing four dimers. The first, representing a nitrilase with a $\Delta\varphi = 77°$ is shown (I, II, III and IV) in green, yellow, red and cyan respectively. The second with a $\Delta\varphi = 66°$ is shown in outline. *Bp*CnD has a smaller diameter and each subunit is more compressed at the interface between adjacent subunits (C-interface).

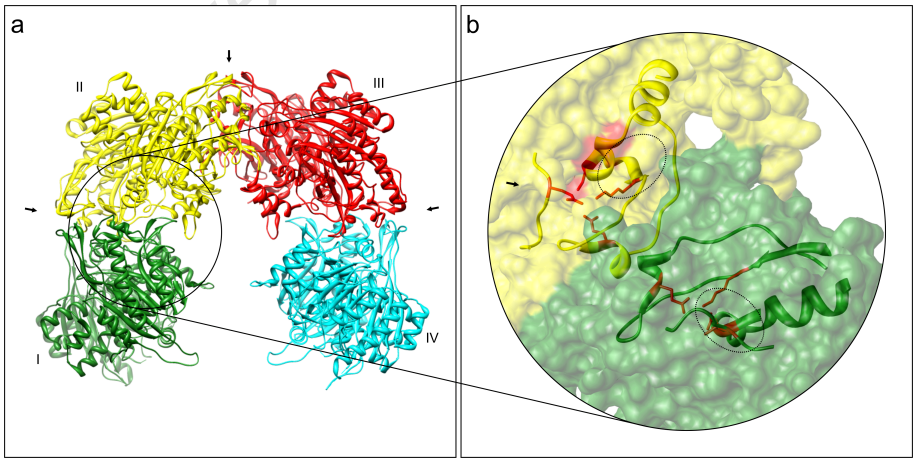


Figure 28. The effect of helical twist on the active site pocket of nitrilase enzymes. (**a**) Four subunits (dimers) of the crystal structure of beta-alanine synthase from *Drosophila melanogaster* (Lundgren *et al*., 2008) are shown in green, yellow, red and cyan (as in Fig. 27). Arrows indicate the two-fold symmetric C-interface, the active site occurs at the interface of two subunits. (**b**) Detail from (a), the active site residues are indicated in red and the position of the two active site pockets are shown by dashed circles. Increasing the angle between two subunits (Fig. 27) compresses the enzyme in the region of the substrate-binding pocket, limiting its extent.

1.2.8 Active-site pocket mutations

In sections 1.2.5, 1.2.6 and 1.2.7 the role of helical twist in defining substrate specificity is shown to have some generality for a sample of plant nitrilases. This effect is not the only determinant of substrate specificity, however. Two plant nitrilases were identified that have the same helical twist and yet exhibit distinct substrate preference. The first is *Capsella rubella* nitrilase 1a (*Cr*NIT1a); this enzyme has a helical twist (Δφ) of -69.9º and a preference for 3-butenitrile. The second is *Sinapis alba* nitrilase 1c (*Sal*NIT1c), (Δφ = -70.6º) with a preference for 4-hydroxyphenylacetonitrile (Fig. 29). These two enzymes differ by ~0.7º and yet show a completely different substrate preference. Furthermore, while the twist appears to define the overall size of the active-site pocket, the details of the interactions that influence binding are likely to be specified by the residues lining the substrate-binding pocket.

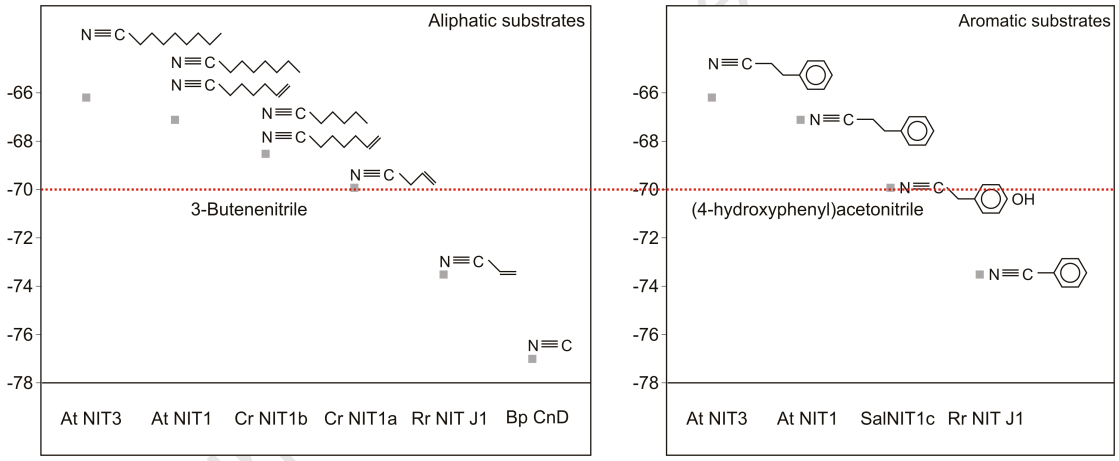


Figure 29. Substrate specificity as a function of helical twist for a sample of plant nitrilases. Only the preferred substrate (aliphatic or aromatic) is indicated, but in most cases these enzymes have a range of acceptable substrates. A red line indicates a helical twist of approximately -70°, *Cr*NIT1a has a twist just below this value (-69.9°) while *Sal*NIT1c has a helical twist just above this value (-70.6°). *Cr*NIT1a has a specific activity of only 1.87 nKat per mg protein against 4-hydroxyphenylacetonitrile (4HPAN) compared to 89.6 nKat per mg protein with 3-butenitrile (3BUT). *Sal*NIT1c on the other hand has a specific activity of 498 nKat per mg protein with 4HPAN compared to just 1.75 nKat per mg protein with 3BUT.

Amino acid residues lining the active site pocket were identified by homology modelling on the basis of the C171A/ V236A mutant of N-carbamyl-D-amino acid amidohydrolase with bound N-carbamyl-D-methionine (1Uf5: Hashimoto *et al.*, unpublished); the amidase from *Pseudomonas aeruginosa* with trapped acyl-transfer

reaction intermediate (2UXY: Andrade *et al*., 2007) and a C165A mutant of the amidase from *Nesterenkonia* AN1 with propionamide trapped in the active site pocket (S. Kimani, unpublished). Homology modelling of the nitrilases under investigation identified four regions that line the substrate-binding pocket (Fig. 30).

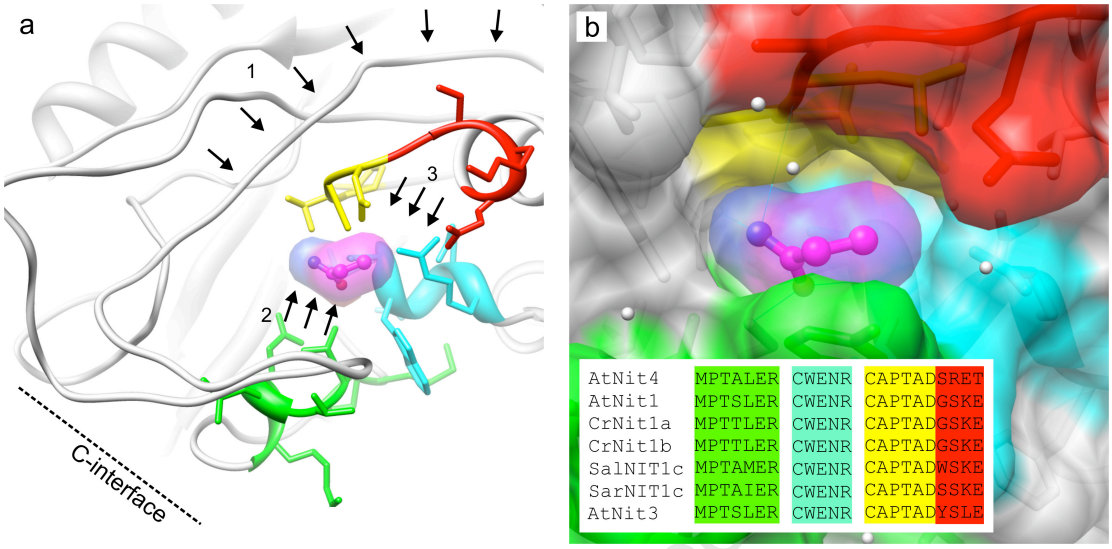


Figure 30. Nitrilase regions in close proximity to the bound substrate. (a) Homology model of *Cr*NIT1a with superimposed propionamide from the active site of the C165A mutant of the amidase from *Nesterenkonia* AN1. Regions that interact directly with the substrate are coloured green (α4), cyan (loop before α5), yellow (loop after β8) and red (α6). Increasing the helical twist most likely results in the movements illustrated by 1, 2 and 3. (b) Space-filling view of the active site pocket of *Cr*NIT1a with bound propionamide. Three of the four regions identified vary between different plant nitrilases (inset).

Two amino acid motifs differ between *Cr*NIT1a and *Sal*NIT1c, the first shown in green (Fig. 30) defines the outer edge of the substrate-binding pocket. In the case of *Cr*NIT1a this motif reads: MPTTLER while *Sal*NIT1c reads: MPTAMER. The second region in red (Fig. 30) differs by a single amino acid substitution reading: GSKE (*Cr*NIT1a) or WSKE (*Sal*NIT1c). These motifs were exchanged by site directed mutagenesis between *Cr*NIT1a and *Sal*NIT1c; the effect of each mutation on substrate specificity was assessed by measuring the specific activity of the mutants against 4-hydroxyphenolacetonitrile and 3-butenitrile. Exchanging the residues in the active site between *Cr*NIT1a and *Sal*NIT1c by site directed mutagenesis has the general effect of decreasing activity for the wild-type substrate and increasing activity against the preferred substrate of the other enzyme (Fig. 31, 32).

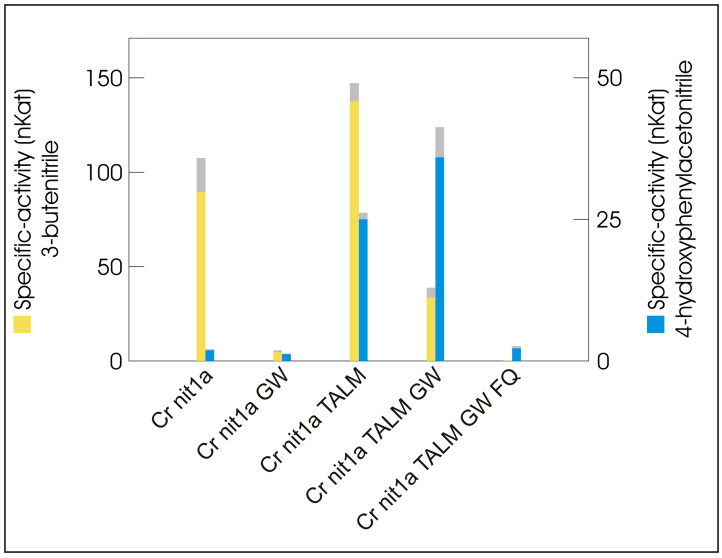


Figure 31. *Cr*NIT1a to *Sal*NIT1c exchanges. (specific activity is measured in nKat per mg protein±SD). Wild type *Cr*NIT1a shows a specific activity of 89±18 against 3BUT, but only 1.9±0.11 against 4HPAN. *Cr*NIT1a G209W decreases the activity for 3BUT, with no corresponding increase in 4HPAN. *Cr*NIT1a T153A L154M increases total activity, but increases the activity of 4HPAN preferentially (3BUT: 137±10, 4HPAN: 25±1.0). Exchanging both regions has a combined effect: 3BUT activity is reduced compared to the WT (33±5.0) and 4HPAN is relatively high (36±5.0). Finally, exchanging the helical symmetry determining loop (*Cr*NIT1aF80Q) resulted in a dramatic decrease in total activity, no activity against 3BUT was measured and very low activity 2.2±0.32 was measured against 4HPAN.

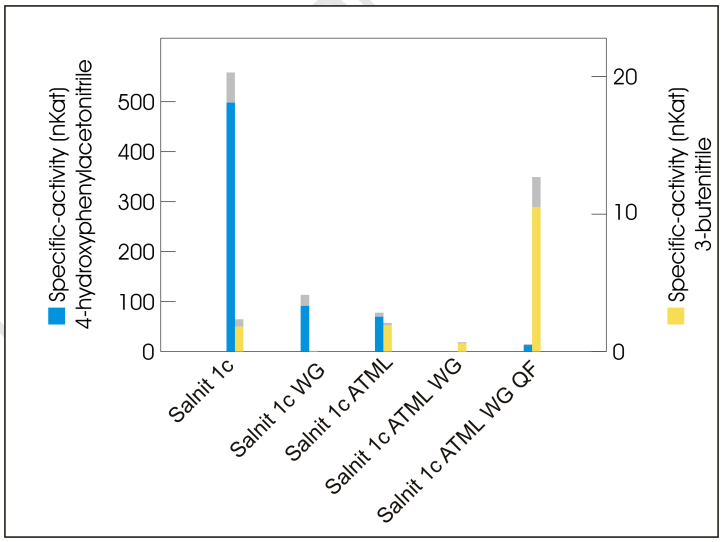


Figure 32. *Sal*NIT1c to *Cr*NIT1a exchanges. (specific activity is measured in nKat per mg protein±SD). A specific activity of 498±60 against 4HPAN and 1.8±0.48 was measured for *Sal*NIT1c WT. *Sal*NIT1c W209G showed a decrease in total activity. *Sal*NIT1c A153T M154L showed a decrease in specific activity for 4HPAN (69.9±7.9) while not affecting 3BUT activity levels. The double mutant virtually eliminates 4HPAN activity (0.20±0.07) but also decreases activity against 3BUT, but not to the same extent. The final quadruple mutant (*Sal*NIT1c A153T M154L W209G F80Q) increases the total enzyme activity, but favours 3BUT activity.

This analysis was extended to a large sample of plant nitrilases with known specificity (Fig. 33). In the green (α4) (Fig. 30) position the nitrilase 4 group generally contains the motif: MPTALER, whereas the nitrilase 1 group have: MPTSLER (Fig. 33). In the cyan position (loop before α5) (Fig. 30), nitrilase 4 enzymes have either an arginine or a lysine in the last position; nitrilase 1 enzymes always have an arginine (Fig. 33). The red motif (α6) (Fig. 30) is the most interesting: nitrilase 4 enzymes have a distinct S/A R E/D T/V motif, while nitrilase 1 enzymes (with typical nitrilase 1 activity) have: GSKE. Nitrilase 3 enzymes have: YSLE and *Sal*NIT1c and *Sar*NIT1c have WSKE and SSKE respectively (Fig 33).

```
AtNIT4     MPTALER  CWENR  CAPTADSRET
ClNIT4     MPTGLER  CWENR  CAPTADAREV
CpNIT4     MPTALER  CWENK  CAPTADSRET
CrNIT4     MPTALER  CWENR  CAPTADARET
CsNIT4     MPTALER  CWENR  CAPTADARET
GmNIT4     MPTALER  CWENR  CAPTADARDV
VvNIT4     MPTSLER  CWENR  CAPTADARDI
PtNIT4     MPSSLER  CWENR  CAPTADARES
SmNIT4-1   VPTAAER  CWENR  CAPTADSRET
SmNIT4-2   VPTAAER  CWENR  CAPTADSREA
TmNIT4     MPTGLER  CWENR  CAPTADARPL
McNIT4     MPTALER  CWENR  CAPTADSREV
MeNIT4-1   MPTATER  CWENR  CAPTADARDT
MeNIT4-2   MPTATER  CWENR  CAPTADARDT
LsNIT4     MPTALER  CWENK  CAPTADSREI
LuNIT4     MPTSVER  CWENR  CAPTADSRPT
LeNIT4     MPTALER  CWENR  CAPTADAREV
MtNIT4     MPTALER  CWENK  CAPTADSREL
PotrNIT4   MPTAVER  CWENR  CAPTADSRDT
PpNIT4     MPTASER  CWENR  CAPTADARRS
ZmNIT4A    MPTALER  CWENK  CAPTADSRPV
SoNIT4A    MPTALER  CWENK  CAPTADSRPV
SbNIT4A    MPTALER  CWENK  CAPTADSRPV
LaNIT4A    MPTALER  CWENK  CAPTADSREV
TaNIT4A    MPTALER  CWENK  CAPTADSRPV
LjNIT4A    MPTAMER  CWENR  CAPTADAREV
OsNIT4A    MPTALER  CWENK  CAPTADSRQV
NtNIT4A    MPTALER  CWENR  CAPTADSRDV
HvNIT4A    MPTALER  CWENK  CAPTADSRPV
BrNIT-T2   MPTSLER  CWENR  CAPTADGSKE
BrNIT-T1   MPTSLER  CWENR  CAPTADGSKE
AtNIT2     MPTSLER  CWENR  CAPTADGSKE
AtNIT1     MPTSLER  CWENR  CAPTADGSKE
BnNIT1B    MPTSLER  CWENR  CAPTADGSKE
BnNIT1A    MPTSLER  CWENR  CAPTADGSTE
SalNIT1a   MPTSLER  CWENR  CAPTADGSTE
SarNIT1a   MPTSLER  CWENR  CAPTADGSKE
SarNIT1b   MPTSFER  CWENR  CAPTADGSKE
CrNIT1A    MPTTLER  CWENR  CAPTADGSKE
CrNIT1B    MPTTLER  CWENR  CAPTADGSKE
SalNIT1c   MPTAMER  CWENR  CAPTADWSKE
SarNIT1c   MPTAIER  CWENR  CAPTADSSKE
CrNIT3     MPTSLER  CWENR  CAPTADYSLE
AtNIT3     MPTSLER  CWENR  CAPTADYSLE
```

Figure 33. Comparison of amino-acid motifs surrounding the binding pocket of plant nitrilases See appendix 1 for references. The three-dimensional position of the three regions relative to the substrate is shown in (Fig. 30). All of the nitrilase 4 enzymes included here (including NIT4A) demonstrate high specific activity for β-cyanoalanine (Janowitz *et al*., 2009). The nitrilase 1 enzymes display typical broad substrate specificity, except for *Cr*NIT1a (3-butenitrile), *Cr*NIT1b (6-heptenitrile), *Sal*NIT1c (4-hydroxyphenylacetonitrile), *Sar*NIT1c (2-phenylacetonitrile) and nitrilase 3 (nonanitrile).

The active site pocket residues were exchanged between *Sal*NIT1c and *At*NIT4, a slight increase in 4HPAN activity was observed in *At*NIT4 S224W R225S E226K T227E, with no corresponding decrease in β-cyanoalanine (β(CN)) activity (Fig. 34). Exchanging the active site pocket residues of *Sal*NIT1c with *At*NIT4 (M154L W209S S210R K211E E212T) inactivated the enzyme (Fig. 35).

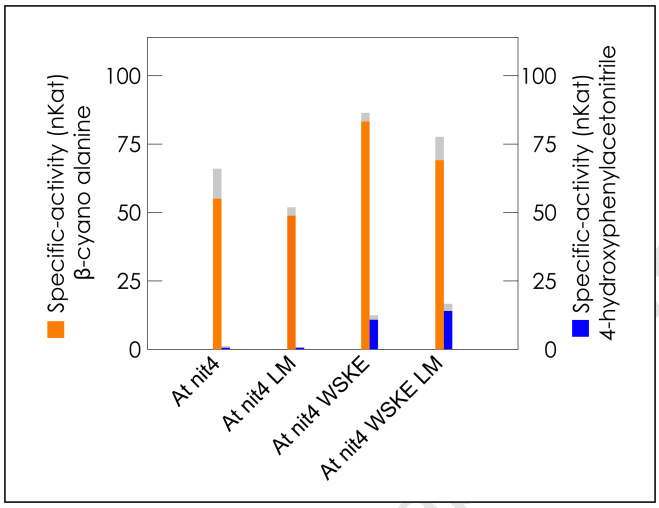


Figure 34. *At*NIT4 to *Sal*NIT1c active-site pocket exchanges. (specific activity is measured in nKat per mg protein±SD). *At*NIT4 has a specific activity of ~50-75 against β(CN) this is not altered by changing the active site pocket residues identified here. 4HPAN activity however, is increased in *At*NIT4 S224W R225S E226K T227E and *At*NIT4 S224W R225S E226K T227E L169M to 16.2±2.4 and 20.9±3.9 respectively.

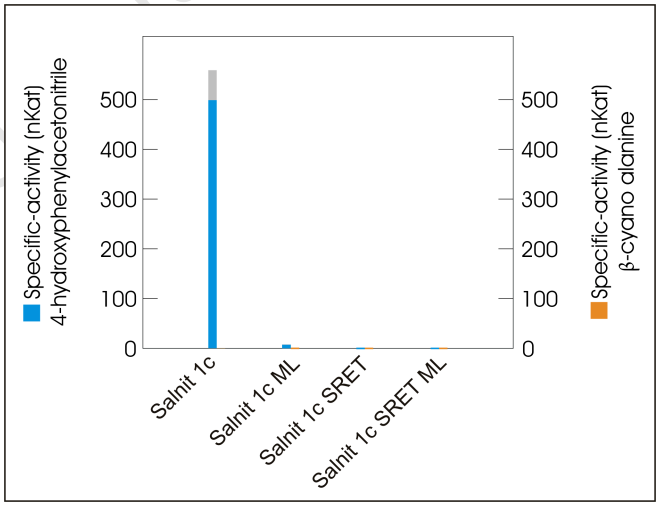


Figure 35. *Sal*NIT1c to *At*NIT4 exchanges. Wild-type *Sal*NIT1c shows no activity for β(CN) and exchanging the active-site pocket residues to those found in *At*NIT4 does not result in an increase in activity against β(CN) but results in a dramatic decrease in 4HPAN activity.

The effect of individual motifs on the specific activity is difficult to interpret however, because of the large number of variables involved. To better understand the effect of each motif on specific activity, the effects of the two regions (α4 and α6) are separated (Fig. 35, 36). For each region, the association between a particular motif and substrate is illustrated for different wild type – as well as mutant enzymes.

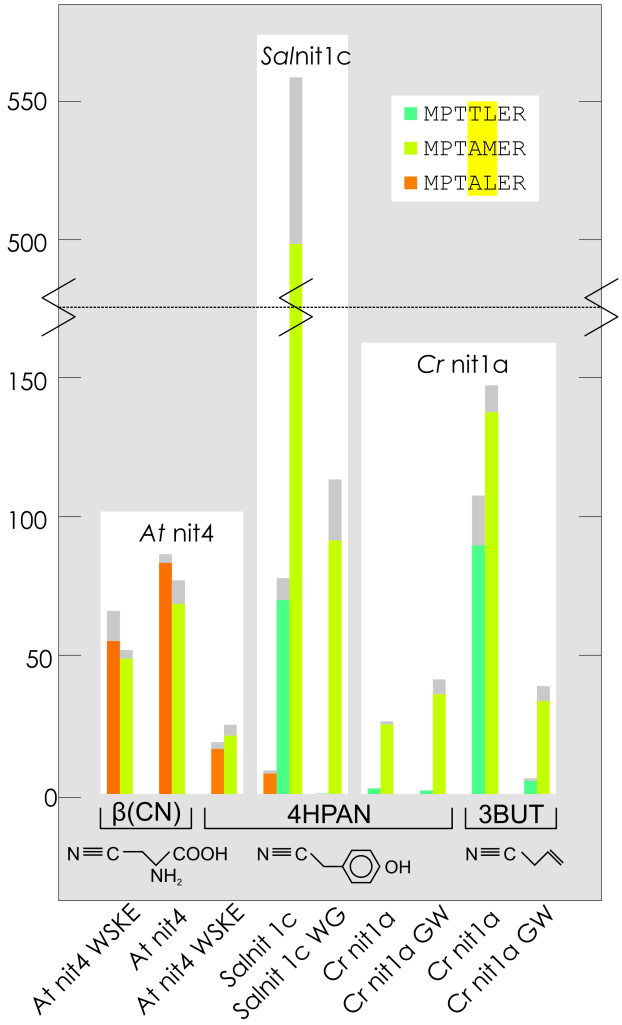


Figure 36. The effect of mutating α4 on the specific activity of plant nitrilases. The three motifs under investigation are shown in the top right-hand corner. Their effect on the specific activity of different mutants (listed at the bottom) is shown, firstly against the substrate β-cyanoalanine (β(CN)); then 4-hydroxyphenolacetonitrile (4HPAN) and then 3-butenitrile (3BUT). β(CN) activity is slightly reduced (first two columns) when "MPTALER" is exchanged with "MPTAMER". However, the specific activities against both 4HPAN and 3BUT are increased when the "MPTAMER" motif is present compared to "MPTTLER" this was found in mutant and WT *Sal*NIT1c, *Cr*NIT1a and *At*NIT4. This effect was also seen with the *Sal*NIT1c M154L single mutant a decrease in specific activity was seen when methionine was exchanged from this position.

In every case where there is a significant ($p<0.01$ from zero) specific activity for 3BUT/ 4HPAN in either wild-type or mutant enzymes, the measured activity is greater for either enzyme (3BUT or 4HPAN) when the motif: "MPTAMER" is present on α4 (Fig. 36). This effect acts both to decrease the activity when the wild-type enzyme has this motif in this position and increase the mutant activity when the wild-type motif is "MPTTLER". This increase in activity does not appear to have a differential effect for 4HPAN compared to 3BUT and so it is probably not relevant with regards to substrate specificity (Fig. 36). The second region (α6) (Fig. 37) has a

different effect on 4HPAN- compared to 3BUT activity. The "WSKE" motif favours an increase in 4HPAN activity, while "GSKE" results in an increase in 3BUT specific activity (Fig. 37).

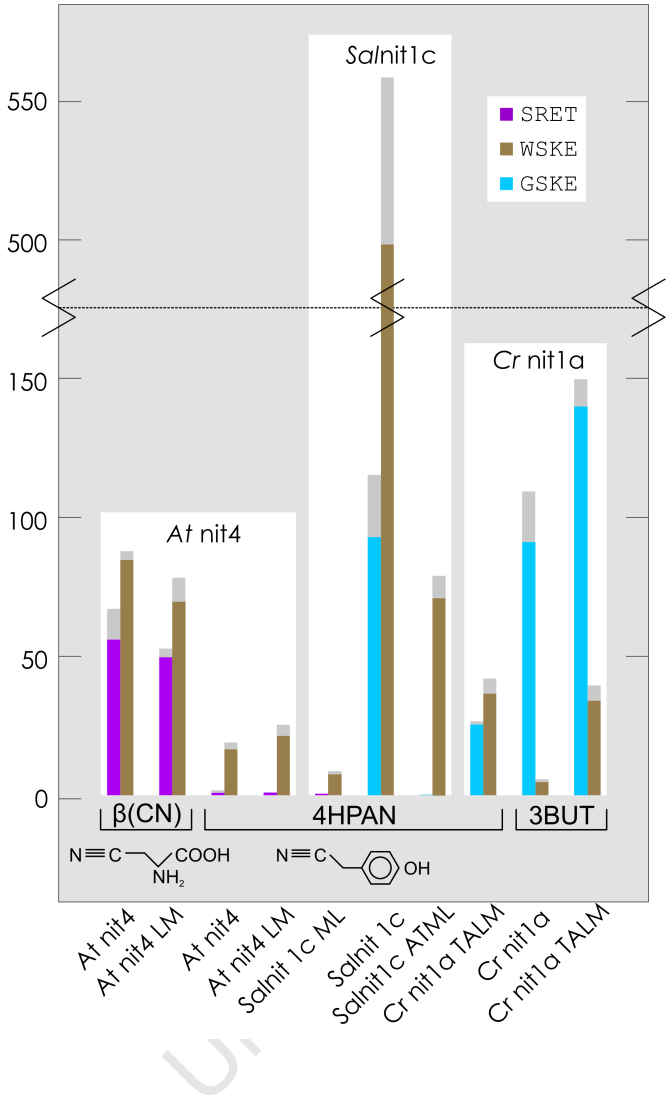


Figure 37. The effect of mutating the α6 on the specific activity of plant nitrilases. Interestingly, β(CN) activity is slightly increased when the WT motif "SRET" is exchanged with "WSKE". In every case where the WSKE motif is present, 4HPAN activity is strongly increased. Likewise, the activity of *Cr*NIT1a for 3BUT activity is decreased when the GSKE motif is exchanged for WSKE.

1.2.9 Nitrilase 4 β-cyanoalanine binding

*At*NIT4 active site pocket residue exchanges do not greatly affect the specific activity of the enzyme for β-cyanoalanine (β(CN)). This suggests that β(CN) binds to a region that is unaffected by the exchanges performed (Fig. 34). The binding of β(CN) was modelled on the crystal structure of 1UF5 (Hashimoto *et al.*, unpublished), where the carboxyl group of the substrate hydrogen binds to a *At*NIT4 R201 homologue (Fig. 38). The substrate visualised in 1UF5 (N-carbamyl-D-methionine) is similar in many respects to β(CN) insofar as it illustrates how the Arginine binds to the carboxyl-group of the substrate (Fig. 39).

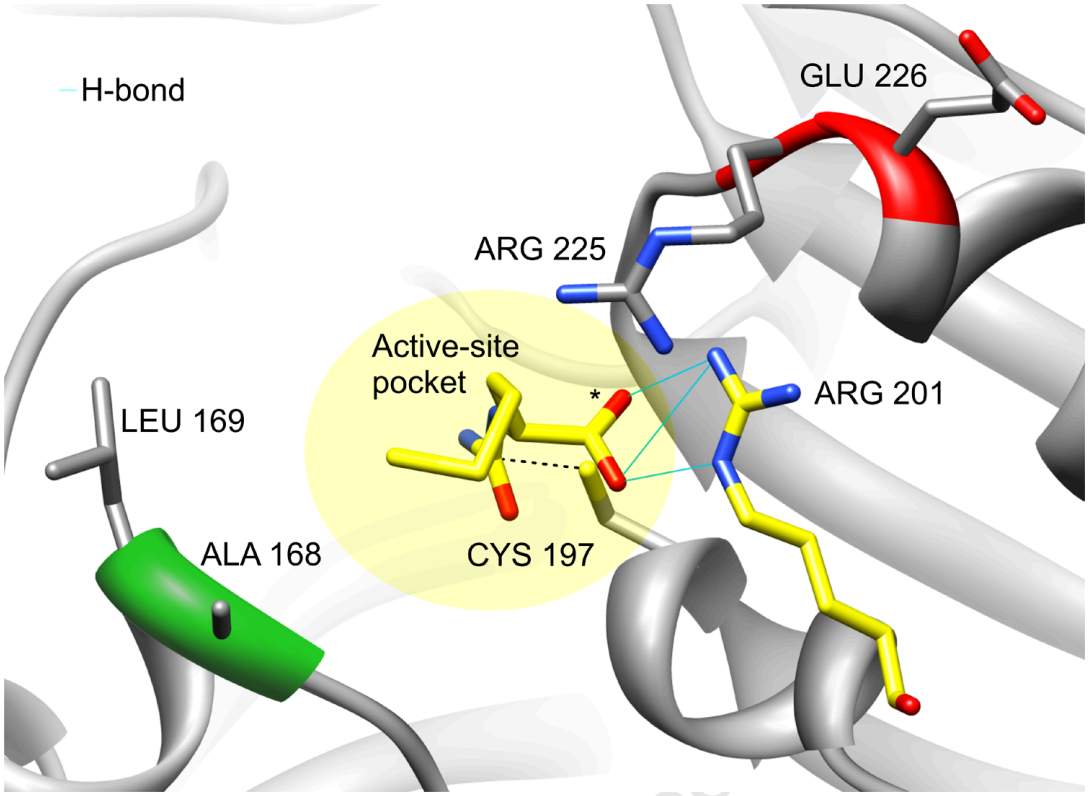


Figure 38. Homology model of *At*NIT4 superimposed on the crystal structure of an active-site mutant of N-carbamyl-D-amino acid amidohydrolase with bound N-carbamyl-D-methionine (1UF5: Hashimoto *et al*., unpublished). The active-site pocket is highlighted in yellow. A dashed line indicates the position of bond responsible for the formation of the tetrahedral covalent reaction intermediate. The carboxyl group of the substrate (*) forms hydrogen bonds with the homologue *At*NIT4 R201; this is the main interaction responsible for maintaining the substrate orientation in 1UF5. Two regions (shown in red and green) bounding the active-site pocket are conserved amongst distantly related nitrilase 4 enzymes (Fig. 33). But after mutating these regions β(CN) can still bind (Fig. 34).

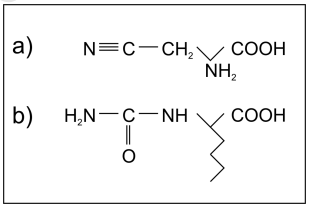


Figure 39. Structural comparison of the substrates of *At*NIT4 and that co-crystallized with the N-carbamyl-D-amino acid amidohydrolase (1UF5). (**a**) β(CN) is similar in some aspects to N-carbamyl-D-methionine (**b**) The major difference between the two substrates is the amino group of beta-cyanoalanine compared to the four carbon aliphatic chain of N-carbamyl-D-methionine. In 1UF5, this aliphatic chain is directed out of the binding pocket and makes no interactions. This explains why the N-carbamyl-D-amino acid amidohydrolase is active against a wide variety of substrates with different groups in this position.

On the basis of this model, it was proposed that *At*NIT4 might display activity against additional substrates, as the carboxyl group is predicted to be the only interaction involved in substrate binding. The substrates 3-aminopropionitrile and 3-

cyanopropionic acid were tested, but only 3-cyanopropionic acid was hydrolysed (Fig. 40) confirming the importance of the carboxyl group in β(CN) binding.

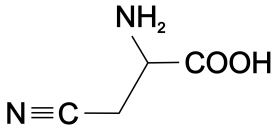
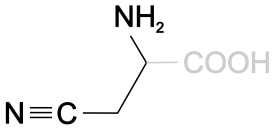
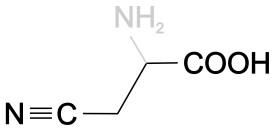
ß-cyanoalanine		Active
3-aminopropionitrile		Inactive
3-cyanopropionic acid		Active

Figure 40. The role of the β(CN) carboxyl group in *At*NIT4 binding. The natural substrate of *At*NIT4 (β(CN)) has an amino –, as well as a carboxyl functional group at the third position and shows good activity with *At*NIT4. According to the model (Fig. 38) only the carboxyl group is involved in substrate binding. The substrates 3-aminopropionitrile and 3-cyanopropionic acid were tested for nitrilase activity with *At*NIT4. As expected, only 3-cyanopropionic acid was hydrolysed.

1.3 Discussion

1.3.1 Substrate specificity

Genome library screening is a powerful method for identifying enzymes with the required substrate specificity for a particular application (Short, 1997). The Diversa Corporation (Diversa Corp. and Celunol Corp. merged to become Verenium in 2007) has collected and characterised over 200 nitrilases from environmental sampling and discovered several nitrilases with potential commercial applications (reviewed by DeSantis *et al*., 2002). These include nitrilases that are active against cyanohydrins to produce alpha-hydroxy acids or aryllactic acid derivatives. Several nitrilases capable of hydrolysing mandelonitrile to produce (R)- mandelic acid, one of which has been further characterised and found to hydrolyse mandelonitrile derivatives with a high efficiency. And an enzyme capable of small-scale enantioselective desymmetrisation of 3-hydroxyglutaryl nitrile to produce (*R*)-4-cyano-3-hydroxybutyric acid by selective hydrolysis (DeSantis *et al*., 2002).

The shortcoming of this approach is that these wild-type enzymes are often unsuitable for large-scale industrial production. For instance, industrial-scale trials at high substrate concentration produced low enantiomeric excess of (*R*)-4-cyano-3-hydroxybutyric, making the process impractical. Directed evolution by gene site saturation mutagenesis (Gray *et al*., 2001) led to the identification of a single amino acid variant capable of producing 95% enantiomeric excess at 100 mM (DeSantis *et al*., 2003). Every amino acid in the protein was exchanged using degenerate primer sets to produce a comprehensive library of 10,528 single-site mutants which were screened with three times oversampling (DeSantis *et al*., 2003). There are two obvious shortcomings to this approach, the first is the expense of producing and screening such a large number of clones, the second is the limited sampling space afforded by this technique: only single mutants are investigated.

This work is an attempt to rationalise the search for novel substrate specificities by using the nitrilase superfamily crystal structures to identify potential substrate binding regions, identifying sequences that correlate in these regions with specificity and testing the effect on specific activity by exchanging potential “substrate determining” residues by site directed mutagenesis. Furthermore, electron microscopy and image

processing have demonstrated the unexpected role that the quaternary-structure of plant nitrilases plays with regards to substrate specificity.

1.3.2 Methodology

1.3.2.1 Helical reconstruction

Real-space helical reconstruction was used to generate volumes using the IHRSR algorithm (Egelman, 2000). The advantage of this technique is that samples comprising very short or even curved fibres can be used without the need for straightening (Egelman, 2000). All of the samples used here were visualised under negative-stain, despite the limits imposed by this technique on attainable resolution to >2 nm and the possibility of sample distortion ("flattening"). Despite these shortcomings, negative stain was used firstly for the sake of simplicity as a large number of fibres had to be reconstructed and secondly, in our experience, nitrilase samples reconstructed by negative stain have compared favourably with those from cryo-electron microscopy. Thirdly, the information required from the reconstructions, namely the helical twist can be acquired with a high degree of precision from the volumes. Placement accuracy for a 2 nm resolution reconstructed volume is generally ~0.2 nm (Volkmann and Hanein, 1999).

1.3.2.2 Specific activity measurement

The preference of an enzyme for a given substrate was estimated here by measuring specific activity. This value is influenced by both the substrate-enzyme affinity as well as the maximum turnover rate of the enzyme. Enzyme kinetics can reveal the influence of each of these two factors: the theoretical reaction rate as substrate concentration approaches infinity (V_{max}) shows the reaction rate attained when substrate is bound to an enzyme. The Michaelis constant K_M indicates the affinity of the substrate for the enzyme; a small K_M indicates that the rate will approach V_{max} more quickly (Menten and Michaelis, 1913). The implicit assumption when estimating substrate specificity differences using specific activity is that V_{max} remains constant. This implies that once the substrate is bound to the active site, it will be hydrolysed at the same rate as the test condition. It is known however, that the chemical properties of certain nitriles increase their hydrolysis rate (e.g. Stevenson *et al*., 1992). Certain mutations have also been found to change the overall rate of

reaction, by for instance, promoting fibre stability (Thuku *et al*., 2007). The data presented here therefore need to be interpreted carefully and each case examined individually.

1.3.3 Active site pocket specificity determinants

In the case of active site pocket mutations, the "preferred substrate" is the substrate that shows the highest specific activity under the test conditions. Where an attempt is made to exchange the substrate specificity of two enzymes, the goal is to decrease the specific activity for the native substrate and increase it for the preferred substrate of the other enzyme. The primary component of this change may be a change in the binding affinity, but remains an assumption in the absence of kinetic analysis. Small substrates have a tendency to have higher specificity. This can be explained in terms of the relative difference in size of a substrate with every additional atom. For instance, the R-group of propionitrile is 100% bigger than acetonitrile, while decanitrile is only 12.5% bigger than nonanitrile. This difference is clearly illustrated by the difference in substrate specificity between *B. pumillus* CynD and *At*NIT3 (Fig. 10). In the case of *Cr*NIT1a and *Sal*NIT1c, exchanging the residues in the active site pocket led to the desired effect: the wild type activity was lowered and activity against the preferred substrate of *Sal*NIT1c (4-hydroxyphenolacetonitrile) increased.

This effect was further explored by defining the determinants of substrate specificity in *Sal*NIT1c by exploring the characteristics of converted substrates (Trompetter, 2010). The importance of the hydroxyl-group is illustrated by the low activity of 2-phenylacetonitrile and after exchanging the hydoxyl- with a methyl-group (4-methylphenolacetonitrile) (Fig. 11). A hydrogen bond donor in the substrate-binding pocket is most likely responsible for this role, because high nitrilase activity was observed with a hydrogen bond acceptor (4-methoxyphenylacetonitrile) (Fig. 11). The length of the substrate is also of critical importance, one carbon bond longer (3-(4-hydroxyphenyl)propionitrile) and one carbon bond shorter (4-hydroxybenzonitrile) shows very low activity (Fig. 11). The importance of the aromatic ring is illustrated by the lack of activity with 4-hydroxypentanitrile (data not shown). *Sar*NIT1c on the other hand is active against a broader range of aromatic substrates of the correct length (Fig. 11). While *Cr*NIT1a catalyses the conversion of short aliphatic substrates (Fig. 10).

Whether or not a substrate is converted depends on factors that either promote or inhibit substrate binding. These factors are not immediately apparent in this case: tryptophan is associated with 4-hydroxyphenolacetonitrile (4HPAN) activity (Fig. 37); exchanging this residue for a glycine decreases 4HPAN activity and increases activity against 3-butenitrile (Fig. 37). Exchanging tryptophan for a serine on the other hand increases specific activity against the preferred substrate of *Sar*NIT1c (2-phenylacetonitrile) (Trompetter, 2010). Explaining the complicated substrate-enzyme interactions that result from these interactions requires further structural data to explain.

While *Cr*NIT1a/ *Sal*NIT1c substrate binding pocket exchange mutations (Fig. 31, 32), resulted in a partial exchange of substrate specificity, a complete exchange was not observed. This may be due to the difference in helical twist between these two enzymes (a difference of ~0.7°), which would presumably have the effect of changing the conformation of the active site pocket (Fig. 28). It is interesting to note that site-directed mutagenesis of the substrate binding pocket residues of *At*NIT4 to match those found in *Sal*NIT1c (difference of ~3.0°) leads to an even less successful slight increase in 4-hydroxyphenolacetonitrile activity (Fig. 34). While, exchanging the substrate binding pocket residues of *Sal*NIT1c to match those of *At*NIT4, leads to a complete loss of activity (Fig. 35). This effect may also be as a result of exchanges occurring at positions distant from the active site pocket. To test whether or not the "twist-specifying residue" identified in *Cr*NIT1a and *Cr*NIT1b (residue 80) accounts for these differences, this amino acid residue was also exchanged in the *Cr*NIT1a/ *Sal*NIT1c active site exchange mutants. In *Cr*NIT1a this active site and "twist specifying residue" double mutant preferred 4-hydroxyphenolacetonitrile (the preferred substrate of *Sal*NIT1c) to 3-butenitrile (the preferred substrate of *Cr*NIT1a). The equivalent mutation in *Sal*NIT1c produced nearly equal activity against 3-butenitrile and 4-hydroxyphenolacetonitrile.

1.3.4 Interface disruption and inactivation

So far every enzyme with significant nitrilase activity that has been investigated by electron microscopy has formed helices or "helix-like" assemblies. These include "terminating spirals" (Sewell *et al*., 2003), rings and figure-of-eights (Williamson *et al*., 2009), and C-shaped assemblies (Thuku *et al*., 2007). In every case, regardless of

the precise nature of the association, dimers associate across the C-interface (Sewell *et al*., 2003) (Fig. 3). One exception to this is an enzyme with ~2.4 nKat of nitrilase activity that forms a dimer in solution from the hyperthermophilic archaeon *Pyrococcus abyssi* (Mueller *et al*., 2006). The structure of this enzyme has been determined at high resolution by X-ray crystallography (Raczynska *et al*., 2011). The position of α4 in this enzyme, in the absence of the C-interface, appears to be maintained by interactions with β3 and α5 (see Fig. 4). In any case, the low nitrilase activity measured for this enzyme may indicate that nitrile hydrolysis is not the primary reaction catalysed by this enzyme.

Disrupting the interfaces of helix-forming nitrilases by changing pH or by site-directed mutagenesis of helical interfaces leads to their inactivation (Sewell *et al*., 2005). However, it was not known before whether or not these mutations lead to nitrilase dimer formation in solution: this has now been conclusively demonstrated by protein purification, gel-filtration and electron microscopy (Fig. 13). It has been hypothesised (Kimani *et al*., 2007) that the role of C-interface association is to maintain the stability of one of the active-site residues (E_2). This work has illuminated another possible role for fibre formation on the catalytic activity of nitrilases. The size of the R-group of substrates hydrolysed by various nitrilases is correlated with the angle across the C-interface between dimers in the helix (Fig. 17). Modifying its angle by as little as 1.0° using site directed mutagenesis completely changes the preferred substrate, disrupting this interface entirely would therefore be expected completely disrupt this fine-tuned mechanism.

SECTION 2: Amide: acid ratio

2.1 Introduction

2.1.1 Mechanism

The nitrilase reaction mechanism is thought to initiate with nucleophilic attack of the slightly positive C-atom of the nitrile by a cysteine to form a covalent enzyme-substrate complex (Mahadevan and Thimann, 1964). This initial reaction-mechanism step was proposed on the basis of the observation that electron-withdrawing R-goups increased the rate of hydrolysis (Mahadevan and Thimann, 1964) presumably leading to an increase in the charge on the C-atom and therefore a higher rate of covalent substrate-enzyme complex formation. Further support for this hypothesis comes from the activation of nitrilases by low concentrations of reducing agents (Mueller *et al*., 2006; Kim *et al*., 2009) and inhibition by high concentrations of thiol reagents (Piotrowski *et al*., 2001). Site-directed mutagenesis of the active-site cysteine leads to a complete loss of nitrilase activity (Kobayashi *et al*., 1992; Piotrowski *et al*., 2001) but interestingly, 5%-10% of the nitrile-hydratase activity remains (Piotrowski *et al*., 2001).

Acid quenching of the enzyme-substrate mixture and subsequent mass-spectroscopy demonstrated that both a covalent thioimidate – or acyl-intermediate is formed during nitrile hydrolysis (Stevenson *et al*., 1992). The reaction requires two hydrolysis reactions and protonation of the N-atom of the nitrile (which ultimately leaves as ammonia). On this basis, Stevenson *et al*. (1992) proposed that hydrolysis of the thioimidate results in a tetrahedral intermediate, protonation of the primary amine results in the loss of ammonia and formation of an acylenzyme which is further hydrolysed to form a second tetrahedral intermediate, reduction of the S-atom of cysteine, results in the formation of carboxylic acid. Jandahyala *et al*. (2005) and Fernandes *et al*. (2006) modified this generally accepted reaction mechanism to explain the formation of amide product (Fig. 41).

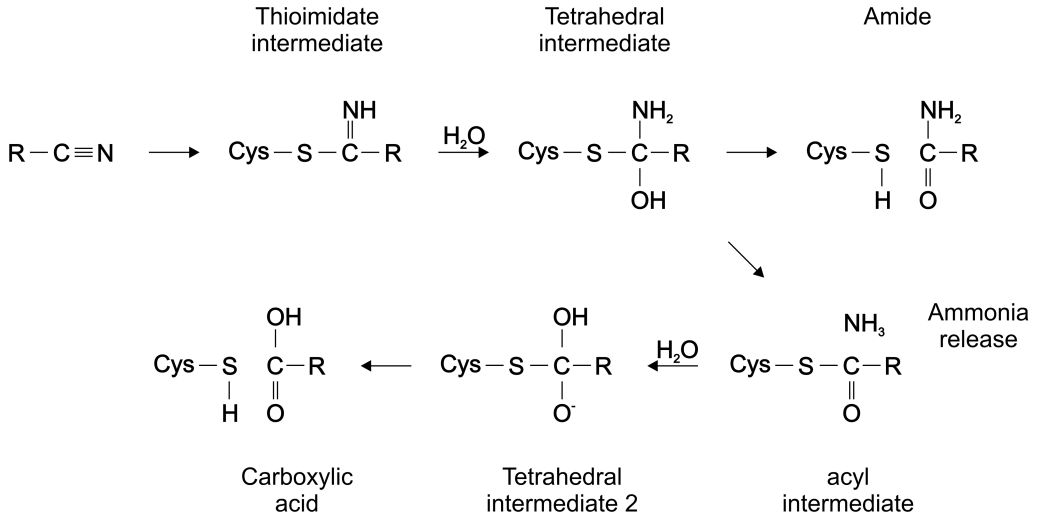


Figure 41. Reaction mechanism. Nucleophilic attack of the nitrile results in the formation of a covalent thioimidate intermediate that is hydrolysed to form a tetrahedral intermediate (Stevenson *et al*., 1992). At this point the reaction can proceed in one of two ways: breakage of the S-C bond leading to the release of amide (nitrile hydratase reaction), or breakage of the N-C bond leading to the release of ammonia and formation of an acyl intermediate (Jandhyala *et al*., 2005; Fernandes *et al*., 2006; Williamson *et al*., 2010). The acylenzyme complex is hydrolysed a second time, resulting in the formation of a second tetrahedral intermediate and the formation of acid (nitrilase reaction) (Stevenson *et al*., 1992).

According to the above scheme, whether a nitrilase – or nitrile hydratase reaction takes place depends on whether the S-C – or N-C bond is broken (Fig. 41). C-N bond dissociation energy (ΔH°) is 300 kJ/mol, while that of C-S is 270 kJ/mol (Sanderson, 1976) and therefore, in the absence of surrounding electronic effects, we would expect cysteine to be the better leaving group (Williamson *et al*., 2010). Fernandes *et al*. (2006) suggest that protonation of the primary amine (tetrahedral intermediate – Fig 41) to form ammonium is necessary for ammonia release and therefore factors that influence amino group protonation determine the amide:acid ratio.

2.1.2 Electronic effects

A correlation between the electron-withdrawing ability of substrates and the amide: acid ratio has been demonstrated in *At*NIT1 (Osswald *et al*., 2002), the percentage of amide product increases with increasing electronegativity of the α-substituent: hydrolysis of (E)-2-butenenitrile produces 1% amide, whereas the electron-deficient 3-nitroacrylonitrile produces 95% amide product. Fernandes *et al*. (2006) tested the recombinant nitrilase from *P. fluorescens* EBC 191 with similar substrates of opposing electronic character and found that 2-phenylpropionitrile formed >99% acid

while the electron-deficient substrate 2-chloro-2-phenylacetonitrile resulted in 89% amide formation. This effect depends on the position of the electron-deficient group relative to the nitrile, for instance: 84% amide was produced with the substrate 2-cyanopyridine, and only 8.9% with 3-cyanopyridine with the nitrilase from *A. niger* K10 (Kaplan *et al*., 2006). See Fig 42.

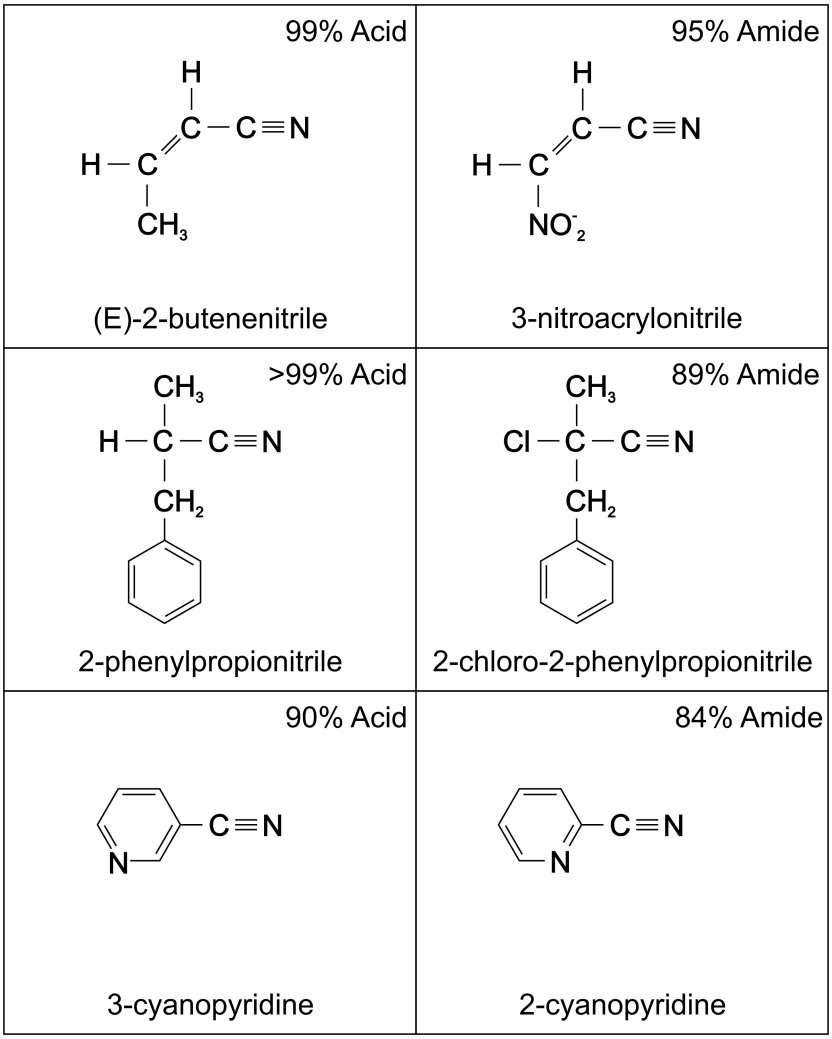


Figure 42. The nitrilase-catalysed hydrolysis of nitriles with electron withdrawing α-substituents results in the formation of a higher proportion of amide product. *At*NIT1 was tested with (E)-2-butenenitrile and 3-nitroacrylonitrile (Osswald *et al*., 2002); *P. fluorescens* EBC 191 was tested with 2-phenylpropionitrile and 2-chloro-2-phenylacetonitrile (Fernandes *et al*., 2006); *A. Niger* was tested with 2-cyanopyridine and 3-cyanopyridine (Kaplan *et al*., 2006).

2.1.3 Amide:acid ratio of nitrilase enzymes

The amide: acid ratio is also dependent on characteristics of the enzyme itself because different nitrilases display different amide:acid ratios with the same substrate e.g.

*At*NIT1, *At*NIT2 and *At*NIT3 with indole-3-acetonitrile (Pollmann *et al*., 2002) and NIT4 enzymes from various plant species with β-cyanoalanine (β(CN)) (Piotrowski *et al*., 2001). Recently it has been demonstrated that the amide: acid ratio as well as enantioselectivity are dependent on the residues directly adjacent to the active-site cysteine (Kiziak and Stoltz, 2009; Sosedov *et al*., 2010) implicating steric factors, as the percentage of amide also depends on whether the (S) – or (R)-enantiomer is tested (Fernandes *et al*., 2006). Interestingly, the proportion of amide also increases with deletion of the C-terminus by ~59 amino acids (Kiziak *et al*., 2007). In most cases, the proportion of nitrilase-catalysed nitrile hydratase activity is unknown, owing to the widespread use of ammonia detection to measure nitrilase catalytic activity (Fernandes *et al*., 2006).

2.1.4 Aims

In this section I aim to:

- Analyse nitrilase homology models to identify amino acid residues with possible catalytic functions
- Conduct a detailed analysis of the crystal structures of the available nitrilase superfamily enzymes to model nitrilase reaction intermediate states
- Formulate a detailed nitrilase mechanism that includes factors that potentially influence the acid:amide reaction product ratio
- Analyse the amino acid sequences of plant nitrilases with differing amide:acid ratios and identify exchanges that potentially lead to these differences
- Test the influence of the amino acid residue exchanges identified by measuring the product ratio of mutant enzymes constructed by site directed mutagenesis

2.1.5 Brief summary of findings

The four active site amino acid residues previously identified (Brenner, 2002; Kimani *et al*., 2007) were identified here in an automated fashion using THEMATICS (Ko *et al*., 2005) and amino acid sequence conservation data. Interestingly, a conserved tyrosine residue lying at the border of the catalytic site was identified as a fifth catalytic amino acid residue. This tyrosine is visualised in two crystal structures of nitrilase superfamily enzymes of unknown function from *Pyrococcus horikoshii*

(Sakai *et al*., 2004: 1J31) and yeast (Kumaran *et al*., 2003), and an enzyme that shows measurable nitrilase activity from *Pyrococcus abyssi* (Raczynska *et al*., 2011). An analysis of nitrilase superfamily crystal structures and homology modelling, led to the formulation of a detailed reaction mechanism that could account either for the formation of carboxylic acid or an amide product. It has been frequently proposed that the hydrolysis of the thioimidate intermediate proceeds via a tetrahedral intermediate that has an amino group (Stevenson *et al*., 1992). If the cysteine leaves then an amide is produced, whereas if the amino group leaves as ammonia, a second hydrolysis of the resulting thioester produces an acid product (Fernandes *et al*., 2006).

The tyrosine identified is predicted to contribute to the formation of a hydrogen-bonding network that promotes protonation of the amino group either indirectly by increasing the nucleophilicity of glutamic acid, or directly through hydrogen bonding to one of the protons on the amino group. The predicted role of this tyrosine is stabilisation of the incipient ammonia. Homology models of two closely related nitrilase 1 enzymes from plants with substantially different amide:acid ratios were compared. In one of the nitrilases, which produced a high proportion amide product, the tyrosine was shifted in the sequence as a result of an insertion. Repositioning this tyrosine by deleting the additional amino acid led to a change in the proportion of acid product from 17% to 97% of the total. Mutating this tyrosine to phenylalanine in a nitrilase 4 resulted in a shift in the percentage of acid product from 44% to 29%.

In section 1, β-cyanoalanine binding is proposed to occur via hydrogen bonding of the carboxyl group of the substrate to a conserved arginine in the binding pocket. Analysis of the reaction product ratios of several nitrilase 4 enzymes revealed that lysine also occurs in this position, when this is the case, the amide:acid ratio is increased. Homology modeling suggested a possible explanation for this effect: the difference in length of arginine compared to lysine results in a rotation of the substrate about the C-S bond. The amino group of the tetrahedral intermediate is therefore shifted away from the optimal location for amino group protonation relative to the active site residues, resulting in an increase in the proportion of amide formed.

2.2 Results

2.2.1 Nitrilase active site

A THEMATICS (Ko *et al*., 2005) prediction was undertaken to identify residues forming an "active site cluster" in a homology model of *At*NIT4. Residues that are not conserved between all known nitrilases were excluded. The theoretical pKa's (Gordon *et al*., 2005; Myers *et al*., 2006) of all identified "active site cluster" residues were calculated in order to determine their most likely protonation state (Table 1).

E76*	pKa: -0.1	E_1
Y82	pKa: 29.1	
K163*	pKa: 32.7	K
E170*	pKa: -10.2	E_2
C197*	pKa: 15.8	C

Table 1. Active-site cluster predicted by THEMATICS with a Z-score >0.99 and distance <9.0 Å (Ko *et al*., 2005) the predicted PKa for each residue is shown (Gordon *et al*., 2005; Myers *et al*., 2006). Previously described nitrilase active site residues are shown (*). According to the predicted pKa's, E_1 and E_2 will remain deprotonated (minus one state), tyrosine 82 and the active site cysteine will remain neutral and the active site lysine will be in the +1 state.

This automated procedure located the three generally accepted nitrilase superfamily active site residues, as well as the fourth, identified by Kimani *et al*. (2007). Interestingly, one additional residue was also identified: Y82 (*At*NIT4), all known nitrilase enzymes, including those from plants, bacteria and fungi have a tyrosine in this position (e.g. Thuku *et al*., 2009).

2.2.2 The unbound enzyme state

Several members of the nitrilase superfamily have been visualised in the unbound state: 1j31, 1erz, 1fo6, 2dyu and 1f89 (Sakai *et al*., 2004, Nakai *et al*., 2000, Wang *et al*., 2001, Hung *et al*., 2008 and Kumaran *et al*., 2003). These enzymes share a common architecture, with four active site residues that superimpose precisely onto one another (Fig. 43a). In the position of *At*NIT4 Y82, other members of the nitrilase superfamily usually contain a tyrosine or phenylalanine, but other amino acids are also found in this location. Of particular relevance in these unbound structures is the presence of two regions (labelled a and b in Fig. 43a). The density in these regions when unoccupied by a ligand is generally interpreted to indicate that water molecules occupy these positions (Fig. 43a). Interestingly, region "b" corresponds to the position

of the carbamyl oxygen of the bound N-carbamyl-D-amino acid in 1uf7 (Fig 43b), leading to the suggestion that this water molecule is responsible for the first hydrolysis reaction of the nitrile (Fig 43:3). The second hydrolysis reaction leading to the loss of ammonia, would therefore involve the water molecule occurring in position "a".

However, tempting this view is, it is unlikely for several reasons. Firstly, it is far from certain that the observed density is correctly interpreted. An analysis of the electron densities of the nitrilase superfamily enzymes for which structure-factors are available reveals that a more likely explanation for the observed density is oxidation of the active-site cysteine. Secondly, the local electrostatic environment in position "a" is influenced by the nearby negatively charged glutamic acid (E_1 – Fig. 43a) and surrounded by multiple hydrogen-bond acceptors. This favours amide binding, but nitriles are strongly polarised toward the nitrogen, with high molecular dipole moments and are more likely to be repelled. Thirdly, while the nitrogen of the tetrahedral intermediate most likely occupies a position close to position "a" when the carbon of the nitrile is sp3 hybridised, this is unlikely to be the case with the linear sp-hybridised nitrile, which on entering the substrate pocket, most likely occupies position "b".

2.2.3 Tetrahedral intermediate state

The amidase from *Pseudomonas aeruginosa* (2uxy) (Andrade *et al*., 2007) was crystallized with a trapped acyl transfer intermediate with distorted tetrahedral sp3 hybridization bound to the active-site cysteine (see Fig. 43c), which arises due to a nucleophilic attack by a hydroxylamine rather than a water. This tetrahedral intermediate is positioned, such that a single-bonded oxygen lies in close proximity to the active-site lysine (2.64 Å) and an aspartic acid (2.95 Å) and the authors predict that a proton is shared between this oxygen and these residues. This configuration is plausible as the active-site lysine is almost certainly protonated (predicted pKa of ~33 Table 1). The nitrogen of the tetrahedral intermediate in 2UXY is bound to a hydroxyl. This differs from the tetrahedral intermediate thought to occur in nitrile hydrolysis.

2.2.4 The nitrilase reaction model

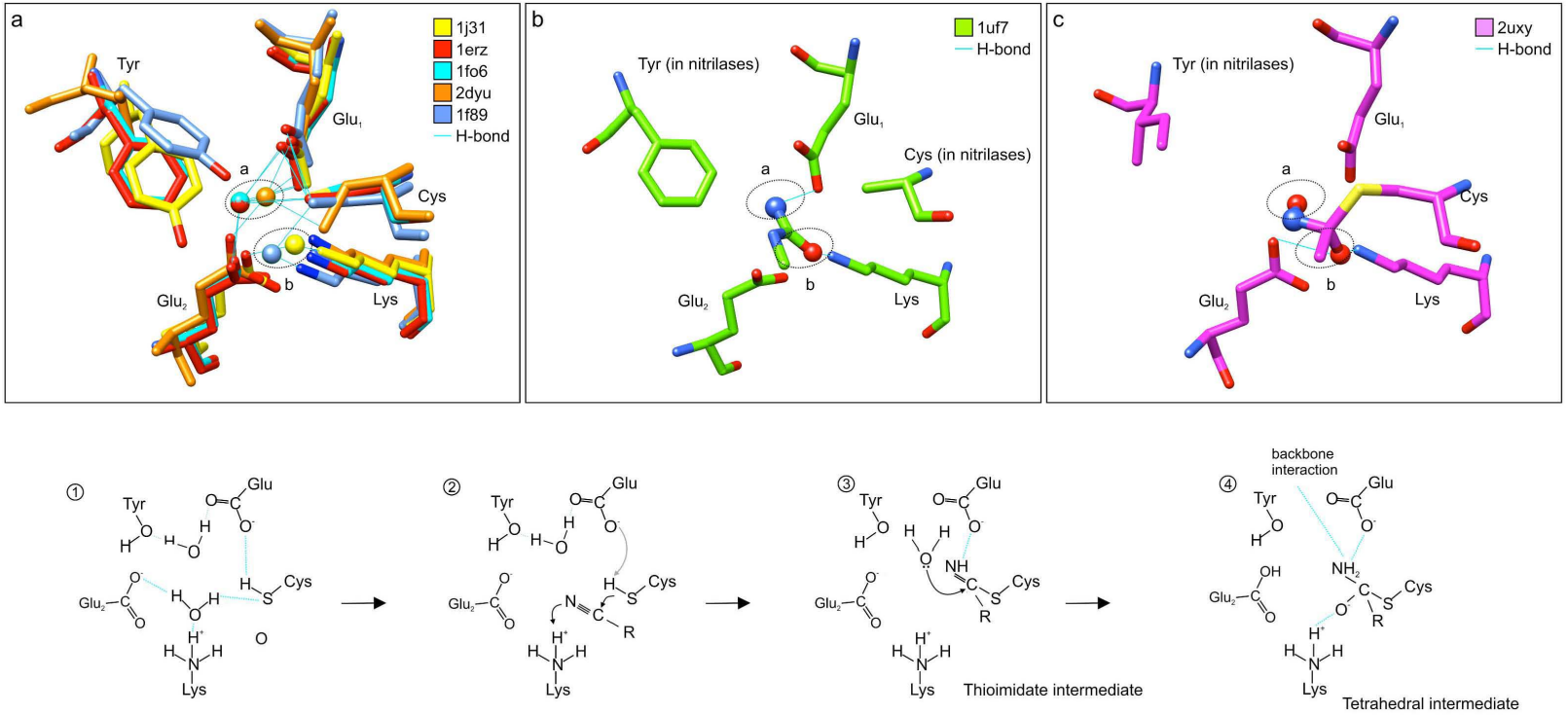

a
1j31
1erz
1fo6
2dyu
1f89
H-bond
Tyr
Glu1
a
Cys
b
Glu2
Lys
b
1uf7
H-bond
Tyr (in nitrilases)
Glu1
Cys (in nitrilases)
a
b
Glu2
Lys
c
2uxy
H-bond
Tyr (in nitrilases)
Glu1
a
Cys
b
Glu2
Lys
①
②
③
④
backbone interaction
Thioimidate intermediate
Tetrahedral intermediate

Figure 42. Proposed nitrilase reaction mechanism: formation of the tetrahedral intermediate.

(**a**) In the unbound state, two well-coordinated waters have been proposed to bind in the active-site pocket. The first, in position "a", hydrogen bonds with Glu_1 and Glu_2 and the active-site cysteine, this interaction is visualised in the crystal structures 2dyu, 1fo6 and 1erz. The second water molecule (in position "b") is purported to hydrogen bond with the active-site lysine, cysteine and Glu_2, the crystal structures 1J31 and 1F89 are interpreted as having a water molecule in this position.

(**1**) One or both of these waters are obvious candidates for the waters responsible for nitrile hydrolysis.

(**b**) The crystal structure of a C171A/V236A mutant of N-carbamyl-D-amino acid amidohydrolase from *Agrobacterium sp*. complexed with N-carbamyl-D-valine (1uf7). The amide of the substrate occupies position "a" and hydrogen bonds to Glu_1; the hydroxyl group occupies position "b" and hydrogen bonds with the active-site lysine.

(**2**) Nucleophilic attack of the carbon atom of the nitrile leading to the formation of a thioimidate intermediate (Hook and Robinson, 1964). The nucleophilicity of the cysteine sulphur atom is increased by close proximity of Glu_1 (e.g. Fernandes *et al*., 2006). The R-group of the substrate is orientated according to 1uf5; the nitrogen of the nitrile occupies position "b" also known as the "oxy-anion hole". The active-site lysine acts as a general acid catalyst, drawing away the electron cloud of the nitrogen of the nitrile and increasing the positive charge on the carbon. The proton of the cysteine is transferred to the nitrogen.

(**c**) The amidase from *Pseudomonas aeruginosa* was fortuitously crystallised with a trapped acyl transfer reaction intermediate state. The hydroxyl group of this tetrahedral intermediate state hydrogen bonds with Glu_2 and shares a proton with the active-site Lys (Andrade *et al*., 2007). A hydroxyl group occupies the position of the "a" water molecule, the nitrogen lies just outside of this pocket.

(**3**) The thioimidate intermediate. The imine group lies in close proximity to the water molecule (in the position of the hydroxyl in C) in position "a". The first hydrolysis reaction therefore involves this water molecule. Glu_2 is correctly placed (Kimani *et al*., 2007) to enhance the nucleophilicity of this water molecule. The lone pair of electrons on the water molecule attack the sp^2-hybridised carbon, leading to the formation of the tetrahedral intermediate.

(**4**) The tetrahedral intermediate. The hydroxyl group lies in position "b" (the oxy-anion hole) and is hydrogen bonded to Glu_2 and Lys. Position "a" is occupied by the amino group, hydrogen bonds are predicted to occur between Glu_1, a backbone interaction and with slight repositioning, a Tyr (shown).

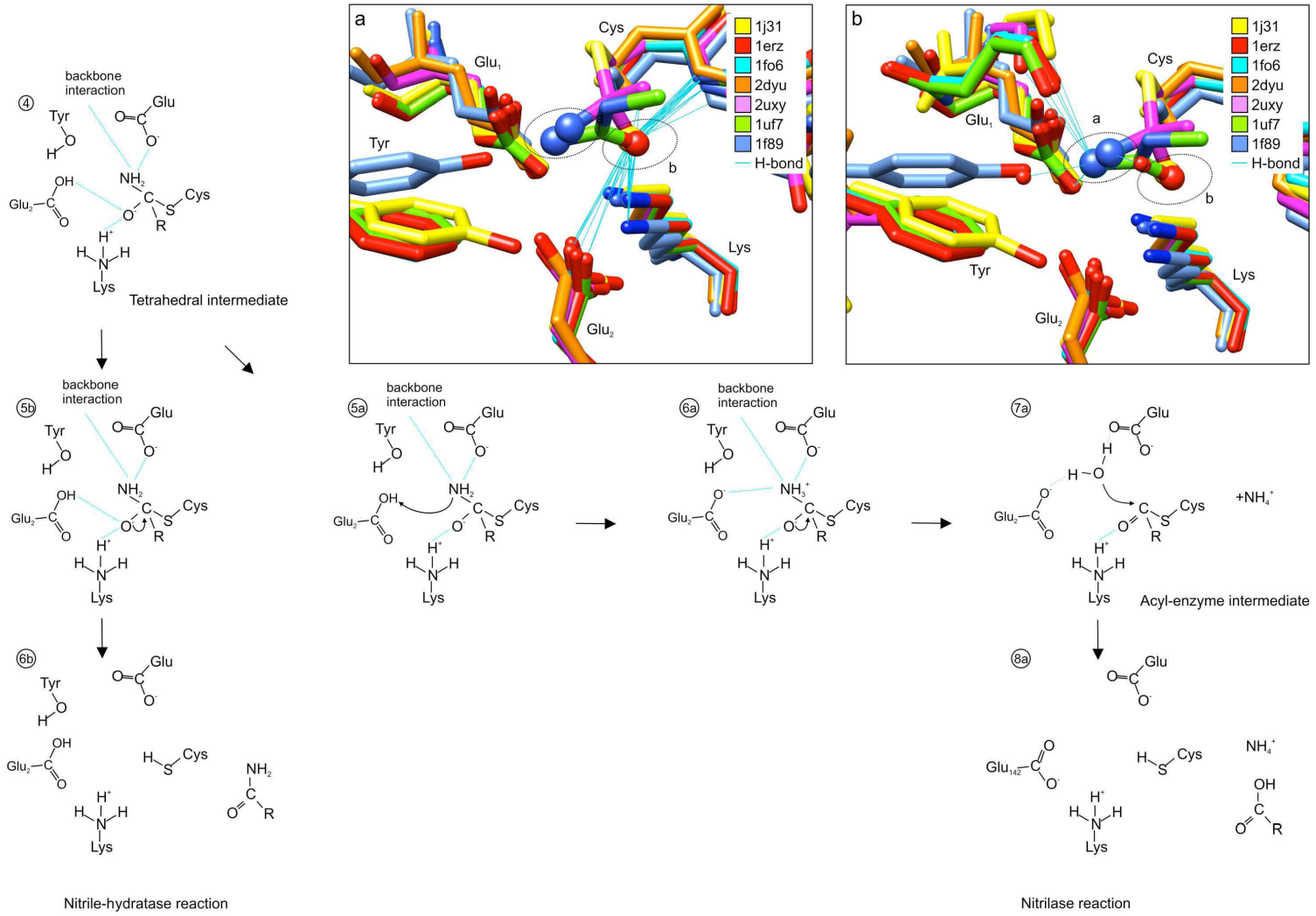

backbone interaction
Tetrahedral intermediate
Nitrile-hydratase reaction
a
b
1j31
1erz
1fo6
2dyu
2uxy
1uf7
1f89
H-bond
Cys
Glu1
Tyr
Lys
Glu2
backbone interaction
Acyl-enzyme intermediate
+NH4+
Nitrilase reaction

Figure 43. Proposed nitrilase reaction mechanism: amide or acid formation determinants

(**a**) Stability of the tetrahedral-intermediate hydroxyl group. Superimposing the nitrilase superfamily crystal structures reveals a conserved substrate intermediate hydroxyl group, which lies in position "b" and hydrogen bonds to Glu_2, Lys and a backbone interaction.

(**b**) Stability of the nitrogen atom of the tetrahedral intermediate. The nitrogen occurs in position "a" and makes hydrogen bonds with Glu_1, a backbone interaction and possibly the conserved tyrosine (shown).

Nitrilase reaction

(**5a**) The amino group is stabilised by a series of hydrogen bonds, which increase the nucleophilicity of the nitrogen. Glu_2 is correctly placed to donate a proton (and according to this scheme, it is protonated at this stage).

(**6a**) Protonation of the amino group leads to the formation of ammonium, the stability of this intermediate is enhanced by the surrounding hydrogen bond network. The oxygen becomes sp2 hybridised, resulting in the formation of the acyl-enzyme intermediate and loss of ammonia.

(**7a**) The final hydrolysis reaction. The water molecule observed in 2DYU (Hung *et al*., 2007) now has sufficient space to occupy the position vacated by ammonia. Glu_2 is correctly placed to act as a general base catalyst, enhancing the nucleophilicity of the water molecule, which hydrolyses the acyl intermediate.

(**8**) The final products of the nitrilase reaction, carboxylic acid and ammonia.

Nitrile-hydratase reaction

(**5b**) In the absence of protonation of the primary amine, sp2 hybridisation of the oxygen leads to breaking the C-S bond (Fernandes *et al*., 2006).

(**6b**) The nitrile-hydratase reaction leads to the formation of an amide product.

2.2.5 Determinants of nitrilase amide:acid product ratio

According to the mechanism proposed in 2.2.4, whether an amide – or acid product is formed depends on whether the C-S – or the C-N bond is broken after the formation of the tetrahedral intermediate. The energy required to break the C-S bond in general is lower than that of the C-N bond and therefore the difference between the two reactions depends on factors that favour ammonia formation. In order for ammonia to be formed the amino group of the tetrahedral intermediate requires further protonation (Jandhyala *et al*., 2005; Fernandes *et al*., 2006). This may be influenced by the stability of the hydrogen-bond network surrounding the amino group of the tetrahedral intermediate. Any disruption will decrease the likelihood of ammonium formation, decreasing ammonia release and increasing the proportion of amide product. The stability of the hydrogen-bonding network is influenced by a diverse range of factors such as; the position of the amino group within the "a" region, the exact orientation of the catalytic residues and the presence or absence of certain residues surrounding the active site pocket.

2.2.6 Sinapis alba and *Sinapis arvensis: the effect of tyrosine*

The NIT1c from *Sinapis alba* (*Sal*NIT1c) and *Sinapis arvensis* (*Sar*NIT1c) show 88% identity at the amino acid level and display very similar substrate specificity (*Sal*NIT1c prefers 4-hydroxyphenolacetonitrile, while *Sar*NIT1c prefers a range of aromatic substrates (Fig. 11) such as, 2-phenolacetonitrile) (Trompetter, 2010). Interestingly, *Sar*NIT1c produces ~17% acid product, while *Sal*NIT1c produces ~91% acid (Niels Agerbirk, personal communication). Homology modelling was undertaken to identify any differences between the two enzymes at the amino acid level within 5 Å of the active site cysteine (Fig. 45). The first difference is in a substrate-specifying region (α6) of the binding pocket (see section 1): *Sal*NIT1c has a tryptophan, while *Sar*NIT1c has a serine in this position and this difference has been shown to account for the substrate preference difference between these enzymes (Trompetter, 2010).

The second difference is that *Sar*NIT1c contains a glycine insertion on a loop lying on the border of the active site between β2 and α2 (Thuku *et al*., 2009). When deciding how this would be modelled, the high homology (structural and sequence) of the region preceding the insertion was taken into consideration and the insertion was

modelled as resulting in the repositioning of the conserved tyrosine residue (Y82 in *At*NIT4) (Fig. 45). It was proposed that this residue is involved in maintaining the stability of the nitrogen, either indirectly by increasing the nucleophilicity of glutamate, or directly through hydrogen bonding to one of the protons on the nitrogen. The difference between *Sal*NIT1c and *Sar*NIT1c could therefore be explained in terms of stability of the nitrogen: the presence of this tyrosine results in increased stability of a protonated nitrogen which therefore allows the nitrogen to leave as ammonia (Fig. 44).

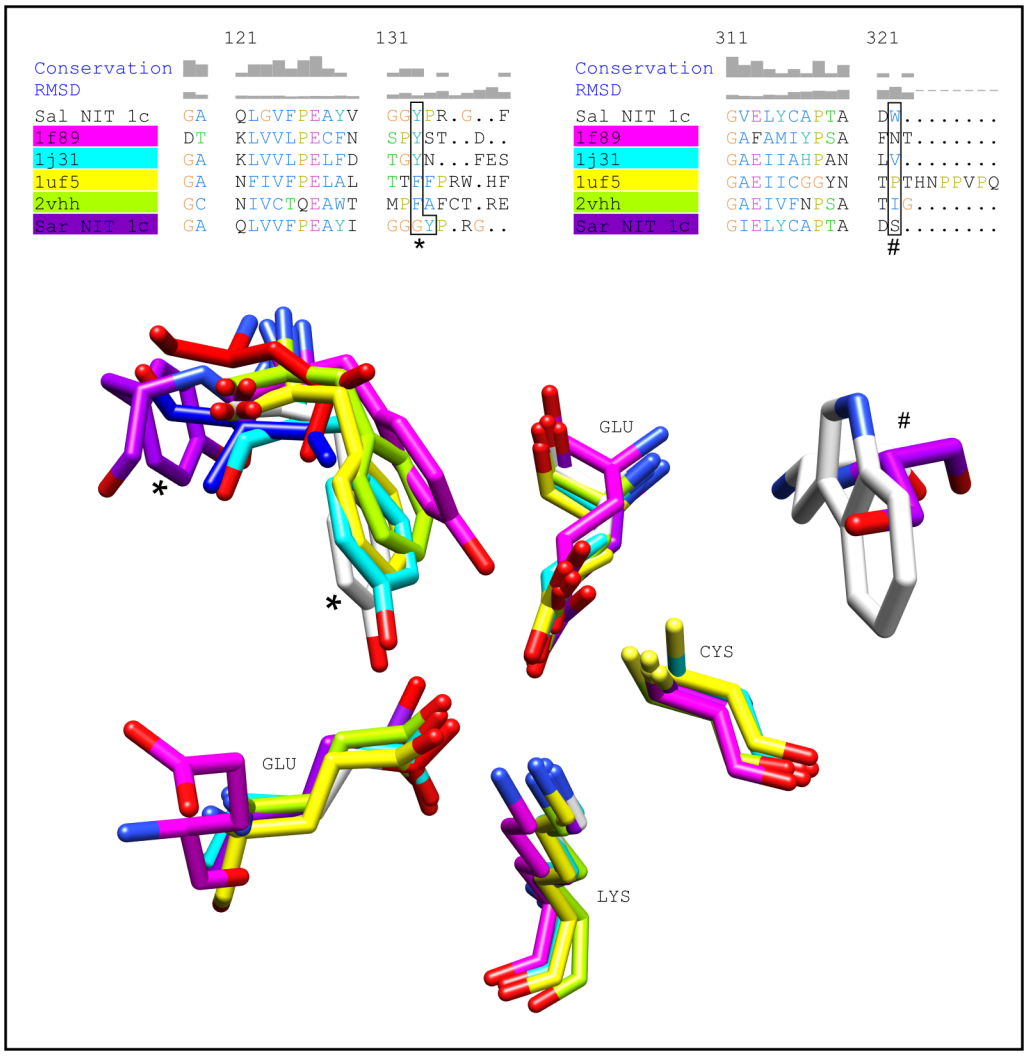


Figure 45. Homology modelling of *Sar*NIT1c and *Sal*NIT1c. The positions of the active site residues of the crystallised homologues (1f89, 1j31, 1uf5 and 2vhh) are shown (the position of the second glutamic acid in the structure 1f89 has not been correctly interpreted in the published structure and has now been reinterpreted by Trevor Sewell). Despite the low sequence alignment of the loop leading to the conserved tyrosine (121-131), the root mean squared difference (RMSD) is very low, indicating that this loop is structurally conserved. The region after the tyrosine (133+) has very low structural conservation: the glycine insertion is therefore predicted to shift the tyrosine by one residue, changing its orientation (*). The serine to tryptophan exchange is also illustrated.

A series of experiments were undertaken to 1) test whether or not the inserted glycine in *Sar*NIT1c is responsible for the difference in the proportion of acid product. 2) Test whether or not the tyrosine is responsible for this difference. 3) Test whether or not this mutation can be applied successfully to a nitrilase 4. The results are summarised in table 2, many thanks to Inga Trompettter from Ruhr University who carried out the site-directed mutagenesis, protein purification and performed the reactions and Niels Agerbirk from the University of Copenhagen who measured the product ratios by HPLC.

Enzyme	*Mutant*	*Proportion of acid*
*Sal*NIT1c	WT	91%
*Sar*NIT1c	WT	17%
*Sar*NIT1c	ΔG66	97%
*Sal*NIT1c	+G66	inactive
*Sar*NIT1c	ΔG66 Y68A	56%
*Sar*NIT1c	ΔG66 Y68F	52%
*Sal*NIT1c	Y68A	inactive
*Sal*NIT1c	Y68F	41%
*At*NIT4	WT	44%
*At*NIT4	Y82F	29%

Table 2. The function of the tyrosine residue in determining the proportion of acid product. *Sal*NIT1c is shown in blue, *Sar*NIT1c in orange and *At*NIT4 in green. Wild-type *Sal*NIT1c produces ~91% acid product, in contrast to wild-type *Sar*NIT1c which produces ~17% acid. When the additional glycine from *Sar*NIT1c is removed by site-directed mutagenesis, the ratio of acid product increases to ~97% of the total. However, when an additional glycine is inserted into *Sal*NIT1c, the enzyme is inactivated. A further four mutants were produced to test whether the presence or absence of the glycine actually has the effect of shifting the tyrosine away from the active site. In the first, the tyrosine was exchanged to alanine and then to phenylalanine in *Sar*NIT1c ΔG66, the proportion of acid product fell from 97% to ~50%. Similarly, the tyrosine in *Sal*NIT1c was exchanged to alanine and phenylalanine, likewise the proportion of acid fell from 91% to ~40 % in the case of phenylalanine, but inactivated the enzyme in the case of alanine. The proportion of acid product changed slightly in the case of *At*NIT4, from 44% for the wild type to 29% for Y82F.

2.2.7 Amide: acid ratio in nitrilase 4: the effect of argine/ lysine

In Section 1.2.9 a model of nitrilase 4 β-cyanoalanine binding is proposed. In the case of *At*NIT4, the carboxyl group of β-cyanoalanine hydrogen bonds to arginine 201 (Fig. 38). This interaction is predicted to be the only interaction required to stabilise the substrate in the active site pocket. However, not all nitrilase 4 enzymes have an

arginine in this position, some have a lysine. Interestingly, in the case of a lysine in this position, the amide:acid ratio is generally increased (Table 3).

NIT	R/K	Amide:acid
*Nt*NIT4A[1]	ICWENR	0,9
*Nt*NIT4B[1]	ICWENR	1,0
*At*NIT4[1]	ICWENR	1,3
*La*NIT4B[2]	ICWENK	3,3
*La*NIT4A[2]	ICWENK	4,0

Table 3. Amide:acid ratio of published NIT4's (total amount of amide – divided by the total amount of acid produced after complete hydrolysis). The proportion of amide is higher with lysine in this position [1]Piotrowski *et al*. (2001), [2]Piotrowski and Volmer (2001). This effect has been demonstrated with a further 18 wild-type NIT4's and 4 mutant NIT4's with exchanged amide:acid ratios.

This phenomenon is consistent with both the model of β-cyanoalanine binding described in section 1.2.9 as well as the nitrilase/ nitrile hydratase mechanism suggested in section 2.2.4. A homology model was generated of *At*NIT4 R201K with β-cyanoalanine bound in the active site pocket (Fig. 46).

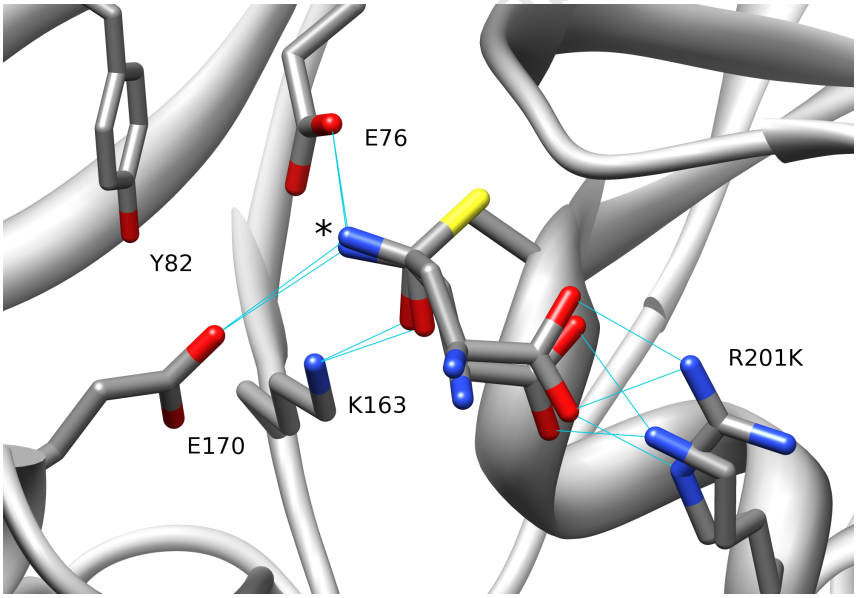


Figure 46. *At*NIT4 homology model of the change in position of β-cyanoalanine when bound to arginine instead of lysine. The tetrahedral intermediate of β-cyanoalanine is required to rotate about the cysteine-carbon bond in order to bring the carboxyl group of the substrate sufficiently close to lysine in order to hydrogen bond. This inevitably comes about because lysine is shorter than arginine. This has the effect of repositioning the amino group of the tetrahedral intermediate (*).

When the carboxyl of the tetrahedral intermediate is bound to lysine instead of arginine a shift in the position of the amide of the tetrahedral intermediate is predicted, presumably this shifts the amide from the optimal position for protonation, thus forming a higher proportion of amide product.

2.3 Discussion

2.3.1 Mechanism modelling

The members of the nitrilase superfamily that are most suitable for understanding the nitrilase catalytic mechanism are the amidases and N-carbamoyl-D-amino acid hydrolases because of the availability of structures from co-crystallized enzyme-substrate complexes, reaction intermediates and ligand-free crystal structures. Amidases (EC 3.5.1.4) catalyse the hydrolysis of amides to the corresponding carboxylic acids and ammonia (Pace and Brenner, 2001). Like all members of the nitrilase superfamily, they contain a conserved (Glu^1-Glu^2-Lys-Cys) catalytic tetrad (Kimani *et al*., 2007). Furthermore, amidase and nitrilase activities are not absolutely mutually exclusive; amidase activity has been measured in the nitrilase from *R. rhodochrous* J1 (Kobayashi *et al*., 1998a) and the nitrilase 4 from *Arabidopsis thaliana* (Piotrowski *et al*., 2001) and nitrilase activity has been detected in the amidase from *R. rhodochrous* J1 (Kobayashi *et al*., 1998b). In terms of modelling the catalytic mechanism of nitrilases, amidases have the advantage that the substrate (amide) is one of the reaction products formed by nitrilases and nitrilases and amidases share carboxylic acid and ammonia as reaction products. Several amidase crystal structures are now available: pdb: 2plq *G. pallidus* RAPc8 (Kimani *et al*., 2007), pdb: 3hkx *Nesterenkonia sp*. (Nel *et al*., 2011), and pdb: 2dyv, 2dyu, 2e2k with bound substrate pdb: 2e2l *H. pylori* (Hung *et al*., 2007), and one trapped reaction intermediate pdb: 2uxy *P. aeruginosa* (Andrade *et al*., 2007).

N-carbamoyl-D-amino acid hydrolases (EC 3.5.1.77) catalyse the hydrolysis of N-carbamoyl-D-amino acids to form D-amino acids, CO_2 and NH_3; there are 5 crystal structures available (1erz, 1uf4, 1uf5, 1uf7, 1uf8). Despite the dissimilarity with nitrilase reaction products, the catalytic residues superimpose precisely onto those of all other crystallized members of the superfamily. The overall protein fold is conserved as are the overall shape of the catalytic pocket and positions of the catalytic pocket residues. With regard to understanding the nitrilase mechanism, 1uf5, 1uf7 and 1uf8 are relevant because they illustrate how large substrates bind into the active site pocket and indicate the positions of the amino- and hydroxyl groups relative to the catalytic residues. A nitrilase superfamily enzyme with some nitrilase activity has recently been crystallized (3ivz: Raczynska *et al*., 2011). While it is unlikely that this

enzyme is primarily a nitrilase (specific activity ~2.4 nKat per mg: Mueller *et al*., 2006) this structure indicates that nitrilase activity is not fundamentally different from that of the crystallized homologues and these can be used to illuminate various aspects of nitrilase catalysis.

2.3.2 Determinants of nitrilase activity

Because the four catalytic residues are present – and superimpose well in all the nitrilase homologue crystal structures solved thus far, the catalytic differences between superfamily members must lie in differences in (hitherto unidentified) residues surrounding the catalytic site. One potential candidate is Tyr82 from *At*NIT4, this residue is present in all known nitrilases, but is not strictly conserved in other superfamily members. A modifying role on nitrilase activity by this residue is suggested by its effect on the nitrile hydratase (amide production) activity of the nitrilase enzymes tested here. The residues that specify other catalytic activities, such as amidase – and N-carbamoyl-D-amino acid hydrolase activity have not been identified. One of the obvious differences between nitrilases and these enzymes is their oligomeric stucture, how this is related to nitrilase catalytic activity compared to for instance, amidase activity remains to be seen. Interestingly, the only plant nitrilase to have significant nitrile hydratase (high amide:acid ratio) activity formed open ring structures and not helices, these structures have also been observed in a bacterial nitrilase 1 (Williamson *et al*., 2007), which also formed a higher than expected proportion of amide.

In section 1 a correlation between the helical association across the C-interface and the preferred substrate size was found. This relationship indicates that the quaternary structure of the enzyme is responsible for maintaining the positions of the amino acid residues lining the binding pocket. An open spiral/lock washer structure, not supported by additional helical interactions, may have the effect of producing a flexible or open substrate-binding pocket. In this scenario, substrates bound in the active site pocket may have the flexibility to reorient the amino group of the tetrahedral intermediate away from the optimal position for protonation; leading to a decrease in the formation of ammonia and therefore acid product. This phenomenon may be related to the observed cyanide hydratase activity (Dumestre *et al*., 1997) of a group of enzymes with a high degree of similarity (Basile *et al*., 2008) to the fungal

nitrilases (Kaplan *et al*., 2011). Two of these so-called cyanide hydratases from fungi have been reconstructed in three dimensions and shown to have small helical twists (Woodward *et al*., 2008; Dent *et al*., 2009). This would predict a preference for large substrates according to the relationship observed in (Section 1). The flexibility of the substrate in the binding pocket may explain why cyanide is hydrolysed to formamide (amide) while larger substrates are converted to an acid product (Kaplan *et al*., 2011).

2.3.3 Nitrilase mechanism

Mahadevan and Thimann (1964) observed that nitrilase activity is inhibited by thiol-specific reagents and electron-donating substrates. On this basis they suggested that the nitrilase catalytic mechanism is initiated by nucleophilic attack of the carbon of the nitrile by a cysteine residue. Hydrolysis proceeds to form a hydroxy-amino intermediate, NH_3 is released and the resulting acyl-enzyme is further hydrolysed to form acid. This reaction mechanism was adapted to account for a small percentage of amide product: in the case of amide formation, the hydroxy-amino covalent tetrahedral intermediate decomposes and the enzyme is the leaving group instead of NH_3 (Hook and Robinson, 1964). Mass spectroscopy indicates that a thioimidate- and acyl-enzyme complexes are found as nitrilase reaction intermediates after acid quenching (Stevenson *et al*., 1992). On this basis they proposed that nucleophilic attack by the cysteine results in a thioimidate intermediate. A glutamic acid-cysteine-lysine catalytic triad was identified by Bork and Koonin, (1994) and these were visualised in nitrilase homologue crystal structures (Pace *et al*., 2000); Nakai *et al*., 2000). Kimani *et al*. (2007) suggested a catalytic role for an additional conserved glutamic acid.

It has been proposed that the role of the glutamic acid (Glu_1) is as a general base catalyst, increasing the nucleophilicity of the cysteine and allowing it to attack the carbon of the nitrile and that the lysine provides electronic stabilisation to the oxyanion of the tetrahedral intermediate (Nakai *et al*., 2000). This arrangement has been visualised in the crystal structure of the amidase from *Pseudomonas aeruginosa* (2uxy) (Andrade *et al*., 2007) that was crystallized with a trapped acyl transfer intermediate with distorted tetrahedral (sp3 hybridization) bound to the active site cysteine. The second glutamic acid is proposed to maintain the side-chain geometry of the catalytic triad and position the substrate within the active site (Hung *et al*., 2007).

Kimani *et al*. (2007) analysed the crystal structure of the amidase from *Geobacillus pallidus* RAPc8 and proposed a further role for Glu_2. The acyl-intermediate appears to block access to Glu_1, while Glu_2 is correctly positioned to act as a general base catalyst for the hydrolysis of the acyl-intermediate.

Nitriles are polar because of the large difference in electronegativity between carbon and nitrogen; which results in a partial negative charge on the nitrogen and a partial positive charge on the carbon atom of a nitrile. On entering the active site, the R-group of the substrate binds to the substrate-binding pocket and the nitrogen of the nitrile is directed close to the oxyanion hole because of the partial negative charge on the nitrogen and the stereochemistry of the nitrile bond. This increases the electrophilic character of the carbon as electrons are drawn away by the catalytic lysine. This arrangement has been visualised in a crystal structure of an amidase with nitrile trapped in the active site (S. Kimani, personal communication). The nucleophilic nature of the catalytic cysteine is enhanced by close proximity to Glu_1 and nucleophilic attack of the carbon by the cysteine results in the formation of a thioimidate intermediate (Fig. 43).

2.3.4 Nitrilase mechanism: amide formation

According to the accepted nitrilase mechanism, the first hydrolysis step produces a covalently bound tetrahedral hydroxy-amino intermediate. Glu_1 has been suggested to be the general base catalyst for this reaction (e.g. Fernandes *et al*., 2006), however in the crystal structure of the amidase from *Geobacillus pallidus* RAPc8, the active site pocket is constricted and access to Glu_1 is blocked by the acyl intermediate (Kimani *et al*., 2007) Glu_2 was therefore proposed as the general base catalyst. At this point the reaction can proceed in one of two ways, the enzyme – or NH_3 can be the leaving group (Hook and Robinson, 1964) and this ultimately depends on whether the C-N or C-S bond is cleaved (Fernandes *et al*., 2006). Protonation of the amino group is required in order for ammonia to be formed (Fernandes *et al*., 2006). Factors, which could potentially lead to amide protonation are therefore interesting from the point of view of determining the reaction product ratio of nitrilases. This explains the observed relationship between electrophilic nature of the substrate R-group and the acid:amide ratio (Fig. 42). An electron-withdrawing R-group decreases the polarity of the C-N

bond, making the amino group less susceptible to protonation, leading to a decrease in the formation of ammonia and a greater proportion of amide reaction product.

It appears as if amide:acid determining mutations fit into one of two classes, those that modify the electrochemical composition of the region surrounding the amino group and those that reposition the amino group of the tetrahedral intermediate away from the optimal region for incipient ammonia stabilisation. Exchanging an alanine for a tryptophan in a position directly adjacent (on the C-terminal side) to the active site cysteine in an (R)-specific acid-forming nitrilase converted it into an enzyme that converted (R,S)-mandelonitrile preferentially to (S)-mandeloamide (Kiziak and Stoltz, 2009). This is presumably because a large amino acid in this position exhibits sufficient steric hindrance towards the (R)-entantiomer of mandelonitrile that conversion is completely inhibited. The stereochemistry of (S)-mandelonitrile presumably allows it to be converted, but repositions the substrate in such a way that amide formation is promoted. Sosedov *et al*. (2010), observed a similar phenomenon with the size of the amino acid residue directly adjacent to the active site cysteine on the N-terminal side.

In this study, a naturally occurring mutation in the substrate-binding pocket product (arginine to lysine exchange) in nitrilase 4 (NIT4) led to differences in the acid:amide ratio of products (table 3). A model of NIT4 binding where the carboxyl-group of β-cyanoalanine binds to R/K in this position was developed based on the crystal structure of an active-site mutant of N-carbamyl-D-amino acid amidohydrolase with bound N-carbamyl-D-methionine (1UF5: Hashimoto *et al*., unpublished) (Fig. 38). The importance of carboxyl-group binding was further demonstrated by showing that 3-cyanopropionic acid and not 3-aminopropionitrile is converted by *At*NIT4 (Fig. 40). The importance of R/K in β-cyanoalanine catalysis by NIT4 was demonstrated by enzyme inactivation with other residues in this position (Markus Piotrowski, personal communication). On the basis of these results, the difference in the length of lysine compared to arginine is expected to lead to a repositioning of the amino group of the tetrahedral intermediate away from the optimal region for ammonia stabilisation (Fig. 44) leading to a decrease in the proportion of acid product.

A THEMATICS (Ko *et al*., 2005) prediction for enzyme active site clusters, based on pKa shifts of amino acid residues <9.0 Å from one another, led to the identification of

a tyrosine residue in a *At*NIT4 homology model as being potentially important for catalytic activity. A tyrosine in this position is universally conserved among nitrilases, but not all nitrilase homologues (e.g. Thuku *et al*., 2009). Interestingly, a shift in the position of this tyrosine as the result of a glycine insertion was observed in a homology model of a nitrilase 1 homologue with a high proportion of amide product (*Sar*NIT1c). Deleting the insertion by site-directed mutagenesis led to a dramatic increase in the proportion of acid product, while additionally mutating the tyrosine to alanine or phenylalanine in this mutant decreased it again (table 2). Potential hydrogen bond acceptors surround the amide of the tetrahedral intermediate, namely Glu_1, a backbone interaction and possibly Tyr (Fig. 44). The role of the tyrosine may therefore be stabilisation of the protonated amino group via hydrogen bonding. This is apparently the first example of a mutation that modifies the acid:amide reaction product ratio by changing the chemical properties of the catalytic site rather than modifying substrate position.

Conclusions and future work

A clear relationship between the helical twist and substrate size was observed in the sample of nitrilases studied here. This correlation is sufficiently strong that, should this relationship prove to be generally applicable, it may prove to be a useful tool for identifying potential substrates of nitrilases of unknown activity. The necessary search space of substrate screens could be limited to those nitriles of approximately the correct size after determining the helical twist. Conversely, the "twist specifying region" which leads to an exchange in helical twist after a single amino acid residue exchange by *in vitro* mutagenesis in *Cr*NIT1a and *Cr*NIT1b may allow nitrilase enzymes to be tailored for a specific substrate. If the underlying principles governing this switch could be better understood there exists the possibility that the twist, and therefore the preferred substrate size, of any nitrilase could be tightly controlled, this may prove to be a fruitful area for future research.

Furthermore, it was demonstrated that exchanging amino acid residues in the region of the predicted substrate-binding pocket altered the specific activity of the nitrilases under study for various substrates. Switching the amino acid residues in two locations of the binding pocket, between two nitrilases with approximately equal helical twist, resulted in a partial substrate specificity exchange. One possibility for future work is to use site-directed mutagenesis with degenerate primer sets in order to produce multiple combinations of mutations at these locations and screen for novel substrate specificity. Another future goal is the rationalisation of substrate-enzyme interactions in order to generate a predictive model of substrate binding; a nitrilase crystal structure would form a key part of this process.

A tyrosine residue was identified that has a modifying effect on the catalytic activity; a high proportion of amide product was formed in a nitrilase with a predicted repositioning of this residue as a result of an insertion. Eliminating this insertion by site-directed mutagenesis led to a large increase in the proportion of acid product. A second mutation, occurring in the predicted substrate-binding pocket, appears to reposition the substrate within the catalytic site. This change is predicted to move the amino group of the tetrahedral intermediate away from the optimal position for protonation, leading to a decrease in ammonia formation and therefore an increased

production of amide. The ability to control the acid:amide ratio of nitrilase reaction products is valuable because it may allow a reduction in the formation of amides which are generally considered to be undesirable side-reaction because their formation reduces the desired carboxylic acid product. On the other hand, the amide forming capacity of nitrilases could be valuble in its own right, as a route to the production of enantiospecific amides.

when applying nitrilase enzymes to biotransformation from an industrial point of view because it enables the products of nitrile hydrolysis to be selected and leads to decrease in waste product.

The ability to control substrate specificity and the reaction product ratio of nitrilase-catalysed biotransformations is attractive. The possibility exists to create a single pH optimised and heat stable nitrilase enzyme that could be specifically tailored to suit particular applications. Work leading to a greater understanding of the details of helical association furthers this end, by leading to the identification of factors that promote helix stability and control of the helical twist. A high-resolution cryo-electron microscopy reconstruction of a nitrilase fibre at sufficient resolution to identify and build individual secondary structure elements into the map would be invaluable in this regard. Crystallographic studies of nitrilases also remain a long-term goal.

References

Agarkar, V.B., Kimani, S.W., Cowan, D.A., Sayed, M.F., Sewell, B.T. (2006). The quaternary structure of the amidase from *Geobacillus pallidus* RAPc8 is revealed by its crystal packing. *Acta Crystallogr F* **62**, 1174-1178.

Andrade, J., Karmali, A., Carrondo, M.A. and Frazao, C. (2007). Structure of amidase from *Pseudomonas aeruginosa* showing a trapped acyl transfer reaction intermediate state. The *J Biol Chem*, **282**, 19598-19605.

Ausubel, R. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., and Struhl, K. (1995). Current protocols in molecular biology. New York: John Wiley & Sons, Inc.

Banerjee, A., Kaul, P., Banerjee, U.C., (2006). Purification and characterization of an enantioselective arylacetonitrilase from *Pseudomonas putida*. *Arch. Microbiol*. **184**, 407–418.

Banerjee, A., Sharma, R. and Banerjee, U.C. (2002). The nitrile-degrading enzymes: current status and future prospects. *Appl Microbiol Biotechnol* **60**, 30-44.

Bartel, B. and Fink, G.R. (1994). Differential regulation of an auxin-producing nitrilase gene family in Arabidopsis thaliana. In: (edn 2 ed.) *Proc Natl Acad Sci* USA **91**, 6649–6653.

Bartling D., Seedorf M., Mithöfer A. and Weiler E.W. (1992). Cloning and expression of an Arabidopsis nitrilase which can convert indole-3-acetonitrile to the plant hormone, indole-3-acetic acid. *Eur J Biochem* **205**, 417–424.

Bartling D., Seedorf M., Schmidt R.C. and Weiler E.W. (1994). Molecular characterization of two cloned nitrilases from Arabidopsis thaliana: key enzymes in the biosynthesis of the plant hormone indole-3-acetic acid. *Proc Natl Acad Sci* USA **91**, 6021–6025.

Basile LJ, Willson RC, Sewell BT & Benedik MJ (2008). Genome mining of cyanide-degrading nitrilases from filamentous fungi. *Appl Microbiol Biotechnol* **80**, 427-35.

Bestwick L., Gronning L., James D., Bones A. and Rossiter J. (1993). Purification and characterization of a nitrilase from Brassica napus. *Physiol Plant* **89**, 611-618.

Bhalla, T.C., Miura, A., Wakamoto, A., Ohba, Y. and Furuhasi, K. (1992). Asymmteric hydrolysis of α-aminonitriles to optically active aminoacids by a nitrilase of *Rhodococcus rhodochrous* PA-34. *Appl Microbiol Biotechnol* **37**, 184-190.

Bradford, 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**, 248-254.

Brady, D., Beeton, A., Zeevaart, J., Kgaje, C., van Rantwijk, F. and Sheldon, R.A. (2004). Characterisation of nitrilase and nitrile hydratase biocatalytic systems. *Appl Microbiol Biotechnol* **64**, 76-85.

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

Bullock, W.O., Fernández, J.M. and Short, J.M. (1987). XL1-Blue: a high efficiency plasmid transforming recA *Escherichia coli* strain with beta-galactosidase selection. *BioTechniques* **5**, 376.

Chan, K-Y., Gumbart, J., McGreevy, R., Watermeyer, J.M., Sewell, B.T. and Schulten, K. (2011). Symmetry-restrained flexible fitting for symmetric EM maps. *Structure*. (In press).

Chiba, R., Dohmoto, M. and Yamaguchi, K. (1999). Oryza sativa, a gene for nitrilase-like protein. GenBank™ accession number: AB027054.

Colebatch, G., Freund, S., Trevaskis, B and Udvardi, M. (2000). Lotus japonicus root nodule ESTs: tools for functional genomics. GenBank™ accession number: AW720658.

Dent KC, Weber BW, Benedik MJ, Sewell BT. (2009). The cyanide hydratase from *Neurospora crassa* forms a helix which has a dimeric repeat. *Appl Microbiol Biotechnol* **82**(**2**), 271-8.

DeSantis, G., Wong, K., Farwell, B., Chatman, K., Zhu, Z., Tomlinson, G., Huang, H., Tan, X., Bibbs, L., Chen, P., Kretz, K. and Burk, M.J. (2003). Creation of a Productive, Highly Enantioselective Nitrilase through Gene Site Saturation Mutagenesis (GSSM) *J Am Chem Soc* **125**, 11476-11477.

DeSantis, G., Zhu, Z., Greenberg, W.A., Wong, K., Chaplin, J., Hanson, S.R., Farewell, B., Nicholson, L.W., Rand, C.L., Weiner, D.P., Robertson, D.E. and Burk, M.J. (2002). An

enzyme library approach to biocatalysis: Development of nitrilases for enantioselective production of carboxylic acid derivatives. *J Am Chem Soc* **124**, 9024-9025.

Dumestre A, Chone T, Portal JM, Gerard M, Berthelin J (1997). Cyanide degradation under alkaline conditions by a strain of *Fusarium solani* from contaminated soils. *Appl Environ Microbiol* **63**, 2729–2734.

Effenberger, E. and Osswald, S. (2001). Selective hydrolysis of aliphatic dinitriles to monocarboxylic acids by a nitrilase from *Arabidopsis thaliana*. *Synthesis* **12**, 1866-1872.

Egelman, E.H. (2000). A Robust Algorithm for the Reconstruction of Helical Filaments Using Single-Particle Methods. *Ultramicroscopy* **85**, 225-234.

Fernandes, B.C.M., Mateo, C., Kiziak, C., Chmura, A., Wacker, J., Van Rantwijk, F., Stolz, A. and Sheldona, R.A. (2006). Nitrile Hydratase Activity of a Recombinant Nitrilase. *Adv Synth Catal* **348**, 2597-2603.

Fiser, A., Do, R.K. and Sali., A. (2000). Modeling of loops in protein structures, *Protein Science* **9** 1753-1773.

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

Gagavan, J.E., DiCosimo, R., Eisenberg, A., Fager, S.K., Folsom, P.W., Hahn, E.C., Schneider, K.J. and Fallon, R.D. (1999). A gram-negative bacterium producing a heat-stable nitrilase highly active on aliphatic dinitriles. *Appl Microbiol Biotechnol* **52**, 654-659.

Ge L, Rudolph P. (1997). Simultaneous introduction of multiple mutations using overlap extension PCR. *Biotechniques*. **22**(**1**), 28–30.

Goddard, J.P. and Reymond, J.L. (2004) Enzyme assays for high-throughput screening. *Curr Opin Biotech* **15**, 314-322.

Gordon JC, Myers JB, Folta T, Shoja V, Heath LS and Onufriev A. (2005). H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **33**, W368-71.

Gray, K.A., Richardson, T.H., Kretz, K., Short, J.M., Bartnek, F., Knowles, R., Kan, L., Swanson, P. E. and Robertson., D.E. (2001). Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Adv Synth Catal* **343**, 607-617.

Griemert, S. (2010). Untersuchungen zu den Nitrilase-4-homologen Enzymen cyanogener Pflanzen. (Diplomarbeit, Ruhr-Universität Bochum)

Hashimoto, H., Aoki, M., Shimizu, T., Nakai, T., Morikawa, H., Ikenaka, Y., Takahashi, S. and Sato, M. (2004). Crystal Structure of C171A⁄V236A Mutant of N-carbamyl-D-amino acid amidohydrolase. RCSB Protein Data Bank (1uf5)

Hiller, A., and Van Slyke, D. (1933). Determination of ammonia in blood. *J Biol Chem* **102**, 499.

Hook, R.H. and Robinson, W.G. (1964). Ricinine Nitrilase II. Purification and properties *J Biol Chem* **239**(**12**), 4263-4267.

Hung, C.L., Liu, J.H., Chiu, W.C., Huang, S.W. and Wang, W-C. (2007). Crystal structure of *Helicobacter pylori* Formamidase AmiF reveals a cysteine-glutamate-lysine catalytic triad. The *J Biol Chem*, **282**, 12220-12229.

Ishikawa, T., Okazaki, K., Kuroda, H., Itoh, K., Mitsui, T. and Hori, H. (2007). Molecular cloning of Brassica rapa nitrilases and their expression during clubroot development. Mol. Plant Pathol. 8 (5), 623-637.

Jandhyala, D.M., Wilson, R.C., Sewell, B.T. and Benedik, M.J. (2005). Comparison of cyanide degrading nitrilases. *Appl Microbiol Biotechnol* **68**, 327-335.

Janowitz, T., Trompetter, I., Piotrowski, M. (2009). Evolution of nitrilases in glucosinolate-containing plants. *Phytochemistry* **70**, 1680-1686.

Jenrich, R., Trompetter, I., Bak, S., Olsen, C.E., Moller, B.L. and Piotrowski, M. (2007). Evolution of heteromeric nitrilase complexes in *Poaceae* with new functions in nitrile metabolism. *Proc Nat Acad Sci* USA **104**, 18848-18853.

Jones, D.T. (1999). GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* **287**, 797-815.

Joyeux, L. and Penczek, P.A. (2002). Efficiency of 2D alignment methods; *Ultramicroscopy* **92**, 33-46.

Kaplan, O., Vejvoda, V., Plihal, O., Pompach, P., Kavan, D., Bojarova, P., Bezouska, K., Mackova, M., Cantarella, M., Jirku, V., Kren, V. and Martinkova, L. (2006). Purification and characterization of a nitrilase from *Aspergillus niger* K10. *Appl Microbiol Biotechnol* **73**, 567-575.

Kaplan O, Bezouška K, Malandra A, Veselá AB, Petříčková A, Felsberg J, Rinágelová A, Křen V, Martínková L. (2011). Genome mining for the discovery of new nitrilases in filamentous fungi. *Biotechnol Lett* **33**(**2**), 309-312.

Kim, J-S, Tiwari, M.K., Moon, H-J., Jeya, M., Ramu, T., Oh, D-K., Kim, I-W., and Lee, J-K. (2009). Identification and characterization of a novel nitrilase from Pseudomonas fluorescens Pf-5. *Appl Microbiol Biotechnol* **83**(**2**), 273-83.

Kimani, S.W., Agarkar, V.B., Cowan, D.A., Sayed, M.F-R. and Sewell, B.T. (2007). Structure of an aliphatic amidase from *Geobacillus pallidus* RAPc8. *Acta Crystallogr* Sect D: *Biol Crystallogr* **63**, 1048-1048.

Kiziak C., Klein J. and Stolz A. (2007). Influence of different carboxy-terminal mutations on the substrate-, reaction- and enantiospecificity of the arylacetonitrilase from Pseudomonas fluorescens EBC191. *Protein Eng Des Sel* **20**, 385-396.

Kiziak, C. and Stolz, A. (2009). Identification of amino acid residues responsible for the enantioselectivity and amide formation capacity of the Arylacetonitrilase from Pseudomonas fluorescens EBC191. *Appl Env Micro* **75**(**17**), 5592-5599.

Ko, J., Murga, L.F., Wei, Y. and Ondrechen, M.J. (2005). Prediction of Active Sites for Protein Structures from Computed Chemical Properties. *Bioinformatics* **21**, i251-i257.

Kobayashi, M., Nagasawa, T. and Yamada, H. (1988). Regiospecific hydrolysis of dinitrile compounds by nitrilase from *Rhodococcus rhodochrous* J1. *Appl Microbiol Biotechnol* **29**, 231-233.

Kobayashi, M., Nagasawa, T. and Yamada, H. (1989). Nitrilase of *Rhodococcus rhodochrous* J1: Purification and characterization. *Eur J Biochem* **182**, 349-356.

Kobayashi, M., Goda, M. and Shimizu, S. (1989a). Nitrilase Catalyzes Amide Hydrolysis as Well as Nitrile Hydrolysis. *Biol. Biophys. Res. Comm* **253**, 662-666.

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

Kobayashi, M., Yanaka, N., Nagasawa, T. and Yamada, H. (1990). Purification and characterization of a novel nitrilase of *Rhodococcus rhodochrous* K22 that acts on aliphatic nitriles. *J Bacteriol* **172**, 4807-4815.

Kobayashi, M., Yanaka, N., Nagasawa, T. and Yamada, H. (1992). Primary structure of an aliphatic nitrile-degrading enzyme, aliphatic nitrilase, from *Rhodococcus rhodochrous* K22 and expression of its gene and identification of the active site residue. *Biochem* **31**, 9000-9007.

Kriechbaumer, V., Park, W.J., Piotrowski, M., Meeley, R.B., Gierl, A. and Glawischnig, E. (2007). Maize nitrilases have a dual role in auxin homeostasis and β-cyanoalanine hydrolysis. *J Exp Bot* **58**, 4225-4233.

Krivov, G.G., Shapovalov, M. V. and Dunbrack, R. L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77**(**4**), 778-95.

Kukushkin, V. Yu., Pombeiro, A.J.L. (2005). Metal-mediated and metal-catalyzed hydrolysis of nitriles (a review). *Inorg. Chim. Acta*, **358**, 1–21.

Kumaran, D., Eswaramoorthy, S., Gerchman, S.E., Kycia, H., Studier, F.W. and Swaminathan, S. (2003). Crystal structure of putative CN hydrolase from yeast. *PROTEINS: Struct Funct Genet* **52**, 283–291.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**(**5259**), 680-685.

Lundgren, S., Lohkamp, B., Andersen, B., Piskur, J. and Dobritzsch, D. (2008). The crystal structure of B-alanine synthase from *Drosophila melanogaster* reveals a homoactameric helical turn-like assembly. *J Mol Biol* **377**, 1544-1559.

Mahadevan, S. and Thimann, K. V. (1964). Nitrilase: II. Substrate specificity and possible mode of action. *Arch Biochem Biophys* **107**, 62-68.

Martínková, L., Křen, V. (2002). Nitrile- and amide-converting microbial enzymes: stereo-, regio- and chemoselectivity. *Biocat Biotrans* **20**(**2**), 73-93.

McGuffin, L.J., Bryson, K. and Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-405.

McGuffin, L.J. and Jones, D.T. (2003). Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics* **19**, 874-881.

Menten, L.; Michaelis, M.I. (1913). "Die Kinetik der Invertinwirkung". Biochem Z 49: 333–369.

Meyers, P.R., Rawlings, D.E., Woods, D.R., and Lindsey, G.G. (1993). Isolation and characterization of a cyanide dihydratase from *Bacillus pumilus* C1. *J Bacteriol* **175**, 6105-6112.

Mueller, P., Egarova, K., Vorgias, E.C., Boutou, E., Trauthwein, H., Verseck, S. and Antranikian, G. (2006). Cloning, expression and characterization of a thermoactive nitrilase from the hyperthermophile archaeon *Pyrococcus abyssi*. *Protein Expression Purif* **47**, 672-681.

Myers J, Grothaus G, Narayanan S, Onufriev A, (2006). A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins*, **63**, 928-938.

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**, 138-144.

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**, 729–739.

Nel, A.J., Tuffin, I.M., Sewell, B.T., Cowan, D.A. (2011). Unique aliphatic amidase from a psychrotrophic and haloalkaliphilic *Nesterenkonia* isolate. *Appl Environ Microbiol* **77**, 3696-3702.

Nolan, L.M., Harnedy, P.A., Turner, P., Hearne, A.B. and O'Reilly, C. (2003). The cyanide hydratase enzyme of *Fusarium lateritium* also has nitrilase activity. *FEMS Microbiol Lett* **221**, 161-165.

O'Reilly, C. and Turner, P.D. (2003). The nitrilase family of CN hydrolyzing enzymes – a comparative study. *J Appl Microbiol* **95**, 1161-1174.

Osswald, S., Wajant, H. and Effenberger, F. (2002). Characterization and synthetic applications of recombinant AtNIT1 from *Arabidopsis thaliana*. *Eur J Biochem* **269**, 680-687.

Pace, H. and Brenner, C. (2001). The nitrilase superfamily: classification, structure and function. *Genome Biol* **2**, 1-9.

Penczek P, Radermacher M, Frank J. (1992). Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy* **40**(**1**), 33-53.

Penczek, P.A., Zhu, J. and Frank J. (1996). A common-lines based method for determining orientations for N > 3 particle projections simultaneously. *Ultramicroscopy* **63**, 205-218.

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**, 1605-1612.

Piotrowski M. (2008). Primary or secondary? Versatile nitrilases in plant metabolism. Phytochemistry **69**(**15**), 2655-2667.

Piotrowski, M., Schonfelder, S. and Weiler, E.R. (2001). The *Arabidospsis thaliana* isogene NIT4 and its orthologs in Tobacco encode β -cyano-L-alanine hydratase/nitrilase. *J. Biol Chem* **276**, 2616-2621.

Piotrowski, M., Volmer, J. J. (2006). Cyanide metabolism in higher plants: Cyanoalanine hydratase is a NIT4 homolog. *Plant Mol Biol* **61**, 111-112.

Pollmann, S., Mueller, A., Piotrowski, M., Weiler, E.W. (2002). Occurance and formation of indole-3-acetamide in Arabidopsis thaliana. *Planta* **216**(**1**), 155-61.

Quillin, M.L. and Matthews, B.W. (2000). Accurate calculation of the density of proteins. *Acta Cryst D* **56**, 791-794.

Raczynska, J.E., Vorgias, C.E., Antranikian, G. and Rypniewski, W. (2011). Crystallographic analysis of a thermoactive nitrilase. *J Struct Biol* **173**, 294-302.

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

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: Struct Funct Bioinform* **57**, 869–873.

Sanderson, R. T. (1976). Chemical Bond and Bond Energies. *Academic Press*: New York

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**, 1-20.

Sewell, B.T, Thuku, R.N., Zhang, X. and Benedik, M.J. (2005). The oligomeric structure of nitrilases: the effect of mutating interfacial residues on activity. *Ann NY Acad Sci* **1056**, 153-159.

Short, J. M. (1997). Recombinant approaches for assessing biodiversity of nitrilases for enantioselective production of carboxylic acid derivatives. *Nat Biotechnol* **15**, 1322-1323.

Singh, R., Sharma, R., Tewari, N., Geetanjali, and Rawat, D.S. (2006). Nitrilase and its application as a 'green' catalyst. *Chem Biodivers* **3**, 1279-1287.

Sosedov, O., Baum, S., Bürger, S., Matzer, K., Kiziak, C. and Stoltz, A. (2010). Construction and Application of Variants of the Pseudomonas fluorescens EBC191 Arylacetonitrilase for Increased Production of Acids or Amides. *Appl Env Micro* **76**(**11**), 3668-3674.

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.

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.

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**, 2099–2018.

Trompetter, I. (2010) Untersuchungen zu den NIT1-homologen Nitrilasen der Brassicaceae (Dissertation, Ruhr-Universität Bochum)

Tsunoda, H. and Yamaguchi, K. (1995). The cDNA sequence of an auxin-producing nitrilase homologue in tobacco (Accession No. D63331). *Plant Physiol* **109**, 339.

Unser, M., Trus, B.L., Steven, A.C. (1987). A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy* **23**, 39-51.

Vejvoda, V., Kaplan, O., Bezouska, K., Pompach, P., Sulc, M., Cantarella, M., Benada, O., Uhnakova, B., Rinagelova, A., Lutz-Wahl, S., Fischer, L., Kren, V. and Martinkova, L. (2008). Purification and characterization of a nitrilase from *Fusarium solani* O1. *J Mol Catal B: Enzym* **50**, 99-106.

Volkmann, N. and Hanein, D. (1999). Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J Struct Biol* **128**, 223-223.

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**, 251-261.

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., Sewell, B.T. (2010). Structural and biochemical characterization of a nitrilase from the thermophilic bacterium, Geobacillus pallidus RAPc8. *Appl Microbiol Biotechnol* **88**,143–153.

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*. *J Struct Biol* **161**, 111-119.

Wurch T., Lestienne F., Pauwels P.J. (1998). A modified overlap extension PCR method to create chimaeric genes in absence of restriction enzymes. *Biotechnol Tech.* **12**, 653-657.

Yamamoto, K., Oishi, K., Kawakami, K. and Komatsu, K. (1991). Production of R-(-)-Mandelic acid from a Mandelonitrile by *Alcaligenes faecalis* ATCC 8750. *Appl Environ Microbiol* **57**, 3028-3032.

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 *App Env Micro* **56**(**10**), 3125-3129.

Zheng L., Baumann U. and Reymond J.L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32**(**14**), e115.

Zhu, D., Murkherjee, C., Biehl, E.R. and Hua, L. (2007). Discovery of a mandelonitrile hydrolase from *Bradyrhizobium japonicum* USDA110 by rational genome mining. *J Biotechnol* **129**, 645-650.

Appendix

1. Nitrilase sequences

Nitrilase 4 group

*At*NIT4 – *Arabidopsis thaliana* nitrilase 4 (Bartel and Fink, 1994)

*Cl*NIT4 - *Curcuma longa* nitrilase 4 (Markus Piotrowski, pers. comm.)

*Cp*NIT4 - *Carica papaya* nitrilase 4 (Papaya genome sequencing consortium)

*Cr*NIT4 - *Capsella rubella* nitrilase 4 (Janowitz *et al*., 2009)

*Cs*NIT4 - *Cleome spinosa* nitrilase 4 (Janowitz *et al*., 2009)

*Gm*NIT4 - *Glycine max* nitrilase 4 (DOE joint genome institute)

*Vv*NIT4 - *Vitis vinifera* nitrilase 4 (International grape genome program)

*Pt*NIT4 - *Pinus taeda* nitrilase 4 (The pine genome initiative)

*Sm*NIT4-1 - *Selaginella moellendorffii* nitrilase 4-1 (Selaginella consortium)

*Sm*NIT4-2 - *Selaginella moellendorffii* nitrilase 4-2 (Selaginella consortium)

*Tm*NIT4 - *Tropaeolum majus* nitrilase 4 (Trompetter, 2010)

*Mc*NIT4 - *Mesembryanthemum crystallinum* NIT4 (Markus Piotrowski, pers. comm.)

*Me*NIT4-1 - *Manihot esculenta* nitrilase 4-1 (Markus Piotrowski, pers. comm.)

*Me*NIT4-2 - *Manihot esculenta* nitrilase 4-2 (Markus Piotrowski, pers. comm.)

*Ls*NIT4 - *Lactuca sativa* nitrilase 4 (Compositae genome project)

*Lu*NIT4 - *Linum usitatissimum* nitrilase 4 (Markus Piotrowski, pers. comm.)

*Le*NIT4 - *Lycopersicon esculentum* nitrilase 4 (Markus Piotrowski, pers. comm.)

*Mt*NIT4 - *Medicago truncatula* nitrilase 4 (Medicago consortium)

*Potr*NIT4 - *Populus trichocarpa* nitrilase 4 (Jenrich *et al*., 2007)

*Pp*NIT4 - *Physcomitrella patens* nitrilase 4 (Moss genome consortium)

*Zm*NIT4A - *Zea mays* nitrilase 4A (Jenrich *et al*., 2007)

*So*NIT4A - *Saccharum officinarum* nitrilase 4A (Jenrich *et al*., 2007)

*Sb*NIT4A - *Sorghum bicolor* nitrilase 4A (M.M. Cordonnier-Pratt, pers. comm.)

*La*NIT4A - *Lupinus angustifolius* nitrilase 4A (Piotrowski and Volmer, 2005)

*Ta*NIT4A - *Triticum aestivum* nitrilase 4A (Jenrich *et al*., 2007)

*Lj*NIT4A - *Lotus japonicus* nitrilase 4A (Colebatch *et al*., 2000)

*Os*NIT4A - *Oryza sativa* nitrilase 4A (Chiba *et al*., 1999)

*Nt*NIT4A - *Nicotiana tabacum* nitrilase 4A (Tsunoda and Yamaguchi, 1995)

*Hv*NIT4A - *Hordeum vulgare* nitrilase 4A (Jenrich *et al*., 2007)

Nitrilase 1 group

*Br*NIT-T1 - *Brassica rapa* nitrilase 1	(Ishikawa *et al*., 2007)
*Br*NIT-T2 – *Brassica rapa* nitrilase 2	(Ishikawa *et al*., 2007)
*At*NIT2 - *Arabidopsis thaliana* nitrilase 2	(Bartling *et al*., 1994)
*At*NIT1 - *Arabidopsis thaliana* nitrilase 1	(Bartling *et al*., 1992)
*Bn*NIT1B - *Brassica napus* nitrilase 1B	(Bestwick *et al*., 1993)
*Bn*NIT1A - *Brassica napus* nitrilase 1A	(Bestwick *et al*., 1993)
*Sal*NIT1A – *Sinapis alba* 1A	(Trompetter, 2010)
*Sar*NIT1A – *Sinapis arvensis* 1A	(Trompetter, 2010)
*Sar*NIT1B – *Sinapis arvensis* 1B	(Trompetter, 2010)
*Cr*NIT1A - *Capsella rubella* nitrilase 1A	(Janowitz *et al*., 2009)
*Cr*NIT1B - *Capsella rubella* nitrilase 1B	(Janowitz *et al*., 2009)
*Sal*NIT1C – *Sinapis alba* nitrilase 1C	(Trompetter, 2010)
*Sar*NIT1C – *Sinapis arvensis* 1C	(Trompetter, 2010)
*Cr*NIT3 - *Capsella rubella* nitrilase 3	(Janowitz *et al*., 2009)
*At*NIT3 – *Arabidopsis thaliana* nitrilase 3	(Bartel and Fink, 1994)